



دانشگاه یزد
دانشکده مهندسی برق و کامپیوتر
گروه مهندسی کامپیوتر

پایان نامه
برای دریافت درجه کارشناسی ارشد
مهندسی کامپیوتر – هوش مصنوعی

رتبه بندی تصاویر بر مبنای پرس و جوی کاربر

استاد راهنما
دکتر علی محمد زارع بیدکی

استاد مشاور
دکتر مهدی رضائیان

پژوهش و نگارش
سید کاظم میردهقان

شهریورماه ۱۳۹۲

تقدیم به

روح پاک پدرم که به من آموخت تا چگونه در عرصه زندگی، ایستادگی را تجربه کنم
به مادرم، دریای بی کران فداکاری و عشق که وجودم برایش همه رنج بود و وجودش
برایم همه مهر و به همسرم، اسطوره زندگی‌ام، پناه خستگی‌ام و امید بودم.

تقدیر و تشکر

شکر شایان نثار ایزد منان که توفیق را رفیق راهم ساخت تا این پایان نامه را به پایان برسانم. سپس از پدر و مادر که نخستین آموزگارانم بودند تا معلمین، دبیران و استادانم که هر یک به زعم خویش بر آموخته هایم افزودند، قدردانی می نمایم. مراتب سپاس خود را از جناب آقای دکتر علی محمد زارع بیدکی به عنوان استاد راهنما که همواره نگارنده را مورد لطف و محبت خود قرار داده اند، ابراز می دارم. همچنین از جناب آقای دکتر مهدی رضائیان، استاد مشاور اینجانب که از رهنمودهای ایشان در این پژوهش بهره برده ام، کمال تشکر را دارم.

چکیده

جستجو در اینترنت کنونی به امری پیچیده تبدیل شده است. دلیل این امر رشد روزافزون در تصاویر موجود در وب می‌باشد. برای جستجوی تصاویر دلخواه کاربر، به یک موتور جستجوی^۱ کارآمد نیاز است تا در کمترین زمان ممکن، متناسب‌ترین تصاویر با پرس‌وجوی^۲ کاربر را به او ارائه نماید. چالش اصلی موتورهای جستجو، رتبه‌بندی^۳ تصاویر بازبازی شده از پرس‌وجوی کاربر است. برای رتبه‌بندی تصاویر الگوریتم‌های مختلفی وجود دارد که می‌توان آن‌ها را به پنج دسته الگوریتم‌های مبتنی بر محتوای متنی^۴، الگوریتم‌های مبتنی بر محتوای دیداری^۵، الگوریتم‌های مبتنی بر اتصال^۶، الگوریتم‌های مبتنی بر رفتار کاربر^۷ و الگوریتم‌های ترکیبی^۸ تقسیم نمود. برخی از این الگوریتم‌ها از دقت کافی برخوردار نیستند و برخی دیگر دچار درجه‌پیچیدگی بالا در پردازش تصویر می‌باشند.

هدف اصلی این پژوهش، مروری بر الگوریتم‌های رتبه‌بندی تصاویر بازبازی شده از پرس‌وجوی متنی کاربر و بررسی روش عملکرد و نقاط ضعف و قوت هر کدام می‌باشد. در ادامه روش جدیدی برای رتبه‌بندی تصاویر برای استفاده در موتورهای جستجو پیشنهاد می‌گردد که ترکیبی از رتبه‌بندی بر اساس محتوای متنی و به کارگیری محتوای دیداری تصاویر به منظور ادغام اسناد با تصاویر مشابه می‌باشد. این ادغام باعث غنی‌تر شدن محتوای متنی اسناد تصویری شده و دقت جستجو را افزایش می‌دهد. سپس با استفاده از یک الگوریتم یادگیری جفتی^۹ و مجموعه داده محک تولیدی که توسط افراد خبره برچسب‌گذاری شده است، وزن فیلدهای متنی مورد جستجو را تنظیم می‌کنیم. نتایج آزمایش‌ها نشان می‌دهد این روش می‌تواند دقت رتبه‌بندی را تا حد بالایی افزایش دهد.

کلمات کلیدی: تصویر، رتبه‌بندی، روش ترکیبی، موتور جستجو

¹ Search engine

² Query

³ Ranking

⁴ Text-based

⁵ Content-based

⁶ Link-based

⁷ User behavior

⁸ Combinational

⁹ Pair-wise Learning

فهرست مطالب

۱	مقدمه	۱
۲	۱-۱ معماری موتور جستجوی تصویر	۲
۵	۲-۱ رتبه‌بندی تصاویر در وب	۵
۷	۳-۱ چالش‌های رتبه‌بندی تصاویر در وب	۷
۷	۱-۳-۱ مفهوم مرتبط بودن	۷
۸	۲-۳-۱ توصیفگرهای تصویر	۸
۸	۳-۳-۱ وابستگی به کاربر	۸
۹	۴-۳-۱ شباهت یا تطبیق	۹
۹	۵-۳-۱ نمایه‌ها	۹
۹	۶-۳-۱ نوع پرس‌وجو	۹
۱۰	۷-۳-۱ بازیابی تصاویر	۱۰
۱۱	۴-۱ ارزیابی	۱۱
۱۱	۱-۴-۱ مجموعه داده محک	۱۱
۱۲	۲-۴-۱ معیارهای ارزیابی	۱۲
۱۳	۵-۱ ساختار پایان‌نامه	۱۳
۱۴	۲ مروری بر کارهای گذشته	۱۴
۱۴	۱-۲ روش‌های مبتنی بر محتوای متنی	۱۴
۱۵	۱-۱-۲ مدل فضای برداری	۱۵
۱۷	۲-۱-۲ مدل احتمالی	۱۷
۱۹	۲-۲ روش‌های مبتنی بر محتوای دیداری تصاویر	۱۹

۲۱	۱-۲-۲ الگوریتم VisualRank
۲۳	۳-۲ روش‌های مبتنی بر اتصال
۲۴	2-3-1 الگوریتم PageRank
۲۶	۲-۳-۲ الگوریتم HITS
۲۹	۳-۳-۲ الگوریتم PicASHOW
۲۹	۴-۲ روش‌های مبتنی بر رفتار کاربر
۳۰	۵-۲ روش‌های ترکیبی
۳۱	۳ روش پیشنهادی
۳۱	۱-۳ مقدمه
۳۱	۲-۳ الگوریتم پیشنهادی
۳۳	۱-۲-۳ گام اول: شناسایی تصاویر یکسان و مقیاس‌شده
۳۷	۲-۲-۳ گام دوم: ادغام اسناد تصویری یکسان و مقیاس‌شده و نمایه‌سازی آن‌ها
۴۲	۳-۲-۳ گام سوم: فرآیند جستجوی تصویر
۴۲	۳-۳ پیاده‌سازی روش پیشنهادی و ایجاد مجموعه داده محک
۴۴	۴-۳ بهبود روش پیشنهادی با روش یادگیری جفتی
۴۶	۵-۳ ارزیابی روش‌های پیشنهادی
۴۹	۴ نتیجه‌گیری و کارهای آینده
۴۹	۱-۴ نتیجه‌گیری
۴۹	۲-۴ دستاوردهای پایان‌نامه
۵۰	۳-۴ کارهای آینده
۵۲	پیوست الف. تعریف واریانس و میانگین

شخص ۶۷

مراجع ۷۰

فهرست شکل‌ها

- شکل ۱-۱: معماری یک موتور جستجوی تصویر ۳
- شکل ۲-۱: مثالی از نمایه‌ساز معکوس ۴
- شکل ۳-۱: فرآیندهای برخط و برون خط موتور جستجو ۵
- شکل ۴-۱: تصویر به عنوان پرس‌وجو در موتور جستجوی گوگل ۱۰
- شکل ۱-۲: دسته‌بندی روش‌های رتبه‌بندی ۱۴
- شکل ۲-۲: گراف شباهت در الگوریتم VisualRank ۲۲
- شکل ۳-۲: مثالی از الگوریتم PageRank ۲۵
- شکل ۴-۲: شمای کلی رتبه‌بندی مبتنی بر رفتار کاربر ۳۰
- شکل ۱-۳: ادغام اسناد تصویری یکسان و مقیاس شده ۳۲
- شکل ۲-۳: روندنمای الگوریتم تولید شناسه برای هر تصویر ۳۴
- شکل ۳-۳: نمونه‌ای از یک تصویر رقمی خاکستری ۳۵
- شکل ۴-۳: نمونه‌ای از یک تصویر رقمی رنگی ۳۵
- شکل ۵-۳: محاسبه شدت نور پیکسل تصویر خاکستری ۳۶
- شکل ۶-۳: تولید کد درهم از تصویر خاکستری ۳۷
- شکل ۷-۳: نمونه‌ای از تصویر درون‌هاست ۳۸
- شکل ۸-۳: نمونه‌ای از تصویر درون‌دامنه ۳۹
- شکل ۹-۳: نمونه‌ای از تصویر خارجی ۳۹
- شکل ۱۰-۳: سطوح مختلف محتوای متنی اطراف تصویر ۴۰
- شکل ۱۱-۳: نحوه قرار گرفتن تصویر در صفحه وب ۴۱
- شکل ۱۲-۳: واسط کاربر سامانه ایجاد مجموعه داده محک ۴۳

شکل ۳-۱۳: ارزیابی روش پیشنهادی پیش و پس از تنظیم وزن‌ها در معیار P@n.....۴۷

شکل ۳-۱۴: ارزیابی روش پیشنهادی پیش و پس از تنظیم وزن‌ها در معیار دقت متوسط ۴۸

فهرست جدول‌ها

جدول ۱-۲: متغیرهای استفاده‌شده در روش BM25[۵]..... ۱۹

۱ مقدمه

استفاده از وب به عنوان منبع اصلی کسب اطلاعات، با رشد شدیدی همراه بوده است. دلیل این امر، رفع نیاز اطلاعاتی^۱ کاربران می باشد. کاربران نیاز دارند تا در کمترین زمان ممکن به اطلاعات مورد نیاز خود دست یابند. از طرف دیگر، اطلاعات موجود روی وب با رشدی نمایی همراه بوده و در قالب های مختلف ظاهر شده اند. این اطلاعات با سرعت بالایی در حال تغییر می باشند. از این رو موتورهای جستجو، ابزارهایی کارآمد در این زمینه هستند. کاربران، پرس و جوی خود را توسط تعداد محدودی کلیدواژه^۲ مطرح کرده و موتور جستجو مرتبط ترین و باکیفیت ترین نتایج با پرس و جوی کاربر را به او ارائه می کند. بر اساس آمار سایت معتبر alexa، موتورهای جستجوی گوگل^۳ و یاهو^۴ به ترتیب در مقام های اول و چهارم پر بازدیدترین پایگاه های اینترنتی جهان [۱] و مقام های اول و دوم پر بازدیدترین پایگاه های ایران قرار گرفته اند [۲]. بر اساس آمار منتشر شده در همین سایت، از نظر رتبه بندی ترافیک منطقه ای^۵، ایران رتبه اول ترافیک جهانی گوگل [۳] و رتبه دوم ترافیک جهانی یاهو [۴] را در اختیار دارد.

یکی از این داده های رقمی، تصویر می باشد. تصویر بخش بزرگی از اطلاعات وب را شامل می شود. در بازیابی تصویر^۶، هدف یافتن تصاویر مرتبط با پرس و جوی کاربر خاص، از میان مجموعه ای معمولاً بزرگ از تصاویر است. بازیابی می تواند بر روی مجموعه های خاص منظوره^۷ و یا از بین مجموعه ای عمومی از تصاویر انجام پذیرد. به عنوان نمونه بازیابی تصاویر ماهواره ای و بازیابی تصاویر پزشکی در دسته تصاویر خاص منظوره قرار می گیرند. کاربر می تواند پرس و جوی خود را توسط تعدادی کلیدواژه و یا یک تصویر به موتور جستجو اعلام کند. ولی در اینجا منظور از پرس و جو، یک عبارت متنی متشکل از تعدادی کلیدواژه می باشد. پس از این که کاربر پرس و جوی خود را توسط تعدادی

¹ Information need

² Keyword

³ Google

⁴ Yahoo

⁵ Regional traffic ranks

⁶ Image Retrieval

⁷ Domain Specific

کلیدواژه به موتور جستجو اعلام می‌کند، موتور جستجو در پاسخ به پرس‌وجوی کاربر ممکن است تعداد زیادی نتیجه را به‌عنوان نتیجه مرتبط کاندید کند. ولی کاربر معمولاً به ۱۰ تا ۲۰ نتیجه اول اکتفا می‌کند [۵] و در صورتی که به هدف خود نرسد، پرس‌وجوی خود را تغییر داده و یا از جستجو منصرف می‌شود. در نتیجه رتبه‌بندی نتایج بازایی شده از اهمیت بالایی برخوردار است تا جایی که می‌تواند میزان محبوبیت^۱ یک موتور جستجو را تعیین کند. به همین منظور در این پایان‌نامه، تمرکز ما بیشتر بر روی الگوریتم‌های رتبه‌بندی خواهد بود. ولی برای آشنایی با چگونگی عملکرد یک موتور جستجو، ابتدا ساختار یک موتور جستجوی تصویر را باهم بررسی خواهیم نمود و سپس به شرح الگوریتم خواهیم پرداخت.

۱-۱ معماری موتور جستجوی تصویر

در شکل ۱-۱، معماری یک موتور جستجوی تصویر ساده را مشاهده می‌کنید. در مرحله اول، یک یا چند خزش‌گر^۲ به‌صورت موازی صفحات وب و تصاویر داخل هر صفحه را خزش نموده و آن‌ها را در مخزن^۳ تصاویر و صفحات وب ذخیره می‌کند. فرآیند خزش به صورت کاملاً هوشمند^۴ و خودکار^۵، انجام می‌گیرد. خزش‌گر، اسناد وب را بر مبنای پوشش^۶ و تازگی^۷ جمع‌آوری می‌کند. پوشش به معنای این است که خزش‌گر بتواند هرچه بیشتر از صفحات و تصاویر موجود روی وب را پوشش دهد. از طرف دیگر یک خزش‌گر باید بتواند صفحات و تصاویر را به‌روز نگه دارد. به دلیل حجم بالا، اطلاعات خزش شده به صورت فشرده^۸ ذخیره می‌شود.

^۱ Popularity

^۲ Crawler

^۳ Repository

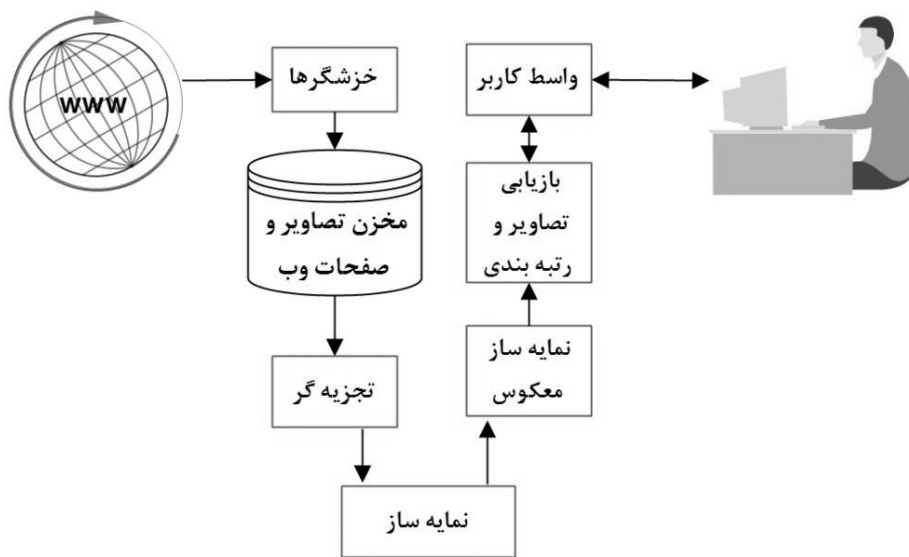
^۴ Intelligent

^۵ Automatic

^۶ Coverage

^۷ Freshness

^۸ Compressed



شکل ۱-۱: معماری یک موتور جستجوی تصویر

پس از این مرحله، محتوای صفحات توسط تجزیه گر^۱، استخراج شده و پیش پردازش^۲هایی روی آن انجام می گیرد. صفحات پیش پردازش شده به نمایه ساز^۳ تحویل داده شده و کلمات داخل صفحات استخراج شده و ساختمان داده ای به نام نمایه^۴ ایجاد می شود. نمایه نشان می دهد در هر صفحه چه واژگانی وجود دارد. ولی اکثر موتورهای جستجو از نمایه معکوس^۵ استفاده می کنند. در نمایه معکوس، مشخص می شود که هر واژه در چه اسنادی^۶ آمده است. با استفاده از نمایه معکوس قادریم تا با سرعت بالا، اسنادی که حاوی یک کلمه خاص هستند را بیابیم. نمایه ساز معکوس از دو قسمت تشکیل شده است. قسمت واژگان، که شامل تمام واژگان داخل اسناد می باشد و لیست رخداد مکانی این واژگان، که شامل تمام اسنادی است که شامل آن واژه هستند. شکل ۱-۲ مثالی از نمایه سازی معکوس برای سه سند را نمایش می دهد.

¹ Parser

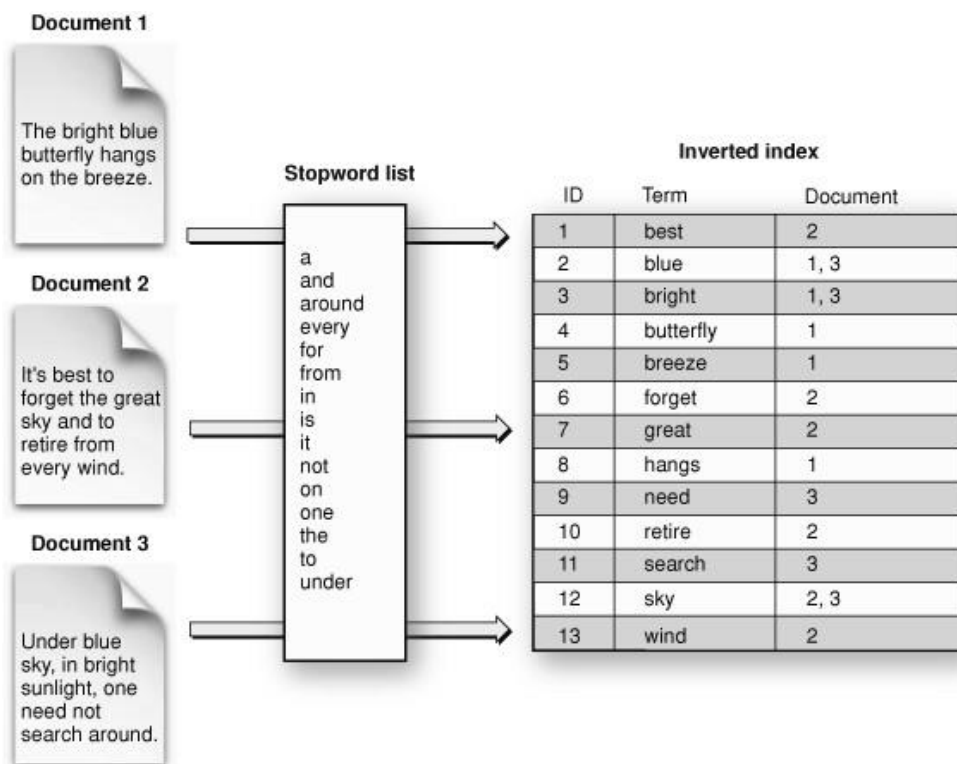
² Preprocessing

³ Indexer

⁴ Index

⁵ Inverted index

⁶ Documents



شکل ۱-۲: مثالی از نمایه‌ساز معکوس

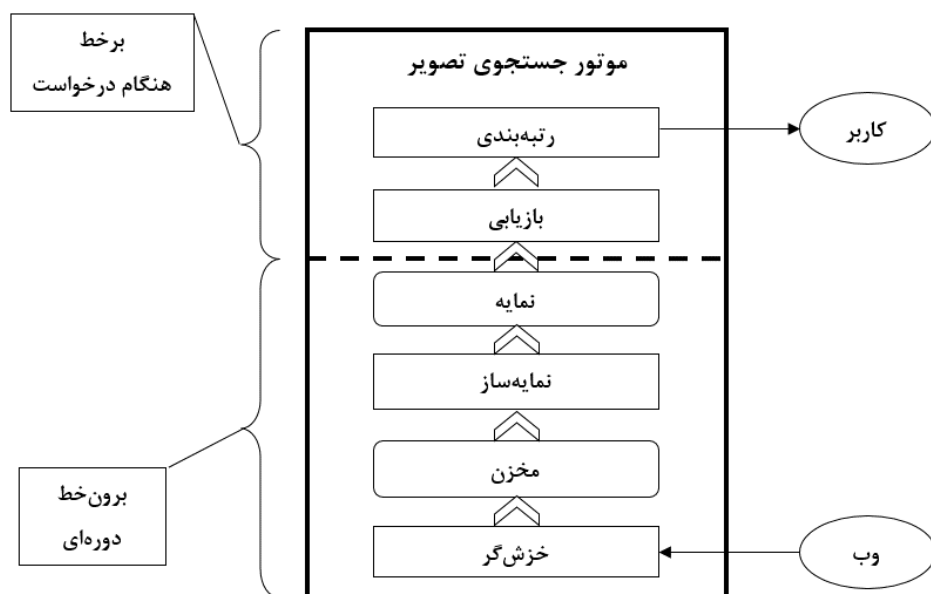
هدف از نمایه‌سازی استخراج یک نمای منطقی از اسناد می‌باشد. مرسوم‌ترین نمای منطقی در موتورهای جستجو، مدل بسته‌ی کلمات^۱ می‌باشد. در این مدل، هر سند از مجموعه‌ای از لیست نامرتب کلمات تشکیل شده است. متن اسناد باید قبل از نمایه‌سازی نرمال شود. هدف از نرمال‌سازی متن، استخراج کلمات کلیدی از سند و حذف حروف اضافی می‌باشد[۶].

