

الحمد لله
البرحمين
م



دانشکده کامپیوتر و فناوری اطلاعات

گروه هوش مصنوعی

بهبود نتایج حاصل از موتورهای جستجو با شبیه‌سازی رفتار هوشمند یک انسان خبره

دانشجو: مرضیه رحیمی

استاد راهنما:

دکتر مرتضی زاهدی

استاد مشاور:

دکتر حمید حسن‌پور

پایان‌نامه ارشد جهت اخذ درجه‌ی کارشناسی ارشد

تیر ۱۳۹۰

دانشگاه صنعتی شاهرود

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

گروه هوش مصنوعی

پایان نامه کارشناسی ارشد آقای / خانم

تحت عنوان:

در تاریخ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مورد ارزیابی و با درجه مورد پذیرش قرار گرفت.

امضاء	اساتید مشاور	امضاء	اساتید راهنما
	نام و نام خانوادگی :		نام و نام خانوادگی :
	نام و نام خانوادگی :		نام و نام خانوادگی :

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی :		نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :
			نام و نام خانوادگی :

به نام خدا

در اینجا لازم می‌دانم که پس از سپاس از خداوند بزرگ به خاطر تمام آنچه به من عطا فرموده و پدر و مادر عزیزم که راهبانی باو پشتیبانی ایشان همیشه باعث دلگرمی و آرامش

بوده است، از استاد راهبانی عزیز و بزرگوارم جناب آقای دکتر زاهدی به دلیل راهبانی‌های بی‌دین و زحمات و سوزنازشان در تمامی دوران تحصیلم در دانشگاه صنعتی

شاهرود و نیز از استاد محترم مشاورم در این پایان‌نامه آقای دکتر حسن پورو همچنین استاد دکترم آقای دکتر پویان به دلیل زحمات و راهبانی‌هایشان کمال تشکر و

سپاس را داشته باشم. همچنین از خواهر خوبم به خاطر همه‌ی همراهی‌هایش تشکر می‌کنم.

تعهد نامه

اینجانب دانشجوی دوره کارشناسی ارشد رشته دانشکده

..... دانشگاه صنعتی شاهرود نویسنده پایان نامه

..... تحت راهنمایی متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است .

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده

امروزه یکی از پرکاربردترین و محبوب‌ترین منابع برای کسب اطلاعات تار جهان‌گستر وب است. وب مجموعه بسیار بزرگی از انواع مختلف داده‌های بدون ساختار است که دائماً در حال تغییر هستند. موتورهای جستجو به عنوان ابزاری برای دسترسی سریع و آسان به اطلاعات مورد نیاز برای کاربران بسیار مورد توجه قرار گرفته‌اند. گسترش روز افزون داده‌ها در وب و توسعه استفاده از آن لزوم بهبود موتورهای جستجو را به گونه‌ای که بتوانند نتایجی را بازگرداند که انطباق بیشتری با اطلاعات مورد نیاز کاربران را داشته باشند، دو چندان می‌کند. هرچه موتور جستجو بتواند با بازگردان نتایج دقیق‌تر رضایت کاربران را بیشتر جلب کند، در رقابت با سایرین موفق‌تر خواهد بود. یکی از دلایلی که در هنگام جستجو در وب باعث بازگرداندن نتایج نه چندان مناسب و نارضایتی کاربران می‌شود درخواست¹ ناکامل و نامناسب است که به دو دلیل ممکن است توسط کاربران صادر شود. یکی اینکه کاربر در جستجوی وب بی‌تجربه باشد و درک دقیقی از روند جستجو نداشته باشد و دیگری اینکه با کلمات کلیدی مناسب در حیطه مورد جستجو آشنا نباشد. یکی از راه‌حلهایی که برای حل این مشکل مطرح می‌شود توسعه درخواست است. رفتار کاربران متخصص، چه در حوزه جستجو در وب و چه در حوزه اطلاعاتی مورد جستجو می‌تواند راهنمای مناسبی برای اثبات مفید بودن توسعه درخواست باشد. در این پایان‌نامه با بهره‌گیری از چنین اطلاعاتی یک روش جدید برای توسعه درخواست پیشنهاد شده است که ترکیبی از دو تکنیک بازخورد ربط² و Latent Semantic Analysis است که هر دو انطباق زیادی با رفتار انسانی دارند و در ضمن ترکیب آنها می‌تواند بخشی از ضعف‌های هر دو را ضمن بهره‌گیری از قدرتهایشان بپوشاند. روش ارائه شده بر اساس یک مجموعه آزمون از درخواستها که توسط کاربران شرکت‌کننده ارائه شده است و با قضاوت آنان در یک محیط واقعی یعنی