در این گام، کاربر پرس‌وجوی خود را از طریق واسط کاربر به موتور جستجو اعلام می‌کند. ماژول بازیابی^۲ و رتبه‌بندی، تصاویر مرتبط با آن پرس‌وجو را بازیابی نموده و بر اساس الگوریتم رتبه‌بندی، نتایج جستجو را مرتب نموده و از طریق واسط کاربر، به کاربر ارائه می‌دهد. واسط کاربر به صورت مستقیم با کاربر رابطه دارد. از یک طرف پرس‌وجوی کاربر را دریافت نموده و از طرف دیگر نتایج بازیابی شده پس از رتبه‌بندی را به کاربر نشان می‌دهد.

¹ Bag of words

² Retrieval

یک موتور جستجو از دو فرآیند برخط^۱ و برون خط^۲ تشکیل شده است. فرآیند برون خط به صورت مداوم و دوره‌ای^۳، تصاویر و صفحات وب را دریافت و ذخیره می‌نماید. سپس نمایه‌ساز این اطلاعات را به نمایه‌های قابل جستجو تبدیل می‌نماید. فرآیند برخط، در هنگام درخواست کاربر اجرا می‌شود. هنگام پرس‌وجوی کاربر، بخش بازیابی تعدادی سند را بازیابی نموده و بخش رتبه‌بندی نتایج بازیابی شده را بر اساس میزان ارتباط با پرس‌وجوی کاربر مرتب نموده و به کاربر ارائه می‌دهد. شکل ۳-۱ این دو فرآیند را در یک موتور جستجو نشان می‌دهد.



شکل ۳-۱: فرآیندهای برخط و برون خط موتور جستجو

۲-۱ رتبه‌بندی تصاویر در وب

رتبه‌بندی یک تصویر به فرآیندی گفته می‌شود که به هر تصویر بازیابی شده با توجه به پرس‌وجوی کاربر، امتیازی نسبت بدهیم که بیان‌کننده میزان مرتبط بودن آن تصویر با پرس‌وجوی کاربر باشد. سپس تصاویر بازیابی شده را بر اساس این امتیاز، مرتب کنیم. با توجه به این که به ازای هر پرس‌وجوی کاربر، هزاران تصویر مرتبط وجود دارد، لازم است تا موتور جستجو نتایج را رتبه‌بندی

¹ Online

² Offline

³ Periodically

نموده و تنها چند نتیجه اول را به کاربر نمایش دهد. در حقیقت رتبه‌بندی، میزان مرتبط بودن هر تصویر نتیجه با پرس‌وجوی کاربر را تخمین خواهد زد.

الگوریتم‌های مختلفی برای رتبه‌بندی تصاویر وجود دارد که به طور کلی می‌توان آن‌ها را در چند دسته زیر تقسیم‌بندی نمود:

- الگوریتم‌های مبتنی بر محتوای متنی
- الگوریتم‌های مبتنی بر محتوای دیداری تصاویر
- الگوریتم‌های مبتنی بر اتصال
- الگوریتم‌های مبتنی بر رفتار کاربر

الگوریتم‌های مبتنی بر محتوای متنی، به هر تصویر درون یک صفحه، بر اساس میزان مرتبط بودن پرس‌وجوی کاربر با محتوای متنی آن صفحه و ویژگی‌های متنی خود تصویر، امتیازی نسبت می‌دهند. یعنی هرچه پرس‌وجوی کاربر با محتوای متنی یک صفحه که حاوی یک تصویر خاص است مرتبط‌تر باشد، آن تصویر امتیاز بیشتری کسب کرده و در رتبه‌بندی در اولویت بالاتری قرار خواهد گرفت [۷، ۸]. در روش‌های مبتنی بر محتوای متنی از مدل‌های بولی^۱، احتمالی^۲ و فضای برداری^۳ جهت رتبه‌بندی اسناد بر مبنای محتوای آن‌ها استفاده می‌شود. مهمترین روش در مدل برداری، الگوریتم TF-IDF [۹] و در مدل احتمالی BM25 [۱۰] می‌باشد.

الگوریتم‌های مبتنی بر محتوای دیداری تصاویر، بر اساس ویژگی‌های محتوای تصویر، میزان شباهت بین دو به دوی تصاویر بازیابی شده را می‌یابند. سپس گرافی تحت عنوان گراف شباهت تشکیل می‌دهند. حال به هر گره از گره‌های این گراف که همان تصاویر ما می‌باشند بر اساس تعداد یال‌های متصل به آن و وزن هر کدام، امتیازی نسبت داده می‌شود. حال بر اساس این امتیازها، رتبه‌بندی مجددی روی تصاویر بازیابی شده انجام می‌شود [۱۱، ۱۲].

^۱ Boolean

^۲ Probabilistic

^۳ Vector space

الگوریتم‌های مبتنی بر اتصال، امتیازدهی به هر تصویر را بر اساس اعتبار^۱ صفحاتی که تصویر درون آن قرار گرفته است، انجام می‌دهند. این اعتبار بر اساس یک رابطه به خود تصویر منتقل می‌شود و به عنوان امتیاز برای آن تلقی می‌شود [۱۳].

الگوریتم‌های مبتنی بر رفتار کاربر، الگوریتم‌هایی هستند که عمل یادگیری^۲ را انجام می‌دهند. این الگوریتم‌ها بر اساس بازخورد^۳هایی که از کاربر دریافت می‌کنند، امتیاز یک تصویر را تغییر می‌دهند. این نوع الگوریتم‌ها، معمولاً به صورت ترکیبی با الگوریتم‌های دیگر مورد استفاده قرار می‌گیرند [۱۴، ۱۵، ۱۶].

الگوریتم‌های ترکیبی، با ترکیب دو یا چند روش فوق، سعی در بهبود نتایج و سرعت جستجو، نموده‌اند [۱۷].

۳-۱ چالش‌های^۴ رتبه‌بندی تصاویر در وب

۱-۳-۱ مفهوم مرتبط بودن

وقتی کاربر پرس‌وجوی خود را اعلام می‌کند، سامانه^۵ باید مجموعه‌ای از تصاویر مرتبط با آن پرس‌وجو را به او ارائه کند. حال منظور از مرتبط بودن پرس‌وجو با یک تصویر چیست؟ در یک دید اولیه، می‌توان گفت اگر محتوای متنی پرس‌وجو با محتوای متن صفحه‌ی در برگیرنده تصویر، اشتراک بیشتری داشته باشد، آن تصویر، تصویر مناسب‌تری برای ارائه به کاربر می‌باشد. مزایای این روش سادگی و سرعت بالای بازیابی نتایج می‌باشد. ولی همواره چنین نیست. صفحاتی که به هر دلیل، از متن و تصویر نامرتب باهم، استفاده نموده‌اند، باعث می‌شوند نتایج نامرتب با پرس‌وجو به عنوان نتیجه جستجو به کاربر ارائه شود [۱۸]. از دیگر مشکلات الگوریتم‌های مبتنی بر محتوای متنی می‌توان به موجود نبودن این اطلاعات برای هر تصویر و دقت پایین نام برد.

¹ Authority

² Learning

³ Feedback

⁴ Challenges

⁵ System

پس الگوریتم‌های مبتنی بر محتوای متنی نمی‌توانند به تنهایی در جستجوی یک تصویر در وب، موفق عمل نمایند. در نتیجه می‌توان از ترکیب این الگوریتم با الگوریتم‌های دیگر بهره برد.

۱-۳-۲ توصیف‌گرهای تصویر^۱

در الگوریتم‌های مبتنی بر محتوای دیداری، به ویژگی‌هایی از تصویر نیاز داریم که بتواند به نوعی بازگوکننده محتوای دیداری تصویر باشد. چنین ویژگی‌هایی با نام توصیفگر تصویر، شناخته می‌شوند. توصیفگرهای تصویر، به طور دقیق‌تری، یک تصویر را توصیف می‌کنند ولی استخراج آن‌ها از تصویر با محدودیت در سرعت همراه است [۱۸]. زیرا در هر تصویر با میلیون‌ها پیکسل به عنوان داده مورد پردازش روبرو هستیم و الگوریتم‌های تعیین میزان شباهت بین تصاویر از درجه پیچیدگی بالایی برخوردارند. مشکل دیگر در استفاده از توصیفگرهای تصویر این است که استنتاج مفاهیم معنایی^۲ از چنین داده‌هایی، کار آسانی نیست. زیرا این توصیفگرها، تصویر را در سطح پایین^۳، توصیف می‌کنند. همچنین یافتن شباهت بین دو به دو تصاویر کاری بس پیچیده است. به همین منظور به راهکارهایی نیاز است تا این فرآیند، با سرعت بالاتری انجام گیرد.

۱-۳-۳ وابستگی به کاربر

وابستگی به بازخورد کاربر در الگوریتم‌هایی که بر اساس رفتار کاربر عمل می‌کنند، خود یک نوع چالش محسوب می‌شود. تجربه نشان داده است که کاربران علاقه‌ای به ارائه بازخورد مستقیم نخواهند داشت. به همین دلیل چنین الگوریتم‌هایی در موتورهای جستجوی تصویر، کاربرد کمی دارند.

^۱ Image Descriptor

^۲ Semantic Concepts

^۳ Low Level

۴-۳-۱ شباهت^۱ یا تطبیق^۲

در جستجوی تصاویر توسط الگوریتم رتبه‌بندی محتوای تصویر، می‌توان با انتخاب کاربر، تصاویر مشابه و یا تصاویر دقیقاً منطبق بر هم را به کاربر ارائه داد [۱۸]. یعنی پس از جستجو توسط کاربر، بتوانیم تصاویر مشابه و یا تصاویر منطبق با یک تصویر نتیجه را به کاربر ارائه کنیم.

۵-۳-۱ نمایه‌ها^۳

یکی از مسائل پیش روی موتورهای جستجو، حجم نمایه‌های حاصل از ویژگی‌های متنی صفحات، ویژگی‌های متنی اختصاص‌یافته^۴ به تصویر و توصیفگرهای تصاویر می‌باشد. بسته به الگوریتم رتبه‌بندی و راهکارهای آن، حجم نمایه‌ها می‌تواند متفاوت باشد. هرچه قابلیت‌های جستجو بالاتر باشد، حجم نمایه‌ها نیز زیادتر می‌شود. به عنوان نمونه اگر بخواهیم مکان واژه‌های پرس‌وجوی کاربر را در نتیجه نشان دهیم و یا این‌که بخواهیم جستجو را بر روی رنگ و اندازه تصاویر انجام دهیم، نیاز به حجم بالاتری برای ذخیره نمایه‌ها داریم. در نتیجه در طراحی الگوریتم رتبه‌بندی باید محدودیت‌های مربوط به فضای مورد نیاز برای نمایه‌ها را هم باید مدنظر قرار دهیم. در صورتی که حجم نمایه‌ها خیلی زیاد باشد، باید اسناد وب را به چندین نمایه در جاهای مختلف تقسیم نمود و عمل جستجو را به صورت موازی انجام داد. برای این منظور می‌بایست از روش‌های توزیع‌شده^۵ بهره برد [۱۸].

۶-۳-۱ نوع پرس‌وجو

نوع پرس‌وجوی کاربر می‌تواند متفاوت باشد. پرس‌وجو توسط متن، پرس‌وجو توسط یک تصویر به‌عنوان ورودی و یا ترکیبی از این دو حالت را به عنوان ورودی داشته‌باشیم. برای هر یک از این حالات، الگوریتم‌های متفاوتی برای رتبه‌بندی به‌کار می‌روند. در این حالت، کاربر پرس‌وجوی خود را

¹ Similarity

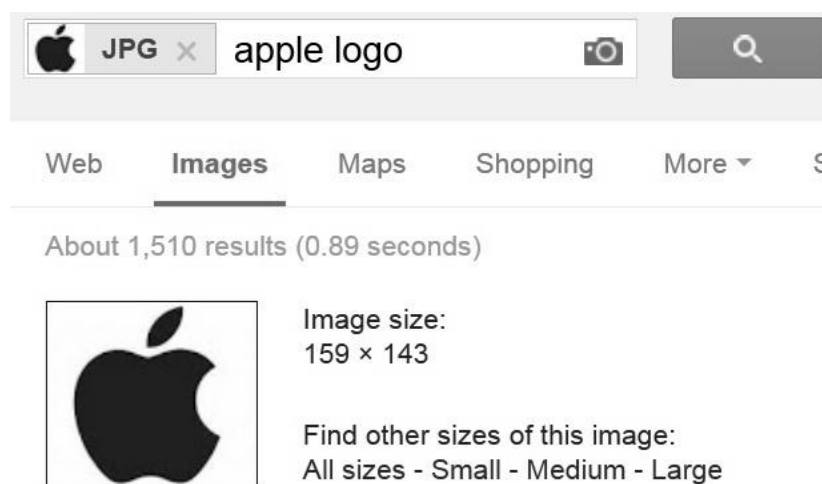
² Matching

³ Indexes

⁴ Associated Text

⁵ Distributed

توسط تعداد محدودی کلیدواژه، انجام می‌دهد و موتور جستجو، مرتبط‌ترین تصاویر را به کاربر ارائه می‌دهد. در حالتی که پرس‌وجو توسط یک تصویر انجام شود، موتور جستجو باید بر اساس ویژگی‌های محتوایی تصویر، مرتبط‌ترین تصاویر را به کاربر ارائه دهد [۱۸]. در این گزارش، پرس‌وجو توسط متن مد نظر می‌باشد.



Best guess for this image: [apple logo](#)

شکل ۴-۱: تصویر به عنوان پرس‌وجو در موتور جستجوی گوگل

۷-۳-۱ بازیابی تصاویر

در مرحله بازیابی تصاویر، یعنی مرحله قبل از رتبه‌بندی، ابتدا بر اساس نوع پرس‌وجو، باید تعدادی تصویر بازیابی شوند. در این مرحله، می‌توان بر اساس میزان فاصله پرس‌وجو از محتوای صفحه و تصویر، تعدادی تصویر به دو روش بازیابی نمود. در روش اول، تعداد K تا از نزدیک‌ترین^۱ تصاویر به پرس‌وجو انتخاب می‌شوند ولی در روش دوم، تمام تصاویری که فاصله‌ای کمتر از ϵ ، از پرس‌وجو دارند، انتخاب می‌گردند [۱۸].

^۱ K-Nearest Neighbor

وجود این مرحله ضروری است. زیرا تعداد تصاویر در مخزن تصاویر خیلی زیاد بوده و الگوریتم‌های رتبه‌بندی از سرعت خیلی کمتری برخوردارند. در نتیجه عبور از این فیلتر ضروری است.

۴-۱ ارزیابی^۱

در بحث ارزیابی یک الگوریتم یا سامانه، با دو مفهوم روبرو هستیم: یکی مجموعه داده محک^۲ و دیگری معیارهای ارزیابی^۳ است.

۱-۴-۱ مجموعه داده محک

برای امکان مقایسه دو روش، باید بستری یکسان فراهم شود که دو روش بر روی آن بستر عمل نموده و نتایج روی بستر یکسان باهم مقایسه گردند.

مجموعه محک برای رتبه‌بندی تصاویر شامل قسمت‌های زیر است:

- مجموعه‌ای از اسناد^۴ شامل صفحات وب و تصاویر داخل آن
- مجموعه‌ای از چندین پرس‌وجو
- مجموعه‌ای از قضاوت‌های انسانی^۵ در مورد زوج تصویر و پرس‌وجو

قضاوت می‌تواند به شکل دودویی^۶ (مرتبط یا غیر مرتبط) و یا به شکل یک عدد باشد که

میزان ارتباط بین تصویر و پرس‌وجو را نشان دهد.

¹ Evaluation

² Benchmark Dataset

³ Evaluation Metrics

⁴ Documents Collection

⁵ Human Judgement

⁶ Binary

۲-۴-۱ معیارهای ارزیابی

برای مقایسه کارایی دو روش متفاوت، باید معیارهایی را برای ارزیابی آن‌ها در نظر بگیریم. دو معیار سنتی در این زمینه، معیارهای دقت^۱ و فراخوانی^۲ است. اگر R مجموعه کل اسناد مرتبط با پرس‌وجو و A مجموعه اسناد بازیابی شده در پاسخ به پرس‌وجو و R_a مجموعه کل اسناد بازیابی شده مرتبط با پرس‌وجو باشد، در این صورت، دقت و فراخوانی به‌صورت زیر تعریف می‌شوند:

دقت

نسبت تعداد اسناد بازیابی شده مرتبط به کل اسناد بازیابی شده [۱۹].

$$Precision = \frac{|R_a|}{|A|} \quad (1-1)$$

فراخوانی

نسبت تعداد اسناد بازیابی شده مرتبط به کل اسناد مرتبط [۱۹].

$$Recall = \frac{|R_a|}{|R|} \quad (2-1)$$

به دلیل این که تعداد اسناد بازیابی شده توسط موتورهای جستجو خیلی زیاد است و محاسبه چنین معیارهایی را دشوار می‌سازد، اغلب از معیارهای دیگری چون $P@n$ ، AP و $NDCG@n$ ^۵ برای این منظور استفاده می‌شود [۱۹].

$$P@n = \frac{\text{Number of Relevant Documents in Top } n \text{ Results}}{n} \quad (3-1)$$

در معیار بالا، تنها مرتبط بودن یا غیر مرتبط بودن در نظر گرفته می‌شود. اگر میزان ارتباط بین تصویر و پرس‌وجو توسط یک عدد مشخص شود، نیازمند معیاری دیگر هستیم که بتوان میزان مرتبط بودن را در معیار اثر داد. اگر تعداد n تصویر بازیابی کنیم و r_i میزان ارتباط نتیجه i ام را با پرس‌وجو نشان دهد، معیار ارزیابی به‌صورت زیر تعریف می‌شود [۱۹]:

¹ Precision

² Recall

³ Precision at n ($P@n$)

⁴ Mean Average Precision (MAP)

⁵ Normalize Discounted Cumulative Gain at n ($NDCG@n$)

$$NDCG@n = \sum_{i=1}^n \frac{2^{r_i} - 1}{\log(1 + i)} \quad (4-1)$$

معیار دیگری به نام دقت متوسط^۱ وجود دارد که برای هر پرس و جو به صورت زیر محاسبه می شود [۱۹]:

$$AP = \frac{\sum_{i=1}^N (P@i * rel(i))}{\text{Number of Total Relevant Documents}} \quad (5-1)$$

در این رابطه N نشان دهنده تعداد اسناد بازیابی شده است و $rel(i)$ به ازای سند i ام مرتبط، ۱ و به ازای سند نامرتبط، ۰ می باشد.

معیار دیگری به نام میانگین دقت متوسط^۲ وجود دارد که میانگین معیار بالا برای تمام پرس و جوها می باشد.

۵-۱ ساختار پایان نامه

این پایان نامه مشتمل بر چهار فصل می باشد. در فصل دوم، روش های رتبه بندی پیشین بر روی تصویر را مرور می کنیم. در فصل سوم روش پیشنهادی برای رتبه بندی تصاویر را معرفی خواهیم نمود و در فصل چهارم به نتیجه گیری و کارهای آینده می پردازیم.

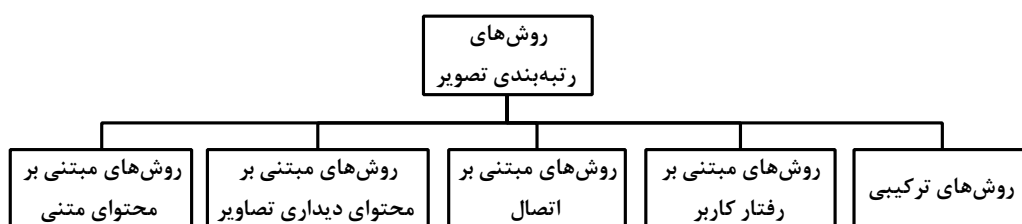
¹ Average Precision

² MAP

۲ مروری بر کارهای گذشته

روش‌های رتبه‌بندی را می‌توان به طور کلی در دسته‌های زیر تقسیم نمود:

- روش‌های مبتنی بر محتوای متنی
- روش‌های مبتنی بر محتوای دیداری تصاویر
- روش‌های مبتنی بر اتصال
- روش‌های مبتنی بر رفتار کاربر
- روش‌های ترکیبی



شکل ۲-۱: دسته‌بندی روش‌های رتبه‌بندی

۲-۱ روش‌های مبتنی بر محتوای متنی

روش‌های مبتنی بر محتوای متنی اسناد، روش‌هایی هستند که تصاویر را بر اساس میزان انطباق پرس‌وجو با محتوای متنی صفحه‌ای که تصویر در آن قرار دارد و متن‌های اختصاص‌یافته به تصویر، رتبه‌بندی می‌کنند. این روش‌ها از جمله روش‌های اولیه برای بازیابی تصاویر وب بوده‌اند و از دقت بسیار پایینی برخوردارند. به همین دلیل معمولاً به‌صورت ترکیبی با روش‌های دیگر مورد استفاده قرار می‌گیرند.

ساختار کلی این روش‌ها، شباهت زیادی به روش‌های مبتنی بر محتوا در بازیابی صفحات وب دارد. برای رتبه‌بندی تصاویر بازیابی شده، تنها میزان مرتبط بودن پرس‌وجو با محتوای متنی صفحات

حاوی تصاویر و متن‌های اختصاص یافته به تصاویر، به عنوان امتیاز برای رتبه‌بندی آن تصویر در نظر گرفته می‌شود. مدل فضای برداری^۱ [۲۰]، روش‌های احتمالی^۲ [۲۱] و روش‌های مبتنی بر مدل‌سازی زبانی^۳ [۵]، از جمله مدل‌هایی هستند که می‌توان توسط آن‌ها میزان ارتباط بین پرس‌وجو با محتوای صفحه حاوی تصویر و متن‌های اختصاص یافته به تصویر را به دست آورد.

۱-۱-۲ مدل فضای برداری

مدل فضای برداری، ارائه شده توسط آقای سالتون^۴ [۹]، یکی از قدیمی‌ترین روش‌های مبتنی بر محتوا در زمینه بازیابی اطلاعات سنتی محسوب می‌شود. نمایش اسناد و پرس‌وجوها به صورت برداری در فضای برداری مشترک، بنام مدل فضای برداری شناخته می‌شود. در این مدل سند و پرس‌وجو در قالب بردارهایی به ابعاد تعداد لغات موجود در واژگان^۵ نمایش داده شده و سپس با استفاده از یک تابع وزن‌دهی^۶، به هر مؤلفه بردار وزنی نسبت داده می‌شود [۵]. سپس با محاسبه کسینوس زاویه بین دو بردار وزن‌دار، درجه شباهت آن‌ها به صورت زیر محاسبه می‌شود. در این رابطه $w_{t,d}$ و $w_{t,q}$ به ترتیب بیانگر وزن کلمه t در سند d و پرس‌وجوی q هستند.

$$Score(q, d) = \frac{\vec{W}_q \cdot \vec{W}_d}{\|\vec{W}_q\| * \|\vec{W}_d\|} = \frac{\sum_t w_{t,q} * w_{t,d}}{\sqrt{\sum_t w_{t,q}^2} * \sqrt{\sum_t w_{t,d}^2}} \quad (1-2)$$

$\vec{W}_q \cdot \vec{W}_d$ ضرب داخلی بردارهای \vec{W}_q و \vec{W}_d بوده و $\|\vec{W}_q\|$ و $\|\vec{W}_d\|$ به ترتیب نشان‌دهنده طول اقلیدسی^۷ بردارهای مربوطه هستند.

این نمایش از پرس‌وجو و سند، ترتیب کلمات را لحاظ نمی‌کند. به بیان دیگر سند و پرس‌وجو بعنوان کیفی از کلمات^۸ در نظر گرفته می‌شوند [۲۲] که در آن ترتیب دقیق کلمات نادیده گرفته

¹ Vector Space Model(VSM)

² Probabilistic

³ Language Modeling (LM)

⁴ Salton

⁵ Lexicon

⁶ Weighting function

⁷ Euclidean length

⁸ Bag of words

شده و تعداد رخدادهای کلمات مد نظر است. بعنوان مثال در این مدل، متن "محمد باهوش تر از علی است" با متن "علی باهوش تر از محمد است" یکسان می‌باشد.

دلیل استفاده از شباهت کسینوسی^۱ نیز غلبه بر تأثیر طول سند (تعداد کلمات در سند) است. زیرا برای مثال اگر اندازه بردارهای سند و پرس‌وجو را مبنای محاسبه امتیاز در نظر می‌گرفتیم، دو سند بسیار مشابه، صرفاً به این دلیل که طول یکی بزرگ‌تر از دیگری است، امتیازات بسیار متفاوتی کسب می‌کردند.

تأثیر تقسیم بردار بر طول اقلیدسی آن، نرمال‌سازی^۲ بردارها به بردار واحد^۳ است. به بیان دیگر اگر $\frac{\vec{w}_q}{\|\vec{w}_q\|}$ را با \vec{w}_q و $\frac{\vec{w}_d}{\|\vec{w}_d\|}$ را با \vec{w}_d نمایش دهیم، رابطه (۱-۲) را می‌توان به صورت زیر نیز نمایش داد که در آن \vec{w}_q و \vec{w}_d بردارهای واحد هستند:

$$Score(q, d) = \vec{w}_q \cdot \vec{w}_d \quad (2-2)$$

نرمال‌سازی باعث می‌شود امتیاز اسناد در یک مقیاس^۴ بوده و قابل مقایسه باشد.

برای وزن‌دهی کلمات داخل متن، معمولاً از خواص آماری متن و پرس‌وجو استفاده می‌شود. مرسوم‌ترین روش وزن‌دهی، روش tf-idf آقای سالتون [۲۳] است که از تکرار کلمات سند و پرس‌وجو استفاده می‌کند. در این روش هدف از tf در نظر گرفتن این مهم است که در صورتی که یک کلمه در یک سند بیشتر ظاهر شود، آن کلمه، سند را بهتر توصیف می‌کند. از آنجاکه صرف تعداد بیشتر رخدادهای یک کلمه در یک سند نمی‌تواند دلیل بر برتری آن باشد، معمولاً تابعی از تواتر کلمه^۵ به عنوان tf آن در نظر گرفته می‌شود.

پارامتر دیگر idf است که نشان‌دهنده عکس تعداد اسناد^۶ شامل کلمه در کل مجموعه است. لازم به تأکید است که در محاسبه idf، تعداد اسنادی که کلمه مورد نظر در آن ظاهر شده (df)^۷ مدنظر است و نه تعداد دفعات کلی تکرار کلمه در مجموعه اسناد (cf)^۸. استفاده از df به این دلیل

^۱ Cosine similarity

^۲ Normalization

^۳ Unit vector

^۴ Scale

^۵ Term Frequency (TF)

^۶ Inverse Document Frequency (IDF)

^۷ Document Frequency (DF)

^۸ Collection Frequency (CF)

مناسب‌تر است که کلمه‌ای که در تعداد اسناد کمتری ظاهر می‌شود، نسبت به کلمه‌ای که در بیشتر اسناد وجود دارد (مثل حروف اضافه)، دارای اطلاعات بیشتری است. در مقابل ممکن است کلمه‌ای تنها در چندین سند ظاهر شده باشد و تعداد تکرارش در این اسناد خیلی زیاد باشد، اما این مسئله دلیل بر کم اهمیت بودن کلمه در کل مجموعه نباشد. جالب‌تر این است که ممکن است df و cf در یک مجموعه برای کلمات مختلف رفتاری متفاوت داشته باشند.

برای محاسبه idf نیز معمولاً تابعی از df استفاده می‌شود که رایج‌ترین آن $\log N/df_t$ است که در آن N نشان دهنده تعداد کل اسناد مجموعه و df_t تعداد اسناد شامل کلمه t در کل مجموعه است [۲۴].

در انتها وزن هر کلمه در بردار، از حاصلضرب tf_t در idf_t محاسبه می‌شود. از ترکیب‌های مختلف توابع مورد استفاده برای محاسبه tf و idf ، شماهای مختلفی بدست می‌آید.

۲-۱-۲ مدل احتمالی

مدل‌های احتمالی نیز از دیگر روش‌های قدیمی رتبه‌بندی مبتنی بر محتوا هستند که در ارزیابی‌ها نتایج چشمگیری را نشان داده‌اند [۲۵]. در این روش‌ها، رتبه‌بندی اسناد بر مبنای احتمال ارتباط سند و پرس‌وجوست. با فرض نمایش دودویی ارتباط^۱ (مرتبط یا نامرتبط)، برای پرس‌وجوی q و سند d ، $R_{q,d}$ بیانگر یک متغیر تصادفی^۲ است که نشان می‌دهد آیا سند با پرس‌وجو مرتبط است یا خیر [۲۲]. با استفاده از مدل احتمالی، واضح است که می‌توان اسناد را براساس احتمال ارتباط آن‌ها با پرس‌وجو رتبه‌بندی نموده و به کاربر نمایش داد. این مهم، مبنای اصل رتبه‌بندی احتمالی^۳ است:

اگر در یک سیستم بازیابی اطلاعات، اسناد را براساس ترتیب نزولی احتمال ارتباط آن‌ها به پرس‌وجو به کاربر ارائه دهیم، به گونه‌ای که احتمالات با حداکثر دقت ممکن براساس داده‌های

^۱ Relevancy

^۲ Random variable

^۳ Probabilistic Ranking Principle (PRP)

موجود در سیستم بدست آمده باشند، کارایی کلی سیستم برای کاربرانش، حداکثر کارایی قابل دسترسی بر مبنای آن داده‌هاست.

با وجود این که این اصل نیاز به فراهم بودن تمامی احتمالات به صورت صحیح دارد، اما همچنان پایه‌ای بسیار مفید برای توسعه دیگر مدل‌ها در بازیابی اطلاعات است.

آقای رابرتسون^۱ در [۲۶، ۲۷] الگوریتم BM25 را ارائه نموده که با ظهور خود انقلابی چشمگیر در مدل‌های احتمالی بازیابی اطلاعات پدید آورد. الگوریتم BM25 بر خلاف روش‌های احتمالی پیش از خود مانند مدل دودویی وابستگی^۲، مفروضات غیر واقعی مانند نمایش اسناد و پرس‌وجو در قالب بردارهای دودویی از کلمات را، کنار گذاشته [۲۲] و یک مکانیزم وزن‌گذاری مبتنی بر "۲-پواسن" به نام اکاپی^۳ که دارای ویرایش‌های متنوعی به صورت BMnn هست را ارائه نموده است. نام‌گذاری این مکانیزم (۲-پواسن) به این خاطر است که توزیع هر واژه در یک مجموعه اسناد دارای توزیع پواسن است. این روش احتمالی از بهترین روش‌های رتبه‌بندی مبتنی بر محتوا به شمار رفته و طبق آزمایش‌های انجام‌شده، دارای دقتی حدود ۲۵٪ است [۲۸]. لازم به ذکر است که الگوریتم BM25 همانند الگوریتم tf-idf به سند و پرس‌وجو در قالب کیفی از کلمات نظر داشته و روابط داخلی بین کلمات را در نظر نمی‌گیرد [۲۲]. متغیرهای مورد نیاز جهت محاسبه امتیاز BM25 در جدول ۱-۲ تعریف شده‌اند.

فرمول BM25 بدین صورت است:

$$S(Q, D) = \sum_{t \in Q} \left(\log \frac{(r + 0.5)/(R - r + 0.5)}{(n - r + 0.5)/(N - n - R + r + 0.5)} \right) \frac{(k_1 + 1)tf}{K + tf} \frac{(k_3 + 1)q}{k_3 + qtf} \quad (3-2)$$

$$+ k_2 |Q| \frac{avdl - dl}{avdl + dl}$$

همان طور که ملاحظه می‌شود مقدار عددی حاصل، برابر جمع وزن دار همه کلمات پرس‌وجو در سند متناظر است. لازم است پارامترهای ثابت در فرمول فوق، برای رسیدن به دقت بالا به صورت مناسب مقداردهی شوند. آزمایش‌ها نشان داده‌اند که بسته به مجموعه داده مورد استفاده، بعضی از

¹ Robertson

² Binary Independence Model (BIM)

³ Okapi

پارامترهای فوق متفاوت خواهند بود. طبق نظر آقای رابرتسون k_1 و b به صورت پیش فرض $1/2$ و $0/75$ مقداردهی می شوند ولیکن برای پرسش های با طول بزرگ k_3 اغلب مساوی 7 مقداردهی می شود و k_2 اغلب صفر است [۲۹].

جدول ۲-۱: متغیرهای استفاده شده در روش BM25 [۵].

نام متغیر	تعریف
r	تعداد اسناد مرتبط به پرس و جوی Q شامل واژه t
R	تعداد اسناد مرتبط به پرس و جوی Q
n	تعداد اسناد شامل واژه t
N	تعداد کل اسناد
tf	فرکانس واژه t در سند D
qtf	فرکانس واژه t در پرس و جو
avdl	میانگین طول اسناد
dl	طول سند D (تعداد واژه های سند)
$ Q $ یا nq	تعداد واژه های پرس و جو
k_1, k_2, k_3 و b $K = k_1((1-b) + b(\frac{dl}{dl_{ave}}))$	ثابت های قابل تنظیم (جهت کوک کردن) ^۱
$S(Q,D)$	شباهت میان پرس و جوی Q با سند D

۲-۲ روش های مبتنی بر محتوای دیداری تصاویر

در روش های رتبه بندی مبتنی بر محتوای دیداری تصاویر که جستجو بر اساس پرس و جوی متنی کاربر انجام می گیرد، ابتدا تعدادی تصویر بر اساس مرتبط بودن با پرس و جوی کاربر، بازیابی می شوند. در هر روش از یک یا چند ویژگی از محتوای تصویر، برای مقایسه دو به دوی تصاویر، استفاده می شود و بر همین مبنا، گرافی تحت عنوان گراف شباهت^۲، شکل می گیرد. حال گره هایی که بیشترین تعداد یال به آن ها متصل است، گره هایی در مرکز تراکم گراف می باشند که تعداد

¹ Tuning

² Similarity Graph

تصاویر مشابه با آن‌ها بیشتر بوده است. در نتیجه می‌توان از این عدد به عنوان معیاری برای نشان دادن میزان شباهت بهره برد.

ویژگی‌های محتوای دیداری تصویر برای تعیین میزان شباهت بین دو تصویر استفاده می‌شود. این ویژگی‌ها خود به دو دسته تقسیم می‌شوند [۱۱]:

- ویژگی‌های سراسری^۱
- ویژگی‌های محلی^۲

ویژگی‌های سراسری

هنگامی که این نوع ویژگی‌ها برای تعیین میزان شباهت بین دو تصویر استفاده شود، تصاویری که متشابه تعیین می‌شوند از تنوع بالایی برخوردار می‌باشند. به عبارت دیگر به هنگام جستجوی یک تصویر خاص، تصاویر با دورنمای سه‌بعدی^۳ متفاوت، زاویه دوربین متفاوت، فاصله کانونی^۴ متفاوت و یا ترکیبی از این‌ها به عنوان نتیجه به کاربر ارائه شود.

این نوع ویژگی‌ها اغلب در برابر بیشتر تبدیل‌ها^۵، ناپایدار^۶ هستند و نمی‌توانند دو تصویر که یکی تبدیلی از دیگری است را تشخیص دهند.

ویژگی‌های زیر همه از این نوع محسوب می‌شوند:

- بافت‌نما^۷: این ویژگی تعداد تکرار هر شدت روشنایی را در تصویر مشخص می‌کند.
- شکل^۸: مشخص کننده ناحیه یک شی داخل تصویر است.
- بافت^۹: این ویژگی بیان کننده ساختاری در تصویر و یا ناحیه‌ای از آن است که در آن ناحیه تکرار شده است.

^۱ Global Features

^۲ Local Features

^۳ Perspectives

^۴ Focal Lengths

^۵ Transformations

^۶ Unstable

^۷ Histogram

^۸ Shape

^۹ Texture

ویژگی‌های محلی

در برابر ویژگی‌های سراسری، ویژگی‌هایی وجود دارد که اطلاعات غنی‌تری داشته و در برابر تبدیل‌های مختلف و چندین درجه تغییر شدت نور، پایدارند. از جمله ویژگی‌های محلی تصویر می‌توان موارد زیر را نام برد:

- تشخیص گوشه هریس^۱ [۳۰]

- تبدیل ویژگی مقیاس نابسته^۲ [۳۱]

- زمینه شکل^۳ [۳۲]

- تصاویر اسپین^۴ [۳۳]

یکی از معروف‌ترین روش‌های رتبه‌بندی مبتنی بر محتوای دیداری تصویر، الگوریتم VisualRank می‌باشد [۳۴].

۱-۲-۲ الگوریتم VisualRank

رویکرد این الگوریتم بر اساس شباهت بین تصاویر می‌باشد. این روش برای رتبه‌بندی تصاویر، از مفهوم ابرپیوند دیداری^۵ بهره می‌برد. مفهوم ابرپیوند دیداری به این معناست که اگر کاربری در حال مشاهده یک تصویر است، ممکن است علاقه‌مند به مشاهده تصاویر مشابه دیگر نیز باشد.

VisualRank از طریق فرمول تکراری زیر محاسبه می‌شود:

$$VR_{n \times 1} = d \times (S_{n \times n}^* \times VR_{n \times 1}) + (1 - d) \times P_{n \times 1} \quad (1-2)$$

و مقدار بردار P به‌صورت زیر است:

¹ Harris Corner Detection

² Scale Invariant Feature Transform(SIFT)

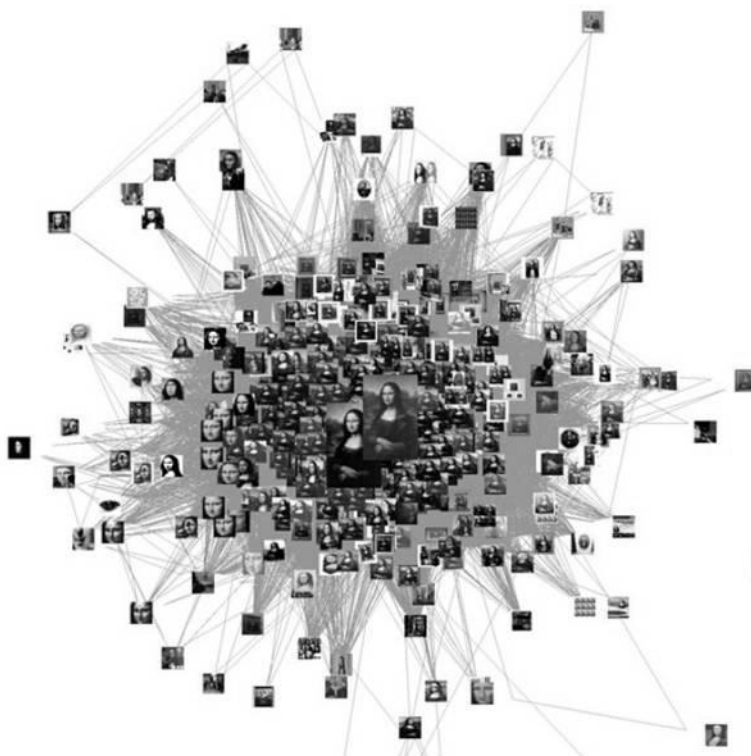
³ Shape Context

⁴ Spin Images

⁵ Visual Hyperlink

$$P = \begin{bmatrix} 1 \\ n \end{bmatrix}_{n \times 1} \quad (2-2)$$

که در آن n تعداد تصاویر بازیابی شده می‌باشد و S^* ماتریس هم‌جواری است که به صورت ستونی نرمال شده است به طوری که $S_{u,v}^*$ مقدار شباهت ظاهری دو تصویر u و v می‌باشد. از d به عنوان ضریب استهلاک^۱ برای همبندسازی گراف شباهت استفاده شده است. شکل ۲-۲ نمونه‌ای از گراف شباهت ساخته شده از ۱۰۰۰ نتیجه اول برای واژه "مونالیزا"^۲ می‌باشد [۳۴].