¹ - Query

² - Relevance Feedback

موتور جستجوی گوگل مورد ارزیابی قرار گرفته است. نتایج این ارزیابی نشان می‌دهد که درخواستهای تولید شده توسط روش پیشنهادی نتایج را به میزان قابل توجهی بهبود می‌بخشد. روش پیشنهادی همچنین با روش Rocchio که یک استاندارد برای بازخورد ربط محسوب می‌شود مقایسه شده و طبق نتایج روش پیشنهادی به مراتب بهتر از Rocchio عمل می‌کند.

کلمات کلیدی: توسعه‌ی درخواست^۳، سیستمهای بازیابی اطلاعات، موتورهای جستجو، latent Semantic Analysis، بازخورد ربط، خوشه‌ی عنوان^۴، خوشه‌بندی، رفتار کاربران وب

³ - Query Expansion

⁴ - Topic Cluster

1. M. Zahedi, M. Rahimi, Query Expansion based on Contextual Meanings of the Query Terms, International Journal of Computer Science Issues, Vol. 8, Issue 5, Aug 2011.

فهرست مطالب

۲	۱- مقدمه
۳	۱-۱- روشهای مبتنی بر اصطلاحنامه
۶	۲-۱- روشهای مبتنی بر تکنیک بازخورد ربط
۷	۱-۲-۱- روش Rocchio
۹	۲-۲-۱- بهره‌گیری از الگوریتمهای تکاملی
۱۴	۲- مدلهای ارزیابی اطلاعات
۱۶	۱-۲- مدل استاندارد بولی
۱۷	۲-۲- مدل فضای برداری
۱۹	۳-۲- مدل توسعه‌یافته‌ی بولی
۲۰	۴-۲- موتورهای جستجو
۲۲	۳- معرفی Latent Semantic Analysis
۲۶	۱-۳- توصیف ریاضی LSA
۲۶	۱-۱-۳- تجزیه‌ی مقادیر تکین
۳۲	۴- تحلیل رفتار کاربران در جستجوی وب
۴۱	۵- معرفی روش پیشنهادی
۴۱	۱-۵- توصیف روش پیشنهادی
۴۲	۱-۱-۵- تعداد اسناد مرتبط
۴۳	۲-۱-۵- داده‌ی ورودی

- ۴۴ ۳-۱-۵- روش تشکیل خوشه‌ها
- ۴۵ ۱-۳-۱-۵- خوشه‌بندی سلسله‌مراتبی
- ۴۶ ۲-۳-۱-۵- خوشه‌بندی K-means
- ۴۷ ۴-۱-۵- انتخاب نهایی کلمات
- ۴۷ ۵-۱-۵- تعداد کلمات
- ۴۸ ۲-۵- ارزیابی
- ۴۹ ۱-۲-۵- معیارهای مبتنی بر رنگ
- ۵۳ ۳-۵- معنی‌داری آماری
- ۵۴ ۴-۵- مجموعه تست
- ۵۵ ۵-۵- بررسی یک مشکل
- ۵۹ ۶-۵- تعداد صفحات مورد ارزیابی در نتیجه‌ی جستجو
- ۶۱ ۶- آزمایشات و نتایج
- ۶۸ ۷- نتیجه‌گیری و کارهای آینده
- ۸- پیوستها
- ۷۳ ۱-۷- پیوست ۱
- ۷۴ ۲-۷- پیوست ۲