شکل ۲-۲: گراف شباهت در الگوریتم VisualRank [۳۴]

با استفاده از الگوریتم SIFT^۳، شباهت دو به دو تصاویر را می‌یابیم. این شباهت، همان تعداد نقاط ویژگی^۴ مشترک بین دو تصویر، تقسیم بر میانگین نقاط مورد علاقه^۵ می‌باشد. بیشتر تصاویر درون صفحات وب توسط متن‌ها، ابرپیوندها^۶ و فراداده‌های^۷ متعدد، احاطه شده است که ویژگی مناسبی برای بازیابی تصاویر در وب می‌باشد. ابتدا بر اساس پرس‌وجوی کاربر،

^۱ Damping Factor

^۲ Mona Lisa

^۳ Scale Invariant Feature Transform

^۴ Feature Point

^۵ Interest Point

^۶ Hyperlink

^۷ Metadata

تعدادی تصویر بازیابی می‌شود. سپس بر اساس الگوریتم VisualRank، به هر تصویر بازیابی شده امتیازی نسبت داده می‌شود. بالاتر بودن امتیاز یک تصویر نسبت به تصویر دیگر بدان معناست که تصویر با امتیاز بیشتر، با تصاویر بیشتری شباهت داشته است.

۳-۲ روش‌های مبتنی بر اتصال

این روش‌ها بر اساس الگوریتم‌های مبتنی بر اتصال در رتبه‌بندی صفحات وب، کار می‌کند. الگوریتم‌های PageRank [۳۵]، HITS [۳۶]، HostRank [۳۷] و DistanceRank [۳۸] از جمله این الگوریتم‌ها می‌باشند. تمام این الگوریتم‌ها با استفاده از گراف وب^۲، رتبه‌بندی هر صفحه را انجام می‌دهند. گراف وب، گرافی جهت‌دار است که گره‌های آن همان صفحات وب است و یال جهت‌دار از گره اول به گره دوم، نشان‌دهنده پیوند از صفحه اول به صفحه دوم می‌باشد. بر اساس اهمیت صفحه اول، یک وزن به یال جهت‌دار اختصاص می‌یابد. هدف تمامی این الگوریتم‌ها، تعیین میزان اعتبار یک صفحه با استفاده از تعداد یال‌های ورودی به گره صفحه، می‌باشد. به عبارت دیگر، کیفیت هر صفحه را از منظر صفحات بیرونی مورد محاسبه قرار داده و از محتوای صفحات دیگر برای ارزیابی کیفیت یک صفحه استفاده می‌کند. اعتبار یک صفحه را می‌توان غنی بودن آن صفحه از نظر بار اطلاعاتی تلقی نمود.

عیب این روش‌ها، پدیده‌ای به نام "غنی‌تر شدن اغنیاء"^۳ می‌باشد. این پدیده به این معنا می‌باشد که تصاویری که از میزان اعتبار بالاتری برخوردارند، در رتبه‌های اول قرار می‌گیرند. بیشتر کاربران موتور جستجو، تنها به ۱۰ تا ۲۰ نتیجه اول اکتفا می‌کنند. همین امر باعث می‌شود اعتبار تصاویر با میزان اعتبار بالا، به طور مدام افزایش یابد. زیرا تعداد پیوندهای ورودی به یک صفحه میزان محبوبیت آن صفحه را نشان می‌دهد و با میزان اعتبار آن صفحه رابطه‌ای ندارد. این امر باعث

^۱ Hyperlink-Induced Topic Search (HITS)

^۲ Web Graph

^۳ Rich-get-richer

می‌شود تا صفحات تازه متولد شده باکیفیت^۱ در نتایج اول ظاهر نشده و یا پس از مدت زمانی طولانی در نتایج اول قرار بگیرد.

روش‌های مبتنی بر اتصال در رتبه‌بندی تصویر، همگی از این فرض پیروی می‌کنند که صفحه‌ای که از اعتبار بیشتری برخوردار است، بار اطلاعاتی غنی‌تری دارد و به تبع آن، تصاویر داخل آن، تصاویر بااهمیت‌تری می‌باشند.

یکی از مهم‌ترین روش‌های مبتنی بر اتصال در رتبه‌بندی تصویر، روش PicASHOW^۲ [۱۳] می‌باشد.

۲-۳-۱ الگوریتم PageRank

PageRank یک الگوریتم مستقل از پرس وجو است که در موتور جستجوی گوگل استفاده شده است و بر اساس اتصال بین صفحات عمل می‌کند. برای مثال اگر صفحه p_1 به صفحه p_2 اشاره کند، موضوع p_2 برای ایجاد کننده p_1 جذاب می‌باشد. بنابراین تعداد پیوندهای ورودی به یک صفحه درجه جذابیت آن صفحه برای دیگران را نشان می‌دهد. در نتیجه درجه جذابیت یک صفحه با تعداد پیوندهای ورودی آن افزایش می‌یابد. به علاوه وقتی به یک صفحه از صفحات مهم (با تعداد پیوند زیاد) اشاره شود، آن صفحه نیز رتبه بالایی خواهد داشت. به عبارت دیگر وزن هر صفحه در PageRank جمع وزن‌دار صفحاتی است که به آن اشاره می‌کنند. بنابراین الگوریتم PageRank بازگشتی بوده و می‌توان آن را با استفاده از زنجیره مارکف مدل کرد.

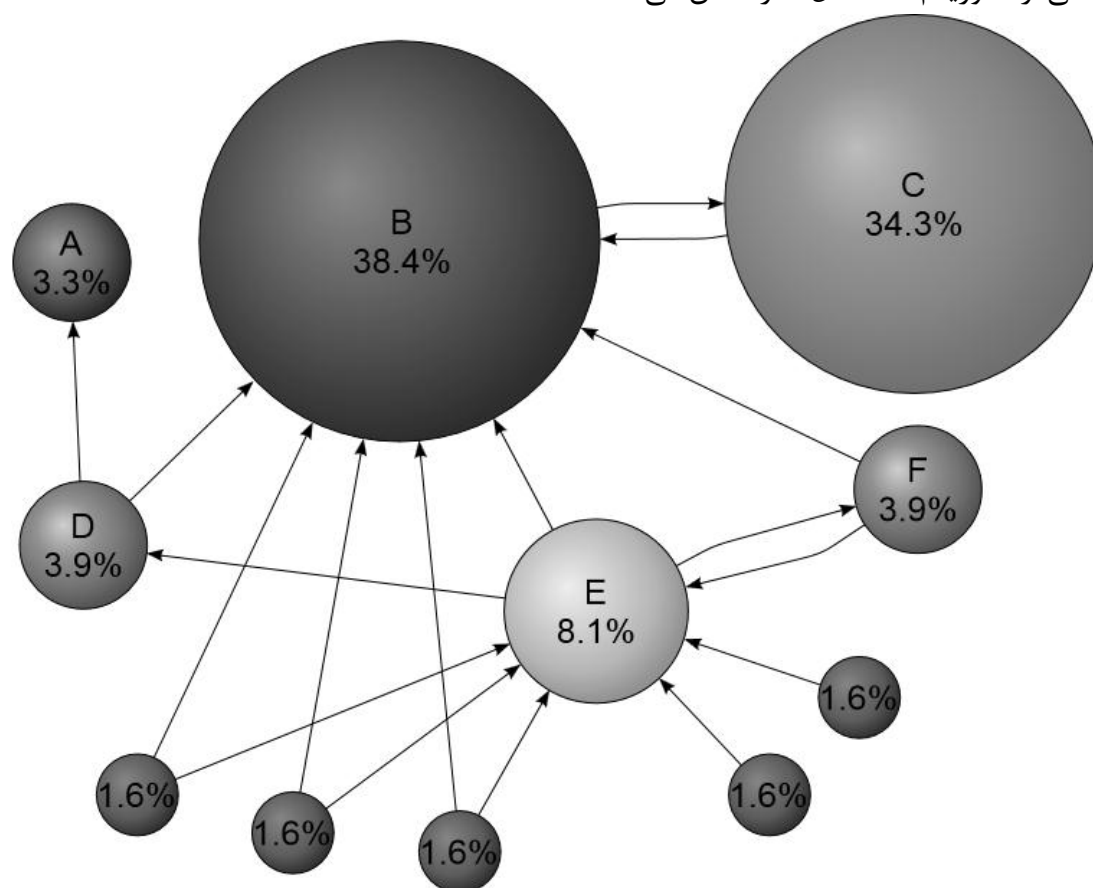
فرض کنید $N(i)$ و $B(i)$ به ترتیب نشان‌دهنده تعداد پیوندهای خروجی و مجموعه صفحات ورودی صفحه i باشند. رتبه صفحه i با استفاده از PageRank به صورت زیر محاسبه می‌شود:

$$r(i) = \sum_{j \in B(i)} r(j)/N(j) \quad (۴-۲)$$

^۱ High quality

^۲ Pictorial Authority Search by Hyperlinks On the Web

در نتیجه رتبه صفحه i مساوی جمع رتبه صفحات ورودی تقسیم بر درجه خروجی آن‌ها می‌باشد. تقسیم بر درجه خروجی باعث می‌شود تا اولاً رتبه صفحه به صورت عادلانه بین بچه‌هایش (خروجی‌ها) تقسیم شود و ثانیاً جمع رتبه همه صفحات به عدد ثابت (یک) نرمال شود. شکل ۳-۲ مثالی از الگوریتم PageRank را نشان می‌دهد.



شکل ۳-۲: مثالی از الگوریتم PageRank

رابطه (۴-۲)، برای حالتی که گراف همبند باشد مناسب است (هر گره به تمام گره‌ها دسترسی داشته باشد). در صورتی که گراف وب پیوسته نبوده یا صفحات بدون ورودی یا خروجی موجود باشند، الگوریتم موجب ایجاد اشکال می‌شود (الگوریتم همگرا نخواهد شد). به عبارت دیگر بعد از اجرای کامل الگوریتم تعداد زیادی از صفحات دارای مقدار PageRank صفر خواهند بود. برای حل این مشکل از پارامتر d به نام ضریب استهلاک به صورت زیر استفاده شده است که n نشانگر تعداد کل صفحات می‌باشد.

$$r(i) = \frac{1-d}{n} + d \sum_{j \in B(i)} r(j)/N(j) \quad (5-2)$$

بنابراین هر صفحه به تمام صفحات با احتمال $\frac{1-d}{n}$ یک پیوند خروجی خواهد داشت. مکانیزم فوق معادل یک موج سوار تصادفی که در وب قدم می‌زند و به صورت تصادفی روی پیوندها کلیک می‌کند، می‌باشد. زمانی که به یک صفحه با درجه خروجی صفر یا به حلقه بسته می‌رسد به یک صفحه دیگر پرش خواهد کرد. لذا وقتی کاربر در یک صفحه است با احتمال d یکی از پیوندهای آن را به صورت تصادفی انتخاب، یا با احتمال $1-d$ به صفحات دیگر پرش می‌کند.

معادله PageRank را می‌توان به صورت یک رابطه جبر خطی $r = A^T \times r$ نوشت که r یک بردار n بعدی است و هر عضو i آن نشان‌دهنده رتبه صفحه i می‌باشد. همچنین هر عضو ماتریس A به صورت $a_{ij} = \frac{1}{N(i)}$ است اگر صفحه i به صفحه j اشاره کند و در غیر این صورت $a_{ij} = 0$ خواهد بود. به عبارت دیگر هدف محاسبه بردار ویژه r از ماتریس A^T با مقدار ویژه ۱ است.

با توجه به وجود موضوعات مختلف روی وب و همپوشانی موضوعی آن‌ها با یکدیگر، در بیشتر مواقع باعث می‌شود موضوعات مختلف با یکدیگر رقابت کنند و در نتیجه دقت سیستم پایین بیاید. از نظر تئوری، همگرایی روش توان برای یک ماتریس مفروض، به شکاف مقدار ویژه بستگی دارد. منظور از شکاف مقدار ویژه، تفاضل بین مائولهای دو تا از بزرگترین مقادیر ویژه ماتریس است. به ادعای نویسندگان مقاله، این ویژگی در ماتریس A^T وجود داشته و همگرایی روش توان نسبتاً سریع است (حدود ۱۰۰ تکرار). با این وجود از آنجا که تنها ترتیب نسبی صفحات در رتبه‌بندی مدنظر است، می‌توان تکرار را تا زمانی ادامه داد که ترتیب نسبی صفحات به پایداری برسد و نیازی به ادامه تکرار تا رسیدن به مقادیر واقعی PageRank نیست. آزمایشات نشان داده در این حالت همگرایی بسیار سریع‌تر خواهد بود.

۲-۳-۲ الگوریتم HITS

الگوریتم HITS [۳۶] توسط آقای کلینبرگ برای رتبه بندی صفحات مرتبط با پرس وجو ارائه شد. این الگوریتم از جنبه‌های مختلفی با PageRank متفاوت است: وابسته به پرس وجوست،

بر روی زیر مجموعه‌ای از گراف وب عمل نموده و برای هر صفحه دو امتیاز hub و authority محاسبه می‌کند.

منظور از صفحات authority صفحات مهمی است که احتمال ارتباط آن‌ها با پرس وجو بسیار زیاد باشد. صفحات hub صفحاتی هستند که لزوماً authority محسوب نشده ولی به چندین صفحه authority اشاره می‌کنند. برای مثال صفحه اصلی وب سایت "link.ir" به احتمال زیاد یک hub خوب برای پرس وجوی "دولت" محسوب می‌شود، زیرا به چندین authority خوب در این زمینه اشاره دارد.

به نظر می‌رسد هیچ ویژگی ذاتی در متن یک صفحه وجود نداشته باشد که authority بودن آن را اثبات کند. در ضمن ممکن است کلمات پرس وجو در متن صفحات authority ظاهر نشوند. برای نمونه در جستجوی "search engine"، موتورهای جستجوی مشهوری مانند گوگل و یاهو، از این متن در صفحه اصلی خود استفاده نمی‌کنند.

به منظور غلبه بر این مشکلات می‌توان از ساختار پیوند بهره گرفت. قاعده‌تاً پیوند از صفحه p به q به معنای تأیید ضمنی صفحه q از جانب p است. البته استفاده صرف از ساختار پیوند، می‌تواند مشکلاتی نیز به همراه داشته باشد. برای مثال بسیاری از پیوندها صرفاً جهت اهدافی دیگر مانند راهبری، تبلیغات و غیره بوده و نمی‌توان حضور آن‌ها را دلیلی بر authority بودن صفحه مقصد دانست.

بین صفحات hub و authority رابطه تقویت دوجانبه برقرار است. hub های خوب به تعداد زیادی authority خوب اشاره نموده و authority های خوب از تعداد زیادی hub خوب پیوند دریافت می‌کنند.

الگوریتم HITS بر روی زیرگراف کوچکی از وب عمل نموده و با استفاده از تحلیل پیوند بر روی این گراف، hub ها و authority ها را تعیین می‌کند. این زیر گراف از مجموعه صفحات مرتبط با پرس وجو بدست آمده و می‌بایست تا حد امکان کوچک باشد. زیرا هم باعث تمرکز هرچه بیشتر

تحلیل پیوند بر روی بخش مرتبط گراف وب شده و هم میزان عملیات لازم در ادامه کار را کاهش می‌دهد. از آنجا که الگوریتم HITS بصورت برخط اجرا می‌شود، لازم است سریع به جواب برسد.

ایجاد زیرگراف:

برای ایجاد زیرگراف ابتدا با استفاده از روش‌های مبتنی بر محتوا، مجموعه‌ای از صفحات را بعنوان مجموعه ریشه که آن را با R نشان می‌دهیم، انتخاب نموده و سپس با افزودن فرزندان (صفحات اشاره شده) و والدین (صفحات اشاره کننده) این صفحات، مجموعه ریشه را به گراف بزرگتری به نام گراف پایه که آن را با S نمایش می‌دهیم، بسط داده می‌شود.

آقای کلینبرگ در مقاله خود، یک سری روش‌های اکتشافی^۱ برای کاهش تعداد پیوندهای بی‌ارزش در گراف پایه ذکر نموده است:

- تقسیم پیوندها به دو دسته پیوندهای درون سایتی و برون سایتی. منظور از پیوندهای درون سایتی، پیوندهای مابین صفحات داخلی یک دامنه است. از آنجا که این پیوندها معمولاً برای راهبری بین صفحات سایت استفاده می‌شوند، از ارزش زیادی برخوردار نبوده و می‌توان از آن‌ها صرف‌نظر کرد.

- اجازه دهیم حداکثر m (معمولاً ۴ تا ۸) صفحه از یک دامنه، به یک صفحه خاص اشاره کنند و پیوندهای اضافی را حذف کنیم. این عمل به این دلیل است که اگر تعداد پیوندها از یک دامنه خاص به یک صفحه زیاد باشد، به احتمال زیاد پیوندهای بی‌ارزش بوده و بیشتر جنبه تبلیغاتی دارند.

این انتظار وجود دارد که در گراف S مجموعه غنی‌ای از authority ها وجود داشته باشند، زیرا به احتمال زیاد، یک authority توسط حداقل تعداد کمی از صفحات مجموعه ریشه، پیوند ورودی دارد. به استدلال مشابه انتظار می‌رود، hub های خوبی نیز در S حضور داشته باشند.

^۱ Heuristic

۳-۳-۲ الگوریتم PicASHOW

این روش تحت تأثیر الگوریتم PageRank عمل می‌کند و از آن برای محاسبه میزان اعتبار صفحات بازیابی شده استفاده می‌کند و بر اساس همین میزان اعتبار، به تصاویر داخل آن، امتیازی تخصیص می‌دهد و رتبه‌بندی را بر اساس این امتیاز انجام می‌دهد. فرض این روش بر این است که صفحه با اعتبار بالاتر، تصاویر با بار اطلاعاتی بیشتر و در نتیجه تصاویر مهم‌تری را در بر دارد. این روش هیچ نیازی به تحلیل محتوای دیداری تصاویر ندارد.