فهرست شکلها

- شکل ۱-۱: تقسیم‌بندی روشهای توسعه‌ی درخواست ۳
- شکل ۲-۱: شمایی از بخشی از یک اصطلاحنامه ۴
- شکل ۳-۱: روند کلی بازخورد ربط ۶
- شکل ۴-۱: روش Rocchio ۸
- شکل ۱-۲: ماتریس T-D ۱۸
- شکل ۲-۳: ماتریسهای حاصل از تجزیه‌ی مقادیر تکین ۲۷
- شکل ۳-۳: تولید ماتریس رنک k ۲۹
- شکل ۴-۳: تجزیه‌ی ماتریس T-D با استفاده از تجزیه‌ی مقادیر تکین ۲۹
- شکل ۱-۴: رفتار کاربران ماهر در اینترنت از دید کلی ۳۲
- شکل ۲-۴: مدل جزئی رفتار کاربران متخصص در جستجوی وب ۳۴
- شکل ۳-۴: مقایسه‌ی فضای اسناد و درخواستها ۳۶
- شکل ۱-۵: شکستن هر سند به پنجره‌های همپوشان ۴۴
- شکل ۲-۵: رابطه‌ی precision و recall ۵۰
- شکل ۳-۵: نمونه‌ای از منحنی recall-precision ۵۲

فهرست جداول

- جدول ۵-۱: حالات ممکن اسناد در یک سیستم بازیابی اطلاعات ۴۹
- جدول ۵-۲: نمونه‌ای از محاسبه‌ی مقادیر تخمینی برای رسم منحنی recall-precision ۵۱
- جدول ۵-۳: مثالی از محاسبات مربوط به معیار MAP ۵۲
- جدول ۵-۴: موضوعات مورد سوال ۵۵
- جدول ۵-۵: نمونه‌ای از درخواستهای مطرح شده ۵۶
- جدول ۵-۶: مثالی از محاسبه‌ی Bpref ۵۷
- جدول ۵-۷: مثالی دیگر از محاسبه‌ی Bpref ۵۸
- جدول ۶-۱: درخواست حاصل از روش Rocchio برای چند نمونه درخواست اولیه ۶۲
- جدول ۶-۲: درخواست حاصل از روش LSA برای چند نمونه درخواست اولیه ۶۲
- جدول ۶-۳: درخواست حاصل از روش LSA - k-means برای چند نمونه درخواست اولیه ۶۲
- جدول ۶-۴: درخواست حاصل از روش LSA - hierarchical clustering برای چند نمونه درخواست اولیه ۶۳
- جدول ۶-۵: درخواست حاصل از روش Rocchio برای درخواست اولیه‌ی rose hybridization ۶۴
- جدول ۶-۶: درخواست حاصل از روش Rocchio برای درخواست اولیه‌ی rose hybridization ۶۵
- جدول ۶-۷: میانگین Bpref روی مجموعه‌ی تست برای هریک از روشها ۶۵
- جدول ۷-۱: برتری روشها نسبت به یکدیگر ۶۹

- جدول ۸-۱: محاسبه‌ی Bpref برای query ذکرشده در حالت اولیه ۷۴
- جدول ۸-۲: محاسبه‌ی Bpref برای query ذکرشده و روش Rocchio ۷۵
- جدول ۸-۳: محاسبه‌ی Bpref برای query ذکرشده و روش LSA ۷۶
- جدول ۸-۴: محاسبه‌ی Bpref برای query ذکرشده و روش LSA – K-means ۷۷
- جدول ۸-۵: محاسبه‌ی Bpref برای query ذکرشده و روش LSA- hierarchical ۷۸
- جدول ۸-۶: محاسبه‌ی Bpref برای query ذکرشده در حالت اولیه ۷۹
- جدول ۸-۷: محاسبه‌ی Bpref برای query ذکرشده و روش Rocchio ۸۰
- جدول ۸-۸: محاسبه‌ی Bpref برای query ذکرشده و روش LSA ۸۱
- جدول ۸-۹: محاسبه‌ی Bpref برای query ذکرشده و روش LSA – K-means ۸۲
- جدول ۸-۱۰: محاسبه‌ی Bpref برای query ذکرشده و روش LSA- hierarchical ۸۳
- جدول ۸-۱۱: محاسبه‌ی Bpref برای query ذکرشده در حالت اولیه ۸۴
- جدول ۸-۱۲: محاسبه‌ی Bpref برای query ذکرشده و روش Rocchio ۸۵
- جدول ۸-۱۳: محاسبه‌ی Bpref برای query ذکرشده و روش LSA ۸۶
- جدول ۸-۱۴: محاسبه‌ی Bpref برای query ذکرشده و روش LSA – K-means ۸۷
- جدول ۸-۱۵: محاسبه‌ی Bpref برای query ذکرشده و روش LSA- hierarchical ۸۸

فصل ۱

مقدمه

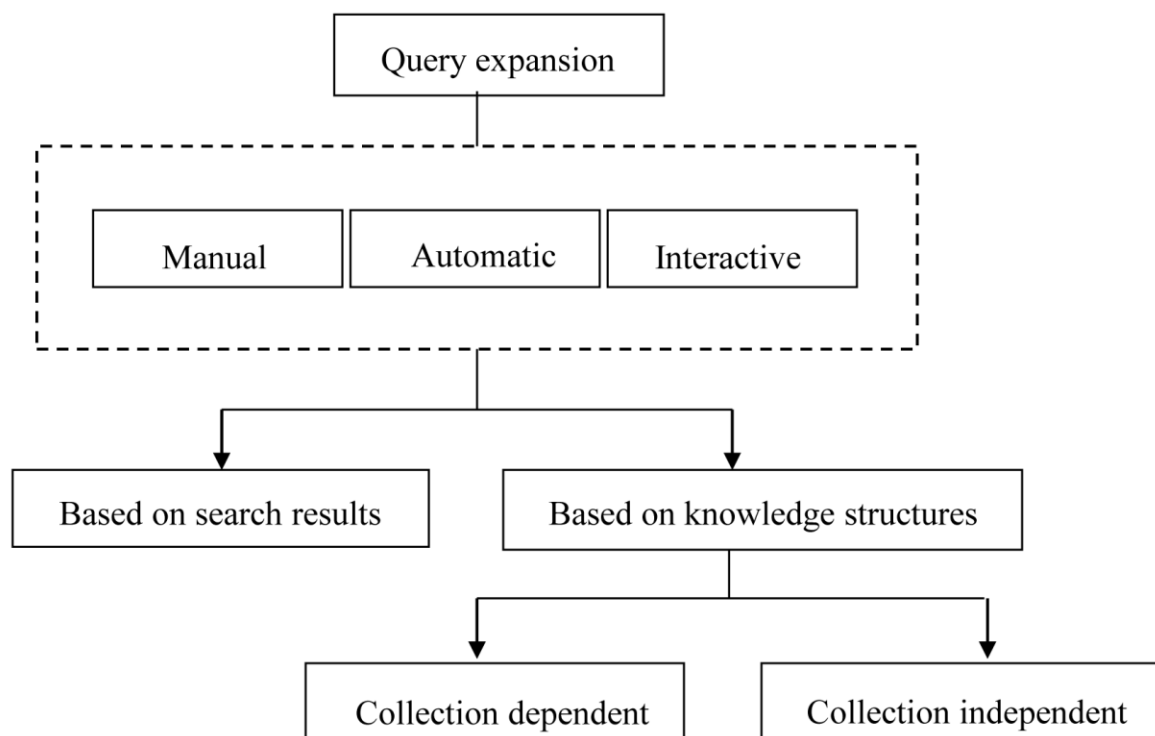
امروزه وب یکی از اساسی‌ترین منابع کسب اطلاعات برای مردم جهان محسوب می‌شود. این پدیده‌ی بسیار گسترده و در حال تغییر، مجموعه‌ای از داده‌های متنوع و بدون ساختار با تعداد ابعاد بالا است که با سرعت فزاینده‌ای در حال گسترش است. این ویژگیها باعث می‌شوند که با توسعه‌ی وب، یافتن اطلاعات مورد نیاز در آن سخت‌تر و سخت‌تر شود و به همین دلیل موتورهای جستجو به عنوان ابزارهایی ساده و قابل دسترس برای جستجوی اطلاعات در وب مورد توجه قرار گیرند.