۴-۲ روش‌های مبتنی بر رفتار کاربر

فرآیند اصلی این روش بر مبنای تغییر رتبه‌بندی تصاویر بر اساس رفتار کاربر می‌باشد. برای این منظور نیاز به جمع‌آوری داده‌هایی از رفتار کاربر داریم. جمع‌آوری داده‌های مربوط به رفتار کاربر، به دو صورت انجام می‌پذیرد:

- بازخورد مستقیم^۱

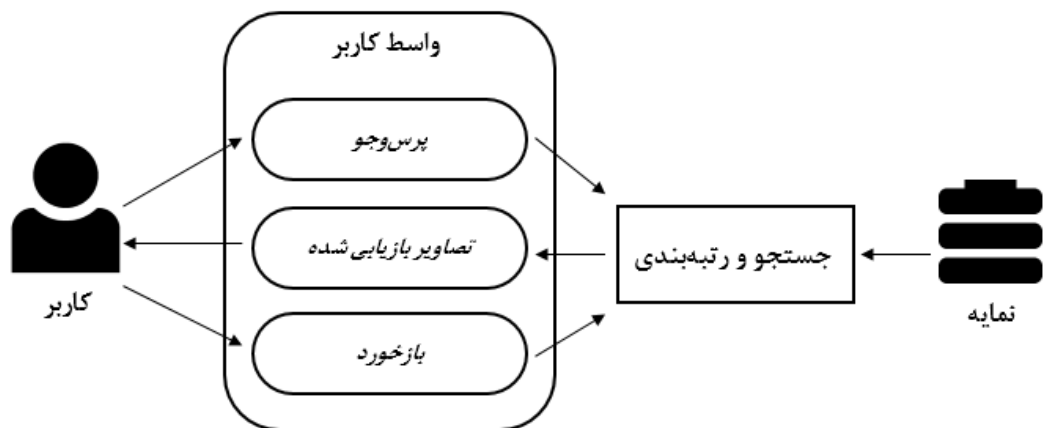
- بازخورد ضمنی^۲ [۵]

در بازخورد مستقیم، از کاربر به طور مستقیم خواسته می‌شود تا در مورد نتیجه اعلام نظر کند. درحالی‌که تجربه نشان داده بیشتر کاربران مایل به ارائه چنین بازخوردی نیستند. در روش بازخورد ضمنی، به جای این که به طور مستقیم نظر کاربر را جویا شویم، از تحلیل فایل‌های موتور جستجو که به ثبت وقایع^۳ می‌پردازند استفاده می‌کنیم. در نتیجه بیشتر از بازخورد ضمنی برای جمع‌آوری داده‌های مربوط به رفتار کاربران استفاده می‌شود. این وقایع شامل اطلاعات ارزشمندی هستند که تحلیل آن‌ها می‌تواند در مدل‌سازی رفتار کاربران در جستجوی تصویر بسیار مفید باشد.

^۱ Explicit feedback

^۲ Implicit feedback

^۳ Log



شکل ۲-۴: شمای کلی رتبه‌بندی مبتنی بر رفتار کاربر

۲-۵ روش‌های ترکیبی

روش‌هایی که در بالا بیان شدند، اگر به تنهایی مورد استفاده قرار گیرند، از دقت پایینی برخوردارند. به همین منظور نیاز داریم که برای رسیدن به دقت بالاتر، از الگوریتم‌های ترکیبی برای رتبه‌بندی تصویر استفاده کنیم.

فرآیند روش‌های ترکیبی به این صورت می‌باشد که امتیاز نهایی هر تصویر بر اساس مجموع وزنی امتیاز هر یک از روش‌های به کار گرفته شده در این روش، به دست می‌آید. اگر $S_{m,i}$ امتیاز روش m به تصویر i ام باشد و w_m وزن مربوط به امتیاز روش m ام باشد، امتیاز نهایی تصویر i ام به صورت زیر محاسبه می‌شود:

$$S_{total,i} = \sum_{m=1}^M S_{m,i} \quad (3-2)$$

از جمله این روش‌ها می‌توان، روش ترکیبی محتوای متنی با دیداری تصویر [۳۹] و روش ترکیبی محتوای دیداری تصویر با بازخورد کاربر [۱۴] را نام برد.

۳ روش پیشنهادی

۳-۱ مقدمه

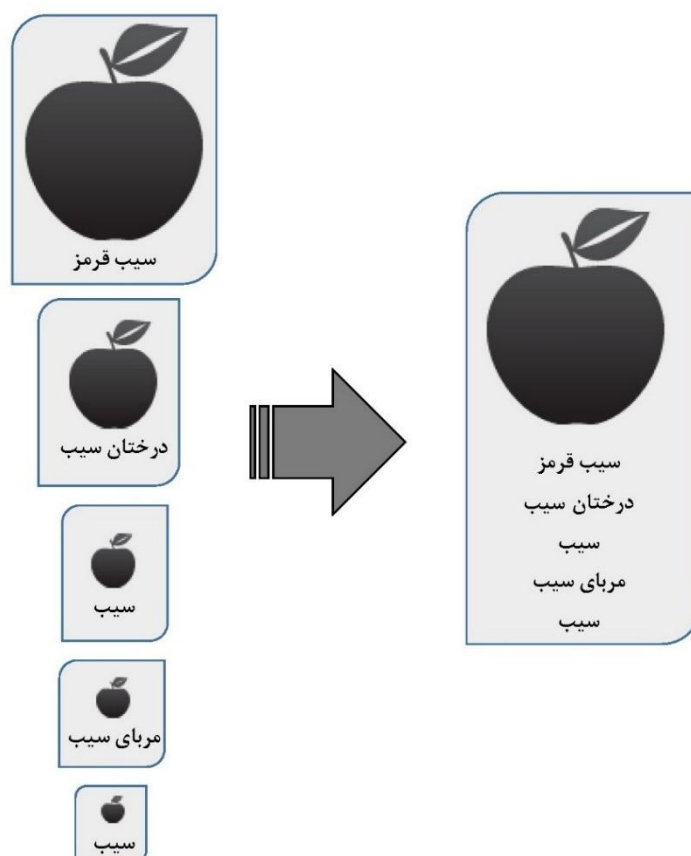
با مطالعه بر روی کارهای پیشین انجام شده برای رتبه‌بندی تصاویر وب، به این نتیجه می‌توان رسید که اغلب کارهای انجام شده در این زمینه یا از دقت کافی برخوردار نیستند و یا این که درجه پیچیدگی بالایی دارند. روش‌های مبتنی بر محتوای متنی، روش‌های مبتنی بر اتصال و روش‌های مبتنی بر رفتار کاربر، اگر به تنهایی برای رتبه‌بندی تصاویر به کار گرفته شوند، از دقت پایینی برخوردار هستند. روش‌های مبتنی بر محتوای دیداری تصاویر اگرچه با در نظر گرفتن شباهت بین تصاویر می‌توانند رتبه‌بندی تصاویر را بهبود ببخشند، ولی برای محاسبه توصیفگرهای تصویر در فاز برون خط و اجرای مقایسه دوه‌دوی تصاویر در فاز برخط نیاز به پردازشی با درجه پیچیدگی بالا دارند. روش‌های ترکیبی نیز برای بالا بردن دقت نیاز دارند تا از محتوای دیداری تصاویر برای رتبه‌بندی در فاز برخط استفاده کنند که همین عامل اگر چه باعث بالا رفتن دقت آن‌ها می‌شود ولی از طرف دیگر از درجه پیچیدگی بالایی برخوردارند.

با این تفاسیر و با توجه به نیاز کاربر در دقت و سرعت موتور جستجو، نیاز به الگوریتمی برای رتبه‌بندی تصاویر داریم تا هر دو نیاز را برآورده سازد. در بخش بعدی به تشریح الگوریتم پیشنهادی خواهیم پرداخت.

۳-۲ الگوریتم پیشنهادی

با توجه به آن‌چه گفته شد، الگوریتم رتبه‌بندی باید بتواند دو ویژگی دقت و سرعت در رتبه‌بندی تصاویر را برآورده سازد. روش پیشنهادی ترکیبی از روش رتبه‌بندی بر اساس محتوای متنی و به‌کارگیری محتوای دیداری تصاویر است. در روش رتبه‌بندی بر اساس محتوای متنی، هر تصویر به همراه ویژگی‌های متنی اطراف آن تصویر، نمایه‌سازی می‌شود. هنگام جستجو، کلیدواژه‌های پرس‌وجوی کاربر، در نمایه‌ها مورد جستجو قرار گرفته و تصاویری که از نظر متنی، با پرس‌وجوی

کاربر مرتبط‌تر باشند با رتبه بالاتر به کاربر نشان داده خواهند شد. بر خلاف جستجو در صفحات وب که پرس‌وجو و نتایج همه از یک نوع هستند، در جستجوی تصویر به دلیل غیر هم نوع بودن پرس‌وجو و نتایج ممکن است تصاویر نمایه‌سازی شده با متن اطرافش نامرتب باشد. به همین دلیل این روش‌ها از دقت کافی برای رتبه‌بندی تصاویر برخوردار نمی‌باشند. به همین منظور در روش پیشنهادی برای بالا بردن دقت، در فاز برون خط، اسناد تصویری دقیقاً مشابه و یا مقیاس^۱ شده را با هم ادغام می‌کنیم. هنگام جستجو، کلیدواژه‌های پرس‌وجوی کاربر، در متن‌های ادغام‌شده هر سند تصویر جستجو شده و رتبه‌بندی هر تصویر بر اساس کلیه متن‌هایی انجام می‌شود که در مورد آن تصویر در صفحات خز شده وب وجود دارد. با این روش می‌توانیم رتبه‌بندی مجدد^۲ را از فرآیندهای فاز برخط حذف کنیم. در ادامه به تشریح گام‌های روش پیشنهادی می‌پردازیم.



شکل ۳-۱: ادغام اسناد تصویری یکسان و مقیاس شده

^۱ Scale

^۲ Rerank

۳-۲-۱ گام اول: شناسایی تصاویر یکسان و مقیاس شده

برای شناسایی تصاویر یکسان و مقیاس شده، راه‌های مختلفی وجود دارد. اغلب این روش‌ها برای این کار، ویژگی‌های مختلفی از تصاویر استخراج می‌کنند. سپس با مقایسه ویژگی‌های دوبه‌دوی تصاویر، میزان شباهت آن‌ها را سنجیده و بر اساس یک حد آستانه^۱ شبیه بودن آن دو تصویر را تعیین می‌کنند. اگر n تعداد تصاویر باشد، درجه پیچیدگی این نوع الگوریتم‌ها جدا از محاسبات لازم برای استخراج ویژگی‌های تصاویر، برای مقایسه دوبه‌دوی تصاویر، $O(n^2)$ می‌باشد. با این شرایط این الگوریتم‌ها برای شناسایی تصاویر مشابه به زمانی بسیار طولانی نیازمندند که غیرقابل اجرا می‌باشد. بنابراین به روشی برای شناسایی تصاویر مشابه نیازمندیم تا بتواند با سرعت خیلی بالاتری این کار را انجام دهد.

اگر بتوانیم از هر تصویر یک شناسه تولید کنیم که برای مقیاس‌های مختلف آن تصویر، این شناسه یکتا باشد در نتیجه می‌توانیم تصاویر مشابه را با درجه پیچیدگی $O(k)$ بیابیم. یعنی با تشکیل جدول درهم^۲، که کلید^۳ آن شناسه و مقدار^۴ آن، تصویر و سایر ویژگی‌های آن سند تصویری باشد، می‌توانیم این کار را انجام دهیم.

برای محاسبه چنین شناسه‌ای نیاز به الگوریتمی است که بر اساس محتوای دیداری هر تصویر، یک کد درهم^۵ تولید کند به طوری که این کد برای تصاویر مشابه، یکتا و برای تصاویر غیرمشابه متفاوت باشد. شکل ۳-۲ روندنمای^۶ تولید این شناسه برای هر تصویر را به صورت کلی نشان می‌دهد. این الگوریتم بهبودیافته الگوریتم یافتن کد درهم بیان شده در [۴۰] می‌باشد.

¹ Threshold

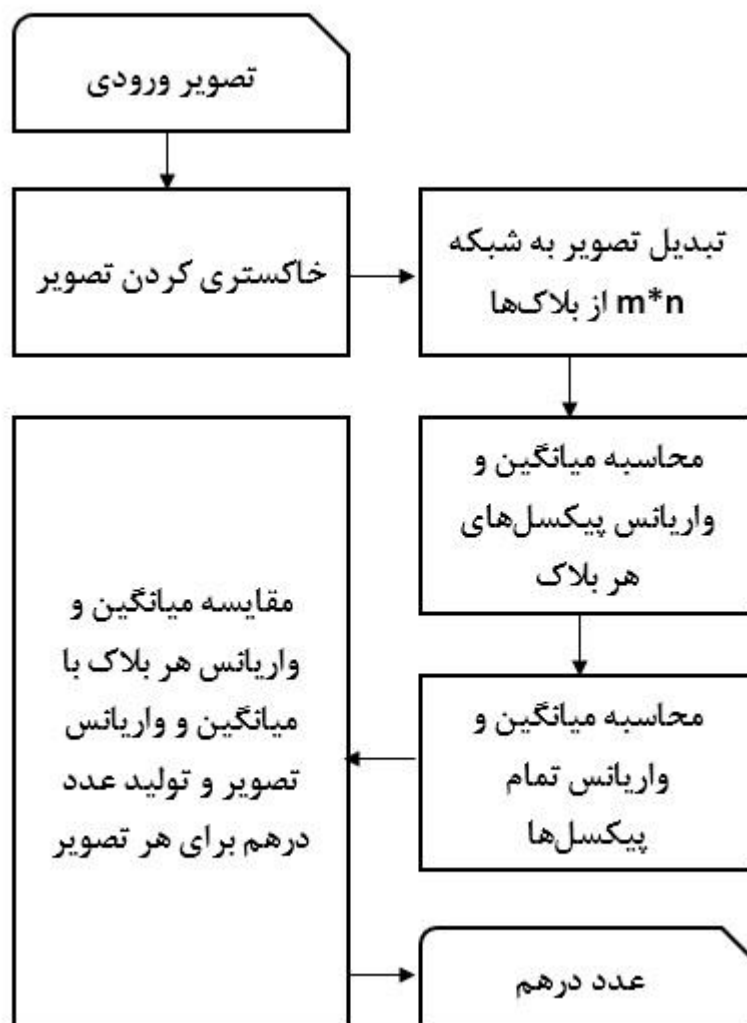
² Hash table

³ Key

⁴ Value

⁵ Hash code

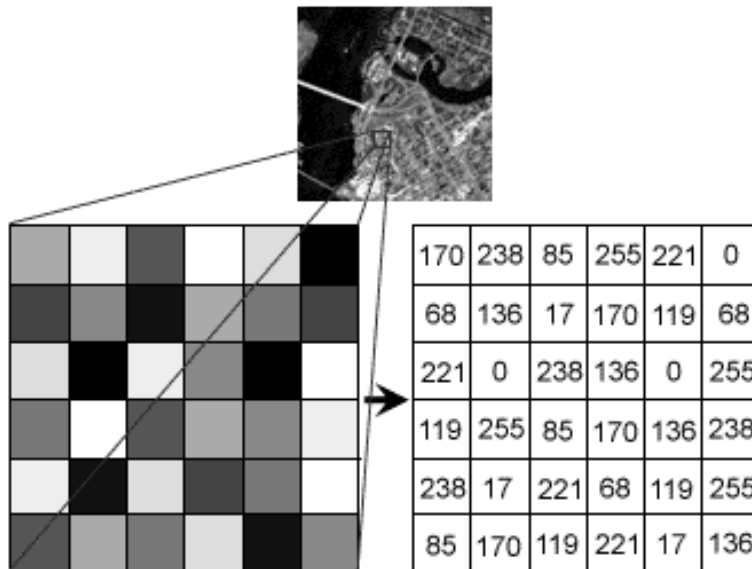
⁶ Flowchart



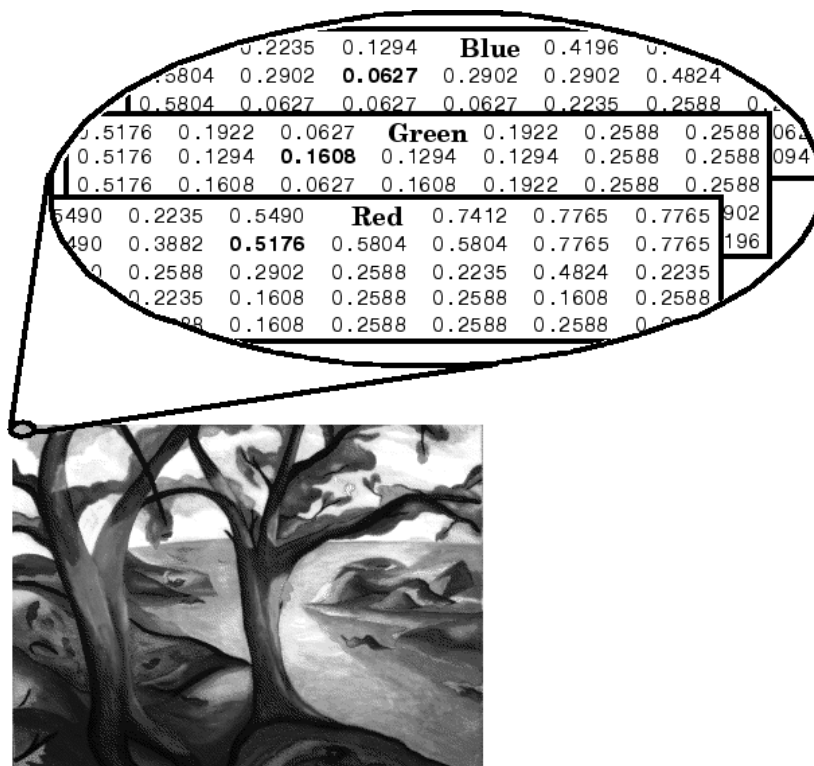
شکل ۳-۲: روندنمای الگوریتم تولید شناسه برای هر تصویر

تصویر ورودی می‌تواند خاکستری^۱ و یا رنگی باشد. ولی ورودی الگوریتم باید یک تصویر خاکستری باشد. هر تصویر رقمی خاکستری، ماتریسی دو بعدی از اعداد ۸ بیتی است. هر عنصر از ماتریس نشان‌دهنده شدت نور پیکسل متناظر است و میزان شدت با عددی بین ۰ تا ۲۵۵ مشخص می‌شود. نمونه‌ای از یک تصویر رقمی خاکستری را در شکل ۳-۳ مشاهده می‌کنید. تصاویر رنگی، با استفاده از ۳ ماتریس دو بعدی از اعداد ۸ بیتی تشکیل می‌شود. ماتریس‌ها به ترتیب شدت رنگ‌های قرمز، سبز و آبی مربوط به کلیه پیکسل‌های تصویر را مشخص می‌کنند. شکل ۳-۴ نمونه‌ای از یک تصویر رقمی رنگی می‌باشد.

^۱ Grayscale



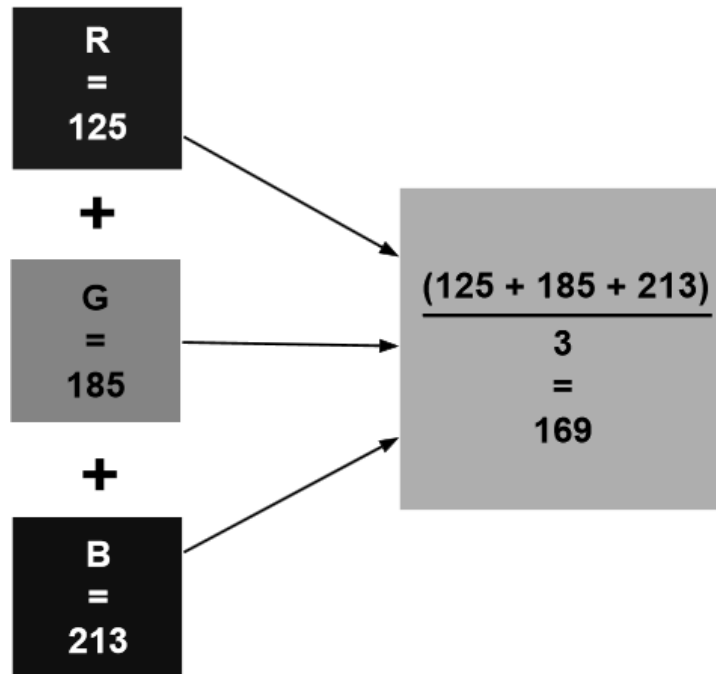
شکل ۳-۳: نمونه‌ای از یک تصویر رقمی خاکستری



شکل ۳-۴: نمونه‌ای از یک تصویر رقمی رنگی

برای تبدیل یک تصویر رقمی رنگی به یک تصویر رقمی خاکستری راه‌های مختلفی وجود دارد. یکی از این روش‌ها گرفتن میانگین از شدت نور هر سه لایه رنگی به ازای هر پیکسل می‌باشد. یعنی یک ماتریس دوبعدی جدید هم اندازه با ماتریس‌های قرمز و سبز و آبی تشکیل داده و شدت

نور هر پیکسل را برابر با میانگین شدت نور پیکسل‌های معادل در سه لایه رنگی قرار می‌دهیم. در شکل ۳-۵ نمونه‌ای از نحوه محاسبه شدت نور پیکسل تصویر خاکستری را مشاهده می‌کنید.



شکل ۳-۵: محاسبه شدت نور پیکسل تصویر خاکستری

در مرحله بعد می‌بایست شبکه $m \times n$ از تصویر خاکستری بسازیم. به این معنی که ماتریس تصویر خاکستری را به گونه‌ای به بلاک‌های $w \times h$ پیکسل تقسیم کنیم که تعداد بلاک‌ها در هر ردیف m و در هر ستون n باشد. حال در هر بلاک به تعداد $w \times h$ پیکسل داریم که مقدار هر کدام عددی بین ۰ تا ۲۵۵ می‌باشد. میانگین پیکسل‌های هر بلاک را محاسبه می‌کنیم. به تعداد $m \times n$ میانگین خواهیم داشت که آن‌ها را به ترتیب در یک آرایه یک بعدی با طول $m \times n$ قرار می‌دهیم. سپس به ازای هر بلاک، مقدار واریانس تمام پیکسل‌های هر بلاک را محاسبه کرده و آن‌ها را به ترتیب در آرایه‌ای دیگر به طول $m \times n$ قرار می‌دهیم.

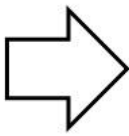
میانگین و واریانس کلیه پیکسل‌های تصویر خاکستری با نام‌های M_T و V_T را محاسبه می‌کنیم. حال دو آرایه یک بعدی با طول $m \times n$ داریم. یکی آرایه میانگین بلاک‌ها با نام M و دیگری آرایه واریانس بلاک‌ها با نام V . به ازای هر کدام از آرایه‌های میانگین‌ها و واریانس‌ها یک آرایه بیتی با نام‌های M' و V' با طول $m \times n$ به دست می‌آوریم به طوری که مقدار هر عنصر آن از طریق روابط (۳-۱) و (۳-۲) به دست می‌آید.

$$M'_i = \begin{cases} 0, & M'_i < M_T \\ 1, & M'_i \geq M_T \end{cases} \quad (1-3)$$

$$V'_i = \begin{cases} 0, & V'_i < V_T \\ 1, & V'_i \geq V_T \end{cases} \quad (2-3)$$

حال دو آرایه بیتی یک بعدی با طول $m \times n$ داریم. اگر بیت‌های دو آرایه را پشت سرهم قرار دهیم به یک کد درهم می‌رسیم که همان شناسه تولیدی برای هر تصویر خواهد بود. در ادامه مثالی از چگونگی تولید این کد آورده شده است.

10	12	13	43
24	1	34	50
98	21	23	20
100	120	10	112



10	12	13	43
24	1	34	50
98	21	23	20
100	120	10	112

$$M = [11, 35, 84, 41] \quad M_T = 43$$

$$V = [67, 193, 1428, 1691] \quad V_T = 1541$$

$$M' = [0, 0, 1, 0]$$

$$V' = [0, 0, 0, 1] \Rightarrow \text{Hash} = 00100001$$

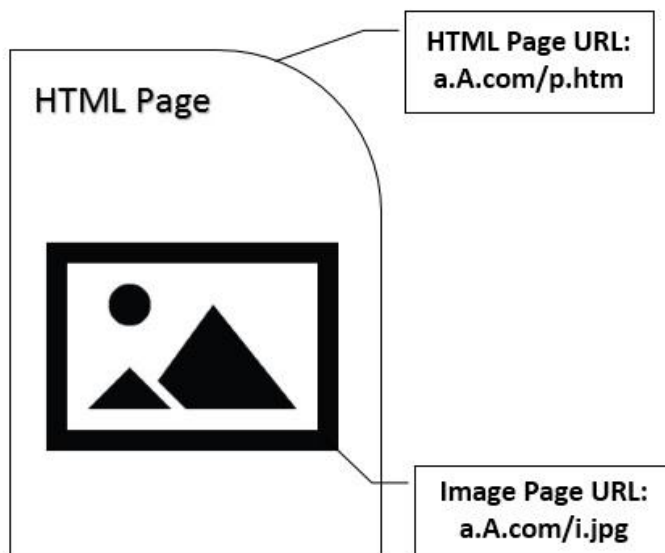
شکل ۳-۶: تولید کد درهم از تصویر خاکستری

۲-۲-۳ گام دوم: ادغام اسناد تصویری یکسان و مقیاس شده و نمایه‌سازی آن‌ها

پس از شناسایی اسناد تصویری با تصاویر یکسان و یا مقیاس شده، نوبت به ادغام این اسناد با یکدیگر می‌رسد. برای بالا بردن دقت در جستجو، علاوه بر محتوای متنی اطراف تصاویر، متن‌های دیگری را نیز به عنوان فیلدهای قابل جستجوی اسناد تصویری در نظر می‌گیریم. قبل از هرگونه توضیح در مورد این فیلدها، با چند مفهوم آشنا می‌شویم.

تصویر درون هاست^۱

اگر تصویر و صفحه محتوای تصویر در یک هاست قرار داشته باشند، آن را تصویر درون هاست می نامیم. شکل ۳-۷ نمونه ای از تصویر درون هاست را نشان می دهد.



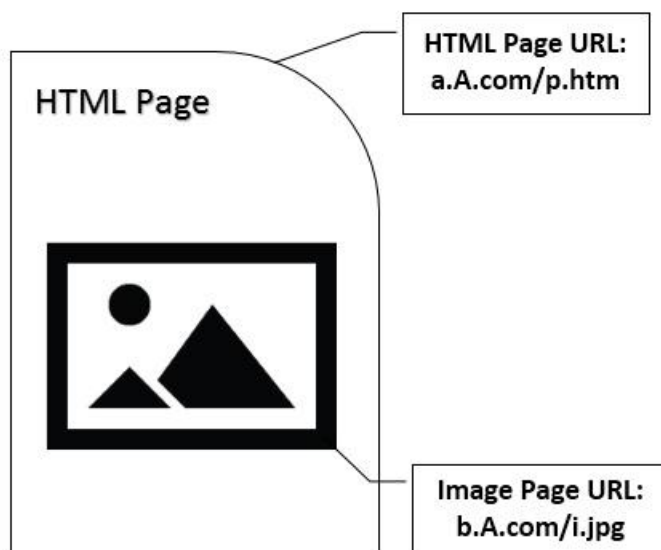
شکل ۳-۷: نمونه ای از تصویر درون هاست

تصویر درون دامنه^۲

اگر تصویر و صفحه محتوای تصویر در یک هاست نباشند ولی در یک دامنه باشند، آن را تصویر درون دامنه می نامیم. شکل ۳-۸ نمونه ای از یک تصویر درون دامنه را نشان می دهد.

¹ Host

² Domain

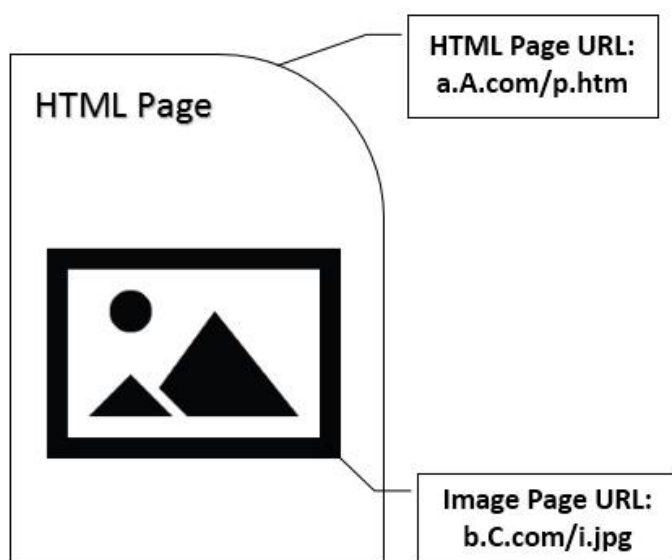


شکل ۳-۸: نمونه‌ای از تصویر درون دامنه

تصویر خارجی^۱

اگر تصویر و صفحه محتوای تصویر در یک دامنه نباشند، آن را تصویر خارجی می‌نامیم. شکل

۳-۹ نمونه‌ای از تصویر خارجی را نشان می‌دهد.

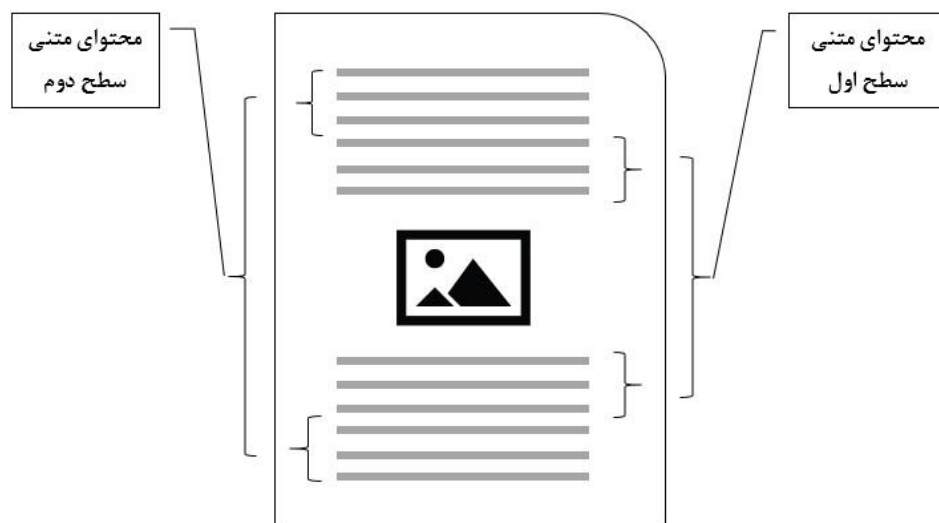


شکل ۳-۹: نمونه‌ای از تصویر خارجی

^۱ External

۱-۲-۲-۳ فیلدهای متنی اسناد تصویر

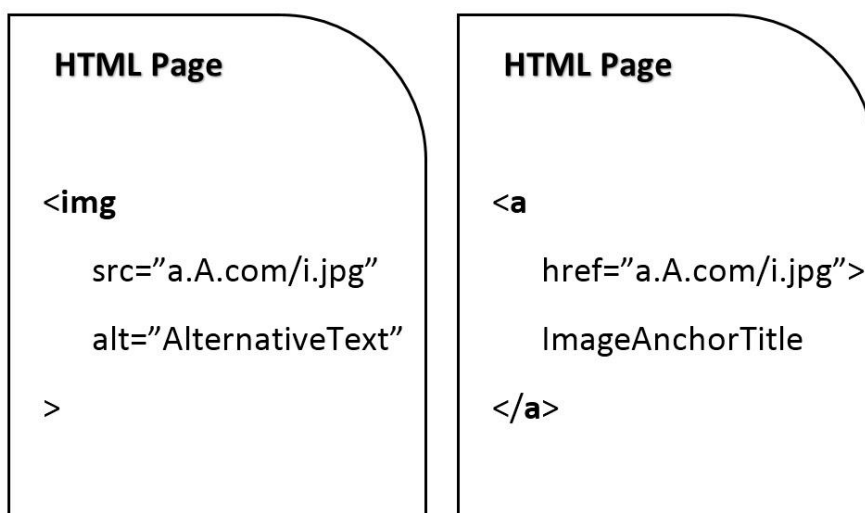
محتوای متنی اطراف تصویر را در دو فیلد محتوای متنی سطح اول و محتوای متنی سطح دوم تقسیم نموده‌ایم. محتوای متنی سطح اول، ۲۰۰ کاراکتر بالا و ۲۰۰ کاراکتر پایین تصویر می‌باشد و از اهمیت بالاتری برخوردار است. محتوای متنی سطح دوم، ۲۰۰ کاراکتر بعدی بالای تصویر و ۲۰۰ کاراکتر بعدی پایین تصویر می‌باشد که از اهمیت کمتری برخوردار است.



شکل ۱۰-۳: سطوح مختلف محتوای متنی اطراف تصویر

همان‌طور که گفته شد، برای بالا بردن دقت در جستجوی تصویر، فیلدهای متنی دیگری را برای هر سند تصویر در نظر می‌گیریم. تصاویر به دو شکل در صفحات وب قرار داده می‌شوند. یکی از طریق ابرپیوند، و دیگری از طریق برچسب تصویر^۱. شکل ۱۱-۳ شکل‌های مختلف قرار گرفتن تصویر در صفحات وب را نشان می‌دهد.

^۱ Img tag



شکل ۳-۱۱: نحوه قرار گرفتن تصویر در صفحه وب

فیلدهای متنی تصویر می‌تواند توسط برچسب پیوند^۱ در صفحه وب، مورد استفاده قرار بگیرد. این نوع فیلدها خود به سه بخش درون‌هاست، درون‌دامنه و خارجی تقسیم می‌شود. فیلدهای متنی دیگر مربوط به حالتی است که تصویر توسط برچسب تصویر در صفحه وب، مورد استفاده قرار گرفته باشد. در این حالت فیلدهای متنی تصویر، از یک طرف به سه بخش درون‌هاست، درون‌دامنه و خارجی تقسیم شده و از طرف دیگر در هر بخش شامل فیلدهای مختلفی می‌باشد. این فیلدها شامل موارد زیر است:

- متن اطراف تصویر سطح اول
- متن اطراف تصویر سطح دوم
- عنوان صفحه حاوی تصویر^۲
- عنوان تصویر^۳
- متن جایگزین تصویر^۴

^۱ a tag

^۲ Page title

^۳ Image title property

^۴ Image alt property

ادغام کردن اسناد تصویری مشابه، به این معنا است که پس از تشخیص مشابه بودن تصاویر دو سند تصویر، به جای اختصاص دادن یک فیلد مشخص به تصویر، آرایه‌ای از آن فیلد مشخص را به تصویر اختصاص دهیم. این فرآیند را در هنگام نمایه‌سازی انجام می‌دهیم.

۳-۲-۳ گام سوم: فرآیند جستجوی تصویر

پس از ادغام اسناد تصویری با تصاویر یکسان و مقیاس شده و نمایه‌سازی تمامی اسناد، نوبت به فرآیند جستجوی تصویر می‌رسد. در این فرآیند، پرس‌وجوی کاربر از طریق واسط کاربر دریافت می‌شود. توسط الگوریتم $tf-idf$ ، اسناد تصویری مرتبط، بازیابی و رتبه‌بندی می‌شوند و سپس توسط همان واسط کاربر، به کاربر نمایش داده می‌شوند.

در گام قبل دیدیم که برای بالا بردن دقت جستجوی تصویر، علاوه بر محتوای متنی اطراف تصویر، از چندین فیلد متنی دیگر مرتبط با تصویر نیز استفاده می‌کنیم. با توجه به این که اهمیت همه این فیلدها با هم برابر نیست، هنگام جستجوی تصویر، جستجو در هر فیلد را با وزنی جداگانه انجام می‌دهیم. ولی در ابتدای جستجو به دلیل عدم آگاهی از اهمیت هر فیلد متنی، مقدار همه را یکسان در نظر می‌گیریم.

۳-۳ پیاده‌سازی روش پیشنهادی و ایجاد مجموعه داده محک

الگوریتم پیشنهادی در محیط جاوا پیاده‌سازی شده و بر روی تعداد زیادی سند تصویر که تعداد آن بالغ بر ۲۰ میلیون سند می‌باشد اجرا شده است. پس از شناسایی تصاویر یکسان و مقیاس‌شده، تعداد اسناد نمایه به ۱۳ میلیون سند کاهش یافته است.

همان‌طور که در بخش قبل گفته شد، در ابتدای جستجو به دلیل عدم آگاهی از اهمیت هر فیلد متنی، در جستجو توسط الگوریتم $tf-idf$ وزن تمام فیلدها را یکسان در نظر می‌گیریم. برای بهبود جستجو، می‌بایست وزن‌ها تنظیم شوند. به همین منظور به مجموعه داده محکی نیاز داریم تا از یک طرف وزن‌ها را با آن تنظیم کنیم و از طرف دیگر بتوانیم میزان بهبود را پس از تنظیم وزن‌ها

محاسبه کنیم. به دلیل عدم وجود چنین مجموعه داده‌ای با این مشخصات، سامانه‌ای را برای ساخت مجموعه داده محک مورد نیاز، ایجاد نموده‌ایم. این سامانه همان موتور جستجوی تصویری است که هنوز وزن فیلدهای آن تنظیم نشده است. کاربر پرس‌جوی خود را به این سامانه وارد و موتور جستجو نتایجی را به کاربران ارائه می‌کند که از دید الگوریتم پیشنهادی، با پرس‌وجوی کاربر مرتبط است. کاربر می‌تواند به هر نتیجه یک درجه از نظر کیفی نسبت دهد. درجه‌های کیفی شامل خوب، متوسط و بد می‌باشد. چندین کاربر با جستجوی پرس‌وجوهای مجزا و درجه‌بندی کیفی آن نتایج براساس مرتبط بودن با پرس‌وجو، مجموعه داده محکی با تعداد رکوردی بالغ بر ۴۶۰۰ رکورد ایجاد نموده‌اند. کاربران به ازای هر پرس‌وجو، ۱۸۰ نتیجه اول را درجه‌بندی کرده‌اند. هر رکورد این مجموعه داده، شامل ۱ ستون برای پرس‌وجوی کاربر، ۱۸ ستون برای امتیاز^۱ هر فیلد در جستجوی کاربر و یک ستون برای مشخص نمودن درجه یا همان کلاسی^۲ است که کاربر به نتیجه اختصاص داده است. شکل ۳-۱۲ واسط کاربر سامانه ایجاد مجموعه داده محک برای جستجوی پرس‌وجوی "سبزه" را نشان می‌دهد.



شکل ۳-۱۲: واسط کاربر سامانه ایجاد مجموعه داده محک

¹ Score

² Class

داده‌های محک مربوط به ۱۰ درصد از پرس‌وجوهای مجموعه داده محک، برای آزمون و ۹۰ درصد دیگر برای آموزش و تنظیم وزن‌ها انتخاب شد. نتایج حاصل روی داده‌های آزمون نشان می‌دهد، دقت متوسط الگوریتم پیشنهادی در حالتی که وزن فیلدها تنظیم نشده است برابر با ۵۴/۰ می‌باشد.

۳-۴ بهبود روش پیشنهادی با روش یادگیری جفتی

مسئله رتبه‌بندی تصاویر، از دیدی دیگر، تابعی است که ورودی آن ویژگی‌های یک سند تصویر است و این ویژگی‌ها بر اساس پرس‌وجوی خاص محاسبه شده است. خروجی این تابع یک عدد به عنوان رتبه تصویر در آن پرس‌وجو می‌باشد. زمانی که تعداد این ویژگی‌ها کم باشد، می‌توان به صورت تجربی وزن این ویژگی‌ها را تعیین نمود. اما در مسئله تصویر تعداد این ویژگی‌ها به حدی است که تنظیم وزن‌ها را دشوار ساخته است. با مجموعه‌ای از قضاوت‌های درست انسانی در مورد نتایج هر پرس‌وجو و با به کارگیری یک الگوریتم یادگیری با ناظر، می‌توان وزن‌ها را به طور دقیق تنظیم نمود. داده‌های ما به صورت مجموعه‌ای از پرس‌وجوها است که به ازای هر پرس‌وجو تعدادی نتیجه با ویژگی‌های مختص خود بازیابی شده و به هر نتیجه بر اساس آن پرس‌وجو، رتبه‌ای تعلق گرفته است. در این روش یادگیری، با دریافت جفت اشیاء (ویژگی‌های اسناد) از پرس‌وجوی ثابت و نیز رتبه نسبی آن‌ها، تلاش می‌شود به هر شیء، رتبه‌ای حتی المقدور نزدیک به رتبه واقعی‌اش نسبت داده شود و بدین ترتیب، نهایتاً اشیاء در دو دسته کلی "به صورت صحیح رتبه‌بندی شده"^۱ و "به صورت نادرست رتبه‌بندی شده"^۲، طبقه‌بندی می‌شوند. اگرچه تعداد حالت انتخاب جفت اشیاء از مجموعه با طول n برابر با $\binom{n}{2}$ می‌باشد، ولی به دلیل این که جفت اشیاء باید از یک پرس‌وجو باشند، n مقدار کوچکی است و تعداد این جفت‌ها نسبت به بزرگی مجموعه داده‌های محک، افزایش نمی‌یابد. شایان ذکر است که اغلب روش‌های موجود رتبه‌بندی مبتنی بر یادگیری از این نوع هستند.

^۱ Correctly-ranked

^۲ Incorrectly-ranked

مجموعه داده محک ایجاد شده که در بخش قبل در مورد نحوه ایجاد آن توضیح داده شد به صورت $DataSet = \{(x, y, q) \mid x = featureVector, y = class, q = userQuery\}$ می‌باشد. یک سه‌تایی است که x بردار ویژگی‌های سند تصویر یا همان امتیاز هر فیلد سند تصویر در جستجوی پرس‌وجوی کاربر است. y کلاس یا درجه‌ای است که کاربر به نتیجه پرس‌وجوی q داده است. وقتی $y_a > y_b$ باشد، بدان معنا است که سند a بر سند b در رتبه‌بندی، برتری دارد. به عبارت دیگر در رتبه‌بندی، سند a رتبه بالاتری نسبت به سند b دارد. دیدیم که در الگوریتم پیشنهادی، برای جستجو در ویژگی‌های متنی، از وزن یکسان برای هر ویژگی استفاده کردیم. یعنی از بردار وزن w که طولی برابر با طول بردار ویژگی‌ها دارد و مقدار تمام عناصر آن مقداری یکسان است.

در این بخش توسط الگوریتم یادگیری جفتی SPD^۱ بیان شده در [۴۱] وزن‌های w را تنظیم می‌کنیم. شبه کد ۳-۱ این الگوریتم را به صورت خلاصه بیان می‌کند.

شبه کد ۳-۱: الگوریتم یادگیری به روش جفتی

```

1:  $w_0 \leftarrow \emptyset$ 
2: for  $i = 1$  to  $t$ 
3:    $((a, y_a, q), (b, y_b, q)) \leftarrow RandomCandidatePair(P)$ 
4:    $x \leftarrow (a - b)$ 
5:    $y \leftarrow (y_a - y_b)$ 
6:    $\eta_i \leftarrow \frac{1}{i\lambda}$ 
7:    $w_i \leftarrow (1 - \eta_i\lambda)w_{i-1} + \eta_i x(y - \langle w_{i-1}, x \rangle)$ 
8: end for
9: return  $w_t$ 

```

$\langle w_{i-1}, x \rangle$ بیان کننده ضرب داخلی بردار وزن‌ها و بردار ویژگی‌ها می‌باشد. λ پارامتر تنظیم و t تعداد تکرار الگوریتم برای زوج‌های کاندید جدیدتر است. با قرار دادن مقدار λ برابر ۵، وزن‌ها در زودترین زمان به مقدار نهایی همگرا شدند.

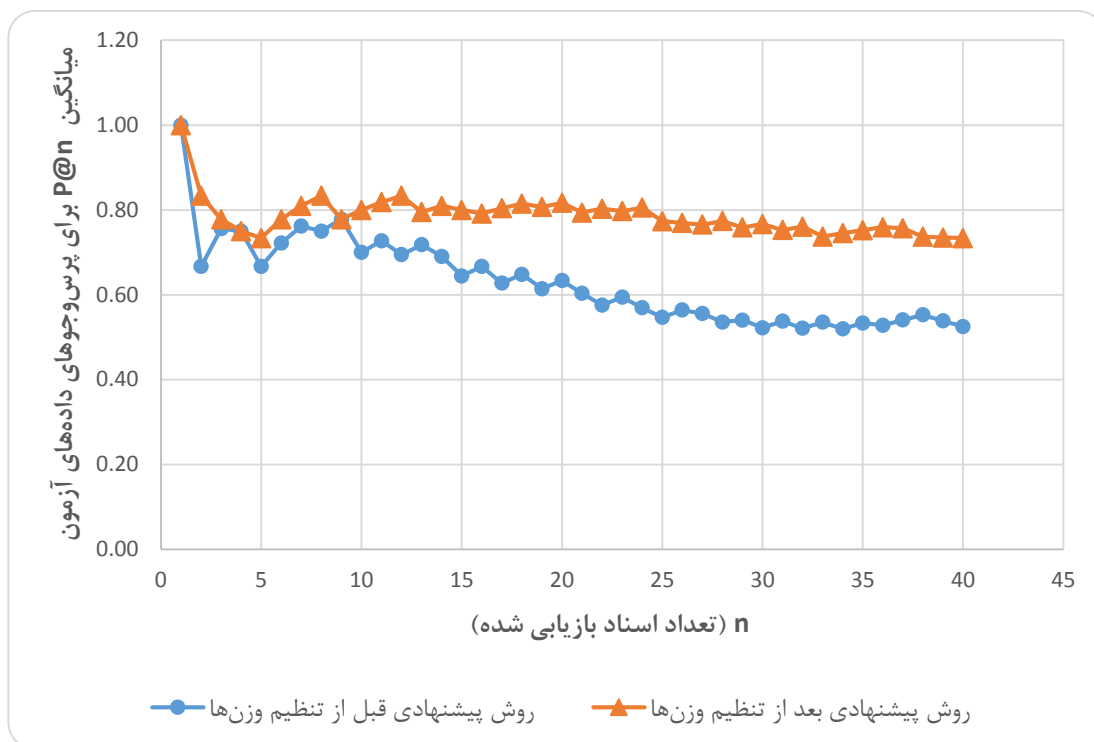
^۱ Stochastic pairwise descent

تعداد پرس‌وجوهای مجموعه داده محک برابر ۴۰ پرس‌وجوی مجزا می‌باشد. یادگیری به ازای جفت اشیا از پرس‌وجوی ثابت انجام می‌گیرد. پس از اتمام یادگیری، آرایه وزن‌های W به دست آمده را در الگوریتم جستجو اعمال می‌کنیم. برای ارزیابی، الگوریتم را روی داده‌های آزمون اجرا کردیم.

۳-۵ ارزیابی روش‌های پیشنهادی

همان‌طور که دیدیم روش پیشنهادی ما با ادغام اسناد تصویری یکسان، بازیابی و رتبه‌بندی تصاویر را بر اساس محتوای متنی اسناد تصویری انجام می‌دهد و جستجو در فیلدهای متنی اسناد تصویری را با وزنی یکسان برای هر فیلد اجرا می‌گردد. در گام بعدی، با استفاده از یک الگوریتم یادگیری جفتی این وزن‌ها تنظیم شدند. این بدان معنا است که جستجو در هر فیلد با وزن مربوط به آن فیلد که نشان دهنده اهمیت محتوای متنی آن فیلد است، انجام می‌گیرد.

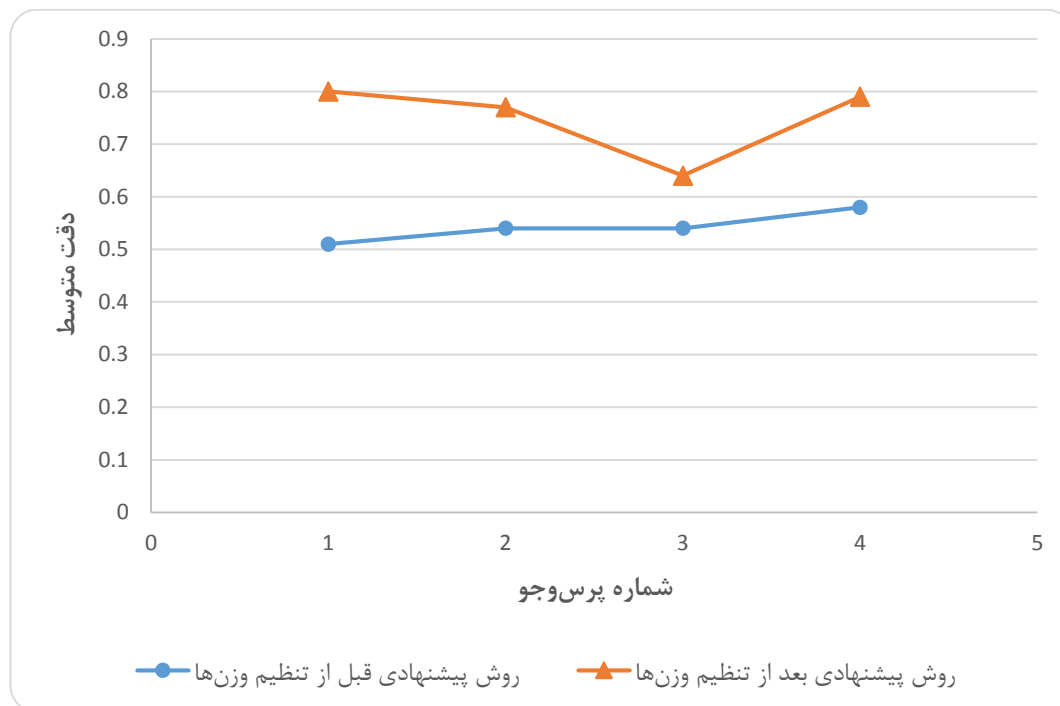
از نظر تئوری، انتظار داریم پس از تنظیم وزن‌ها و مشخص شدن اهمیت هر فیلد، رتبه‌بندی با دقت بالاتری انجام شود. برای این منظور، به انجام آزمایش‌هایی روی داده‌های آزمون پرداختیم تا آن‌چه انتظار می‌رود را به اثبات برسانیم. در ادامه این آزمایش‌ها را مورد بررسی قرار می‌دهیم.



شکل ۳-۱۳: ارزیابی روش پیشنهادی پیش و پس از تنظیم وزن‌ها در معیار $P@n$

شکل ۳-۱۳ نمودار ارزیابی روش پیشنهادی پیش و پس از تنظیم وزن‌ها توسط الگوریتم یادگیری جفتی در معیار ارزیابی $P@n$ می‌باشد. این نمودار نشان می‌دهد روش پیشنهادی پس از تنظیم وزن‌ها، دقت جستجو را تا حد بالایی افزایش دهد. این معیار بر روی پرس‌وجوهای داده‌های آزمون، مورد بررسی و ارزیابی قرار گرفته است.

شکل ۳-۱۴ نمودار ارزیابی روش پیشنهادی پیش و پس از تنظیم وزن‌ها در معیار ارزیابی دقت متوسط را نشان می‌دهد. میانگین دقت متوسط برای روش پیشنهادی پیش از تنظیم وزن‌ها برابر 0.5425 و پس از تنظیم وزن‌ها برابر با 0.75 می‌باشد. همان‌گونه که مشخص است، پس از تنظیم وزن‌ها، دقت متوسط و میانگین دقت متوسط الگوریتم پیشنهادی افزایش چشم‌گیری داشته است.



شکل ۳-۱۴: ارزیابی روش پیشنهادی پیش و پس از تنظیم وزن‌ها در معیار دقت متوسط

۴ نتیجه‌گیری و کارهای آینده

۴-۱ نتیجه‌گیری

رشد سریع در تعداد تصاویر رقمی منتشرشده در وب، پیدا کردن تصاویر خاص را بسیار دشوار نموده است. از طرف دیگر کاربران نیاز دارند تا در کمترین زمان ممکن به تصاویر مورد نیاز خود دست پیدا کنند. از این رو موتورهای جستجو ابزارهایی کارآمد در این زمینه می‌باشند. کاربران پرس‌وجوی خود را توسط تعداد محدودی کلیدواژه مطرح کرده و موتور جستجوی تصویر نتایج مرتبط با این پرس‌وجو را بازایی و رتبه‌بندی کرده و به کاربران ارائه می‌کند. کاربران معمولاً به ۱۰ تا ۲۰ نتیجه اول اکتفا می‌کنند و در صورتی که به نتیجه نرسند از جستجو منصرف شده یا پرس‌وجوی خود را تغییر می‌دهند. در نتیجه رتبه‌بندی تصاویر در یک موتور جستجو نقشی حیاتی داشته و کیفیت یک موتور جستجوی تصویر به این فرآیند وابسته است.

در فصل دوم روش‌های پیشین رتبه‌بندی تصویر در موتورهای جستجو را بررسی کردیم. دیدیم که تمام روش‌های پیشین در چندین دسته کلی تقسیم می‌شوند. هر یک از این روش‌ها مزایا و معایب خود را داشتند. از معایب این روش‌ها می‌توان به نادیده گرفتن محتوای تصویری و دقت پایین و یا درجه پیچیدگی بالا در فاز برخط اشاره کرد.

هدف از انجام این پایان‌نامه، پیشنهاد یک الگوریتم رتبه‌بندی جدید برای موتور جستجوی تصویر است که ترکیبی از رتبه‌بندی بر اساس محتوای متنی و به‌کارگیری محتوای دیداری تصاویر می‌باشد. این روش از یک طرف دقت جستجو را بالا برده و از طرف دیگر با حذف فرآیند رتبه‌بندی مجدد در فاز برخط، به سرعت جستجو افزوده است.

۴-۲ دستاوردهای پایان‌نامه

در فصل سوم، روش پیشنهادی مطرح گردید. ابتدا در گام اول اسناد تصویری مشابه را با تولید یک کد درهم برای هر تصویر، شناسایی نمودیم. این کد برای تصاویر یکسان و یا مقیاس شده عددی

یکسان خواهد بود. سپس در گام دوم اسناد تصویری یکسان و مقیاس شده که در گام اول شناسایی شدند را با هم ادغام نموده و نمایه‌سازی نمودیم. در گام سوم عمل جستجوی تصویر با وزن‌های یکسان روی تمام فیلدهای متنی انجام شد.

در بخش بعدی ایده تنظیم وزن فیلدها توسط یک الگوریتم یادگیری جفتی مطرح شد. برای تنظیم وزن‌ها و اجرای چنین الگوریتمی نیاز به مجموعه داده محکی داشتیم که شامل فیلدهای مورد نظر باشد. به دلیل عدم وجود چنین مجموعه داده‌ای، سامانه‌ای برای جمع‌آوری این مجموعه داده ایجاد شد. چندین کاربر با انجام پرس‌وجو از سامانه، درجه کیفی هر تصویر نتیجه نسبت به پرس‌وجوی خود را مشخص نموده و مجموعه داده محک با تعداد رکوردهای بالغ بر ۴۶۰۰ رکورد ایجاد گردید. الگوریتم یادگیری جفتی بر روی ۹۰ درصد از داده‌های این مجموعه اجرا شد و وزن‌های نهایی تعیین شدند.

نتایج حاصل از بررسی الگوریتم پیشنهادی قبل و بعد از تنظیم وزن روی داده‌های آزمون نشان داد که دقت متوسط الگوریتم پیشنهادی قبل از تنظیم وزن، با وزن یکسان برای تمام فیلدهای متنی برابر با ۰/۵۴ و پس از تنظیم وزن‌ها توسط الگوریتم یادگیری جفتی برابر با ۰/۷۵ می‌باشد.

۴-۳ کارهای آینده

همانطور که در فصل دوم دیدم، روش‌های دیگری نیز برای رتبه‌بندی تصاویر وجود دارد. از جمله این روش‌ها، روش مبتنی بر اتصال و روش مبتنی بر رفتار کاربر بود. در روش‌های مبتنی بر اتصال می‌توان میزان اعتبار صفحات محتوی تصاویر را به دست آورد و از طریق رابطه‌ای میزان اعتبار هر تصویر را محاسبه نمود. از میزان اعتبار هر تصویر می‌توان به عنوان امتیاز دیگری برای رتبه‌بندی مجدد تصاویر استفاده نمود. از طرف دیگر می‌توان بازخوردهای مستقیم یا غیر مستقیم کاربران را بررسی نموده و بتوانیم به صورت برخط، اهمیت فیلدهای مختلف مورد جستجو را با تنظیم وزن‌های متناظر با آن فیلد را تنظیم کنیم.

الگوریتم یافتن تصاویر یکسان و مقیاس شده گرچه در تشخیص تصاویر مشابه خطای کمی دارد ولی نمی‌تواند تمامی آن‌ها را تشخیص دهد. از طرف دیگر این الگوریتم تنها می‌تواند تصاویر یکسان و مقیاس شده را بیابد. در صورتی که تصاویر مشابه دیگری نیز می‌تواند با تبدیل‌هایی غیر از مقیاس در وب موجود باشد. مثلاً تصاویر چرخش‌یافته و تصاویر با زوایای مختلف از یک صحنه، از جمله تصاویری هستند که الگوریتم مورد نظر در تشخیص مشابه بودن آن‌ها از خود ضعف نشان می‌دهد.

پیوست الف. تعریف واریانس و میانگین

الف-۱ واریانس^۱

در آمار و احتمال، واریانس نوعی سنجش پراکندگی^۲ برای یک توزیع احتمال یا متغیر تصادفی بوده، و نماینده پخش‌شدگی^۳ مقادیر آن حول مقدار میانگین است. واریانس را با σ^2 نشان می‌دهند و از رابطه الف-۱-۱ به دست می‌آید.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (\text{الف-۱-۱})$$

الف-۲ میانگین^۴

در آمار و احتمال، میانگین نوعی سنجش برای میزان مرکزیت^۵ داده‌ها است. میانگین را با \bar{x} نشان می‌دهند و از رابطه الف-۲-۱ به دست می‌آید.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (\text{الف-۲-۱})$$

^۱ Variance

^۲ Dispersion

^۳ Diffusion

^۴ Mean

^۵ Centrality

واژه‌نامه فارسی به انگلیسی

الف

Hyperlink	ابرپیوند
Visual Hyperlink	ابرپیوند دیداری
Probabilistic	احتمالی
Probabilistic	احتمالی
Relevancy	ارتباط
Evaluation	ارزیابی
Documents	اسناد
Authority	اعتبار
Score	امتیاز

ب

Feedback	بازخورد
Implicit feedback	بازخورد غیرمستقیم
Explicit feedback	بازخورد مستقیم
Retrieval	بازیابی
Image Retrieval	بازیابی تصویر
Texture	بافت
Histogram	بافت‌نما
High quality	باکیفیت
Online	برخط

Offline	برون خط
Correctly-ranked	به صورت صحیح رتبه‌بندی شده
Incorrectly-ranked	به صورت نادرست رتبه‌بندی شده
Boolean	بولی

پ

Diffusion	پخش‌شدگی
Dispersion	پراکندگی
Query	پرس‌وجو
Preprocessing	پیش‌پردازش

ت

Weighting function	تابع وزن‌دهی
Scale Invariant Feature Transform(SIFT)	تبدیل ویژگی مقیاس‌ناسته
Transformations	تبدیل‌ها
Parser	تجزیه‌گر
Combinational	ترکیبی
Harris Corner Detection	تشخیص گوشه هریس
Spin Images	تصاویر اسپین
Matching	تطبیق
Tuning	تنظیم
Term Frequency (TF)	تواتر کلمه

Distributed	توزیع شده
Image Descriptor	توصیفگرهای تصویر

ج

Hash table	جدول دَرهَم
Pairwise	جفتی

چ

Challenges	چالش‌ها
------------	---------

ح

Threshold	حد آستانه
-----------	-----------

خ

External	خارجی
Domain Specific	خاص منظوره
Grayscale	خاکستری
Crawler	خزشگر

د

Domain	دامنه
Precision	دقت

Average Precision	دقت متوسط
Binary	دودویی
Perspectives	دورنمای سه بعدی
Periodically	دوره‌ای

ر

Ranking	رتبه‌بندی
Probabilistic Ranking Principle (PRP)	رتبه‌بندی احتمالی
Regional traffic ranks	رتبه‌بندی ترافیک جهانی
Rerank	رتبه‌بندی مجدد
User behavior	رفتار کاربر
Flowchart	روندنما

ز

Shape Context	زمینه شکل
---------------	-----------

س

Low Level	سطح پایین
-----------	-----------

ش

Similarity	شباهت
Cosine similarity	شباهت کسینوسی

شکل	Shape
-----	-------

ض

ضرب استهلاک	Damping Factor
-------------	----------------

ط

طول اقلیدسی	Euclidean length
-------------	------------------

ع

عنوان تصویر	Image title
-------------	-------------

عنوان صفحه	Page title
------------	------------

غ

غنی تر شدن اغنیا	Rich-get-richer
------------------	-----------------

ف

فاصله کانونی	Focal Lengths
--------------	---------------

فراخوانی	Recall
----------	--------

فراداده	Metadata
---------	----------

فضای برداری	Vector space
-------------	--------------

فضای ویژه	Eigenspace
-----------	------------

ق

Human Judgement	قضاوت انسانی
-----------------	--------------

ک

Hash code	کد دَرهَم
Class	کلاس
Key	کلید
Keyword	کلیدواژه

گ

Similarity Graph	گراف شباهت
Web Graph	گراف وب

م

Link-based	مبتنی بر اتصال
Content-based	مبتنی بر محتوای دیداری
Text-based	مبتنی بر محتوای متنی
Random variable	متغیر تصادفی
Associated Text	متن اختصاص یافته
Image alt	متن جایگزین تصویر
Documents Collection	مجموعه اسناد

Benchmark Dataset	مجموعه داده محک
Popularity	محبوبیت
Repository	مخزن
Binary Independence Model (BIM)	مدل دودویی وابستگی
Language Modeling (LM)	مدل زبانی
Vector Space Model(VSM)	مدل فضای برداری
Centrality	مرکزیت
Evaluation Metrics	معیارهای ارزیابی
Semantic Concepts	مفاهیم معنایی
Value	مقدار
Scale	مقیاس
Scale	مقیاس
Search engine	موتور جستجو
Mean	میانگین
MAP	میانگین دقت متوسط

ن

Unstable	ناپایدار
Normalization	نرمال سازی
Interest Points	نقاط موردعلاقه
Feature Points	نقاط ویژگی
Index	نمایه
Inverted index	نمایه معکوس

Indexer	نمایه‌ساز
Information need	نیاز اطلاعاتی

و

Variance	واریانس
Lexicon	واژگان
Global Features	ویژگی‌های سراسری
Local Features	ویژگی‌های ناحیه‌ای

ی

Learning	یادگیری
----------	---------

واژه‌نامه انگلیسی به فارسی

A

Associated Text	متن اختصاص یافته
Authority	اعتبار
Average Precision	دقت متوسط

B

Benchmark Dataset	مجموعه داده محک
Binary	دودویی
Binary Independence Model (BIM)	مدل دودویی وابستگی
Boolean	بولی

C

Centrality	مرکزیت
Challenges	چالش‌ها
Class	کلاس
Combinational	ترکیبی
Content-based	مبتنی بر محتوای دیداری
Correctly-ranked	به صورت صحیح رتبه‌بندی شده
Cosine similarity	شباهت کسینوسی
Crawler	خزشگر

D

Damping Factor	ضریب استهلاک
Diffusion	پخش‌شدگی
Dispersion	پراکندگی
Distributed	توزیع شده
Documents	اسناد
Documents Collection	مجموعه اسناد
Domain	دامنه
Domain Specific	خاص منظوره

E

Eigenspace	فضای ویژه
Euclidean length	طول اقلیدسی
Evaluation	ارزیابی
Evaluation Metrics	معیارهای ارزیابی
Explicit feedback	بازخورد مستقیم
External	خارجی

F

Feature Points	نقاط ویژگی
Feedback	بازخورد
Flowchart	روندما
Focal Lengths	فاصله کانونی

G

Global Features	ویژگی‌های سراسری
Grayscale	خاکستری

H

Harris Corner Detection	تشخیص گوشه هریس
Hash code	کد دَرهم
Hash table	جدول دَرهم
High quality	باکیفیت
Histogram	بافت‌ما
Human Judgement	قضاوت انسانی
Hyperlink	اِبرپیوند

I

Image alt	متن جایگزین تصویر
Image Descriptor	توصیفگرهای تصویر
Image Retrieval	بازیابی تصویر

Image title	عنوان تصویر
Implicit feedback	بازخورد غیرمستقیم
Incorrectly-ranked	به صورت نادرست رتبه‌بندی شده
Index	نمایه
Indexer	نمایه‌ساز
Information need	نیاز اطلاعاتی
Interest Points	نقاط موردعلاقه
Inverted index	نمایه معکوس

K

Key	کلید
Keyword	کلیدواژه

L

Language Modeling (LM)	مدل زبانی
Learning	یادگیری
Lexicon	واژگان
Link-based	مبتنی بر اتصال
Local Features	ویژگی‌های ناحیه‌ای
Low Level	سطح پایین

M

MAP	میانگین دقت متوسط
Matching	تطبیق
Mean	میانگین
Metadata	فراداده

N

Normalization	نرمال‌سازی
---------------	------------

O

Offline	برون خط
---------	---------

Online

برخط

P

Page title	عنوان صفحه
Pairwise	جفتی
Parser	تجزیه‌گر
Periodically	دوره‌ای
Perspectives	دورنمای سه‌بعدی
Popularity	محبوبیت
Precision	دقت
Preprocessing	پیش‌پردازش
Probabilistic	احتمالی
Probabilistic	احتمالی
Probabilistic Ranking Principle (PRP)	رتبه‌بندی احتمالی

Q

Query	پرس‌وجو
-------	---------

R

Random variable	متغیر تصادفی
Ranking	رتبه‌بندی
Recall	فراخوانی
Regional traffic ranks	رتبه‌بندی ترافیک جهانی
Relevancy	ارتباط
Repository	مخزن
Rerank	رتبه‌بندی مجدد
Retrieval	بازیابی
Rich-get-richer	غنی‌تر شدن اغنیا

S

Scale	مقیاس
-------	-------

Scale	مقیاس
Scale Invariant Feature Transform(SIFT)	تبدیل ویژگی مقیاس ناپسته
Score	امتیاز
Search engine	موتور جستجو
Semantic Concepts	مفاهیم معنایی
Shape	شکل
Shape Context	زمینه شکل
Similarity	شباهت
Similarity Graph	گراف شباهت
Spin Images	تصاویر اسپین

T

Term Frequency (TF)	تواتر کلمه
Text-based	مبتنی بر محتوای متنی
Texture	بافت
Threshold	حد آستانه
Transformations	تبدیل‌ها
Tuning	تنظیم

U

Unstable	ناپایدار
User behavior	رفتار کاربر

V

Value	مقدار
Variance	واریانس
Vector space	فضای برداری
Vector Space Model(VSM)	مدل فضای برداری
Visual Hyperlink	ابریوند دیداری

W

Web Graph	گراف وب
-----------	---------

Weighting function

تابع وزن دهی

شاخص

پیکسل.....۸, ۳۴, ۳۵, ۳۶

۱

اتصال.....۳۱, ۵۰, ۵۸, ۶۳
احتمالی.....۶, ۱۵, ۱۷, ۵۳, ۵۶, ۶۴
ارزیابی.....۱۱, ۱۲
اسناد.....۳, ۴, ۶, ۹, ۱۱, ۱۲, ۱۳, ۱۴, ۱۵, ۱۶, ۱۷, ۱۸, ۱۹, ۳۲, ۳۷, ۴۰, ۴۲, ۴۴, ۴۹, ۵۳, ۵۸, ۶۱
اعتبار.....۷, ۲۳, ۲۴, ۲۹, ۵۰, ۵۳, ۶۱
الگوریتم‌های ترکیبی.....۷, ۳۰
امتیاز.....۵, ۶, ۷, ۱۵, ۲۳, ۲۹, ۳۰, ۴۳, ۴۵, ۵۰, ۵۳, ۶۵

آ

آزمون.....۴۴

ب

بازخورد.....۷, ۸, ۲۹, ۳۰, ۵۳, ۶۲, ۶۳
بازیابی.....۱, ۴, ۵, ۶, ۷, ۱۰, ۱۲, ۱۳, ۱۴, ۱۹, ۲۲, ۲۹, ۴۲, ۴۹, ۵۳, ۶۲, ۶۴
بافت.....۲۰
بافت‌نما.....۲۰
برخط.....۵, ۳۱, ۵۰, ۵۳, ۶۴
برون خط.....۵, ۳۱, ۳۲, ۵۴, ۶۳

پ

پخش‌شدگی.....۵۲, ۵۴, ۶۱
پراکندگی.....۵۲, ۵۴, ۶۱
پرس‌وجو.....۴, ۷, ۹, ۱۰, ۱۱, ۱۲, ۱۳, ۱۴, ۱۵, ۱۶, ۱۷, ۱۸, ۱۹, ۳۲, ۴۳, ۴۴, ۴۹, ۵۰, ۵۴, ۶۴
پرس‌وجوی ۱, ۴, ۵, ۶, ۷, ۹, ۱۹, ۲۲, ۳۱, ۴۲, ۴۳, ۴۵, ۴۹, ۵۰
پیش‌پردازش.....۳, ۵۴, ۶۴

ت

تبدیل ویژگی مقیاس نابسته.....۲۱
تجزیه‌گر.....۳, ۵۴, ۶۴
ترافیک منطقه‌ای.....۱
تشخیص گوشه هریس.....۲۱
تصاویر اسپین.....۲۱
تصویر.....۱, ۲, ۳, ۵, ۶, ۷, ۸, ۹, ۱۰, ۱۱, ۱۲, ۱۳, ۱۴, ۱۵, ۱۹, ۲۰, ۲۱, ۲۲, ۲۳, ۲۴, ۲۹, ۳۰, ۳۱, ۳۳, ۳۴, ۳۵, ۳۶, ۳۷, ۳۸, ۳۹, ۴۰, ۴۱, ۴۲, ۴۵, ۴۹, ۵۰, ۵۳, ۵۵, ۵۷, ۵۸, ۶۲, ۶۳
تطبیق.....۹
تنظیم وزن.....۵۰
توزیع احتمال.....۵۲
توصیفگر.....۸

ج

جاوا.....۴۲

چ

چالش.....۸
چرخش‌یافته.....۵۱

خ

خارجی.....۳۹, ۴۱, ۵۵, ۶۲
خاص‌منظوره.....۱
خزش.....۲, ۷۰

د

داده‌های آزمون.....۴۴, ۴۶, ۵۰
درجه پیچیدگی.....۸, ۳۱, ۳۳, ۴۹

دقت ۱۲, ۱۳, ۱۴, ۳۰
دوره‌ای ۵, ۵۶, ۶۴
دیجیتال ۱, ۳۴, ۴۹

ر

رتبه‌بندی ۱, ۲, ۴, ۵, ۶, ۷, ۹, ۱۰, ۱۱, ۱۳, ۱۴, ۱۵, ۱۹, ۲۱, ۲۳, ۲۴, ۲۹, ۳۰, ۳۱, ۴۲, ۴۴, ۴۹, ۵۰, ۵۶, ۶۴
رفتار کاربر ۸, ۲۹

س

سند ۳, ۴, ۵, ۱۳, ۱۵, ۱۶, ۱۷, ۱۸, ۱۹, ۳۲, ۳۳, ۴۰, ۴۲, ۴۵

ش

شباهت ۶, ۸, ۹, ۱۴, ۲۰, ۲۱, ۲۲, ۲۳
شکل ۱۱, ۱۹, ۲۰, ۲۱

ف

فاز بر خط ۳۱, ۳۲, ۴۹
فراخوانی ۱۲
فضای برداری ۶, ۱۵, ۵۷, ۵۹, ۶۵
فیلد ۴۰, ۴۲, ۴۵, ۵۰

ک

کلیدواژه ۱, ۱۰, ۴۹, ۵۸, ۶۳

گ

گراف شباهت ۶, ۱۹, ۲۲
گوگل ۱, ۱۰

م

مبتنی بر اتصال ۶, ۷, ۱۴, ۲۳, ۲۴
مبتنی بر رفتار کاربر ۶, ۷, ۱۴, ۲۹

مبتنی بر محتوای دیداری ۶, ۸, ۱۴, ۱۹, ۲۱
مبتنی بر محتوای متنی ۶, ۸, ۱۴
مبتنی بر محتوا ۱۵, ۱۷, ۱۸
متغیر تصادفی ۵۲, ۵۸, ۶۴
مجموعه داده محک ۱۱, ۴۳, ۴۵, ۵۰, ۵۹, ۶۱
مجموعه محک ۱۱
محبوبیت ۲, ۲۳, ۵۹, ۶۴
مخزن ۲, ۱۱, ۵۹, ۶۴
مدل فضای برداری ۱۵
مدل کلمات ۴
مرکزیت ۵۲, ۵۹, ۶۱
معماری ۲, ۳
معیارهای ارزیابی ۱۱
مفاهیم معنایی ۸, ۵۹, ۶۵
مقیاس شده ۳۲, ۳۳, ۳۷, ۴۲, ۴۹, ۵۱
موتور جستجو .. ۱, ۴, ۵, ۱۰, ۲۳, ۲۹, ۳۱, ۴۳, ۴۹, ۵۹, ۶۵
موتورهای جستجو ۹, ۱۲
میانگین ۱۳, ۲۲, ۳۵, ۳۶, ۵۲, ۵۹, ۶۳

ن

نرمال ۴, ۲۲
نرمال‌سازی ۱۶
نمایه ۳, ۹, ۴۲, ۵۹, ۶۳
نمایه معکوس ۳, ۵۹, ۶۳
نمایه‌ساز ۳, ۴, ۵, ۶۰, ۶۳
نمایه‌سازی ۳, ۴, ۳۱, ۳۷, ۴۲, ۵۰
نیاز اطلاعاتی ۱, ۶۰, ۶۳

و

واریانس ۳۶, ۵۲, ۶۰, ۶۵
وب ۷, ۸, ۱۱, ۱۴, ۲۲, ۲۳
ویژگی‌های سراسری ۲۰, ۲۱

ی

یادگیری ۷, ۴۴, ۴۵, ۴۶, ۵۰, ۶۰, ۶۳

یاهو.....۱

یادگیری جفتی.....۴۴, ۴۵, ۵۰

مراجع

1. Top Sites. Available from: <http://www.alex.com/topsites/global>, October 2013.
2. Top Sites in Iran. Available from: <http://www.alex.com/topsites/countries/IR>, October 2013.
3. Google.com's Regional Traffic Ranks. Available from: <http://www.alex.com/siteinfo/google.com>, October 2013.
4. Yahoo.com's Regional Traffic Ranks. Available from: <http://www.alex.com/siteinfo/yahoo.com>, October 2013.

۵. زارع بیدکی، ع.م. "رتبه بندی و خزش مؤثر در وب". رساله دکترا، انتشارات دانشگاه تهران،

۱۳۸۸

6. Baeza-Yates, R. "**Challenges in the interaction of information retrieval and natural language processing**". In Proceedings of 5th international conference on Computational Linguistics and Intelligent Text Processing (CICLing), Lecture Notes in Computer Science Springer, Vol. 2945, pp. 445–456., February 2004.
7. Gong, Zhiguo, and Chan Wa Cheang. "**Web image indexing by using associated texts**". Knowledge and information systems 10.2: p. 243-264, 2006.
8. Müller, H. and H. SO, "**Text-based (image) retrieval**". HES SO//Valais, Sierre, Switzerland [Online] http://thomas.deselaers.de/teaching/files/tutorial_icpr08/03_textBasedRetrieval. p df [Accessed 25 July 2010], 2007.

9. Salton, G. **"The SMART retrieval system - experiments in automatic document processing"**. Prentice-Hall, 1971.
10. Robertson, S., and Walker, S. **"Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval"**. In Proceedings of SIGIR, pp. 232-241, 1994.
11. Jing, Y. and S. Baluja. **"Pagerank for product image search"**. ACM, 2008.
12. Ahmed, Gulfishan Firdose, and Raju Barskar. **"A Study on Different Image Retrieval Techniques in Image Processing"**. International Journal of Computing and Engineering (IJSCE) ISSN: p. 2232-2307, 2011.
13. Lempel, R. and A. Soffer. **"PicASHOW: Pictorial authority search by hyperlinks on the web"**. ACM, 2001.
14. Po Leung, A. and P. Auer. **"An efficient search algorithm for content-based image retrieval with user feedback"**. IEEE, 2008.
15. Vasconcelos, N. and A. Lippman, **"Learning from user feedback in image retrieval systems"**. Advances in neural information processing systems. 12: p. 977-983, 1999.
16. Chen, Zheng, Liu Wenyin, Feng Zhang, Mingjing Li, and Hongjiang Zhang. **"Web mining for web image retrieval"**. Journal of the American Society for Information Science and Technology 52, no. p. 831-839, 2001.

17. Tollari, Sabrina, Marcin Detyniecki, Christophe Marsala, Ali Fakeri-Tabrizi, Massih-Reza Amini, and Patrick Gallinari. **"Exploiting visual concepts to improve text-based image retrieval"**. In *Advances in Information Retrieval*, pp. 701-705. Springer Berlin Heidelberg, 2009.
18. Kherfi, M., D. Ziou, and A. Bernardi. **"What is Behind Image Retrieval from the World Wide Web?"**, 2004.
19. Manning, C.D., P. Raghavan, and H. Schütze, **"Introduction to information retrieval"**. Vol. 1. Cambridge University Press Cambridge, 2008.
20. Salton, Gerard. **"The SMART retrieval system—experiments in automatic document processing"**. 1971.
21. Sparck Jones, K., S. Walker, and S.E. Robertson, **"A probabilistic model of information retrieval: Development and comparative experiments"**: Part 1. *Information Processing & Management*. 36(6): p. 779-808, 2000.
22. Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. **"Introduction to information retrieval"**. Vol. 1. Cambridge: Cambridge University Press, 2008.
23. Salton, Gerard, and Christopher Buckley. **"Term-weighting approaches in automatic text retrieval"**. *Information processing & management* 24.5: p. 513-523, 1988.

24. R. Baeza-Yates and B. Ribeiro-Neto. "**Modern Information Retrieval**". Addison-Wesley, 1999.
25. Sparck Jones, Karen, Steve Walker, and Stephen E. Robertson. "**A probabilistic model of information retrieval: development and comparative experiments**": Part 1. Information Processing & Management 36.6: p. 779-808, 2000.
26. Robertson, Stephen E., and Steve Walker. "**Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval**". Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Springer-Verlag New York, Inc. 1994.
27. Okapi at TREC-3. Robertson, S. E., Walker, S., Jones, S., M.Hancock-Beaulieu, M., and Gatford, M. 1995, In Harman, D. K., editor, "**The Third Text REtrieval Conference**", pp. 109-126, 1995.
28. Najork, Marc A., Hugo Zaragoza, and Michael J. Taylor. "**HITS on the Web: How does it Compare?**". Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007.
29. Robertson, Stephen E., and Steve Walker. "**Microsoft Cambridge at TREC-9: Filtering Track**". TREC. Vol. 9. 2000..
30. Harris, C. and M. Stephens. "**A combined corner and edge detector**". Manchester, UK, 1988.

31. Lowe, D.G., "**Distinctive image features from scale-invariant keypoints**".
International journal of computer vision. 60(2): p. 91-110, 2004.
32. Belongie, S., J. Malik, and J. Puzicha, "**Shape matching and object recognition using shape contexts**". Pattern Analysis and Machine Intelligence, IEEE Transactions on. 24(4): p. 509-522, 2002.
33. Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "**A sparse texture representation using affine-invariant regions**". Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 2. IEEE, 2003.
34. Jing, Y. and S. Baluja, "**Visualrank: Applying pagerank to large-scale image search**". Pattern Analysis and Machine Intelligence, IEEE Transactions on. 30(11): p. 1877-1890, 2008.
35. Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. "**The PageRank citation ranking: bringing order to the web**". 1999.
36. Kleinberg, J.M., "**Authoritative sources in a hyperlinked environment**". Journal of the ACM (JACM). 46(5): p. 604-632, 1999.
37. Xue, Gui-Rong, Qiang Yang, Hua-Jun Zeng, Yong Yu, and Zheng Chen. "**Exploiting the hierarchical structure for link analysis**". In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 186-193. ACM, 2005.

38. Zareh Bidoki, A.M. and N. Yazdani, DistanceRank: "**An intelligent ranking algorithm for web pages**". Information Processing & Management. 44(2): p. 877-892, 2008.
39. La Cascia, Marco, Saratendu Sethi, and Stan Sclaroff. "**Combining textual and visual cues for content-based image retrieval on the world wide web**". Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on. IEEE, 1998.
40. Wang, Bin, Zhiwei Li, Mingjing Li, and Wei-Ying Ma. "**Large-scale duplicate detection for web image search**". In *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 353-356. IEEE, 2006.
41. Sculley, D. "**Combined regression and ranking**". Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010.