در بسیاری از موتورهای جستجوی امروزی جستجو بر مبنای کلمات انجام می‌گیرد. به این ترتیب که کاربر بر اساس اطلاعات مورد نیاز خود، کلماتی را به عنوان درخواست وارد می‌کند و جستجو بر مبنای این کلمات صورت می‌پذیرد. به این ترتیب روشن است که دقت نتایج حاصل، بیش از هرچیز دیگری وابسته به درخواست کاربر است. از طرف دیگر بسیاری از کاربران در وب به دلیل عدم آشنایی با موضوع مورد جستجو و یا عدم آشنایی با فرایند جستجو در وب قادر به صدور درخواستهایی که بتوانند توصیف دقیقی از اطلاعات مورد نیاز آنها ارائه دهند، نیستند. ابهام در کلمات نیز مشکل را دوچندان می‌کند.

یکی از روشهایی که برای حل این مشکل مورد توجه است، توسعه‌ی درخواست است. به این معنی که درخواست اولیه‌ی کاربر را با افزودن کلمات متناسب با اطلاعات مورد نیاز کاربر بهبود بخشیم. روشهای گوناگونی تاکنون برای توسعه‌ی درخواست پیشنهاد شده‌اند که در ادامه برخی از رایج‌ترین آنها را معرفی خواهیم نمود.

به طور کلی روشهای توسعه‌ی درخواست را طبق مقاله‌ی [۱] می‌توان به صورت شکل (۱-۱) تقسیم‌بندی کرد. نوعی از این روشها بر مبنای نتایج جستجو هستند که مربوط می‌شوند به روش بازخورد ربط [۲] و گروه دیگر، روشهای مبتنی بر سازه‌های دانشی هستند که مستقل از نتایج جستجو می‌باشند. روشهای نوع

دوم را در وب می‌توان به طور عمده روشهای مبتنی بر اصطلاحنامه دانست که همهی انواع ساخت اصطلاحنامه‌ها؛ دستی یا اتوماتیک را دربرمی‌گیرد.



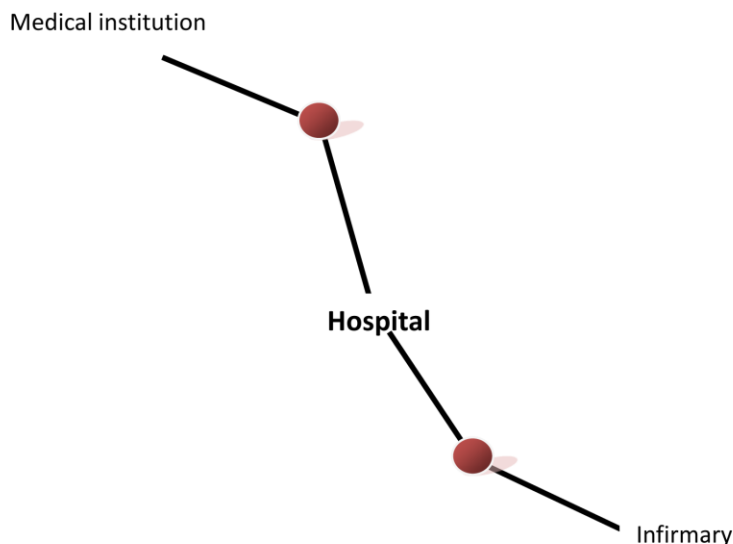
شکل ۱-۱: تقسیم‌بندی روشهای توسعه‌ی درخواست [۱]

چنین تقسیم‌بندی‌ای در [۳] نیز آمده است. در این کتاب روشهای توسعه‌ی درخواست به دو دسته‌ی محلی و سراسری تقسیم شده‌اند که دسته‌ی اول شامل روشهای بازخورد ربط و دسته‌ی دوم شامل روشهای مبتنی بر اصطلاحنامه هستند.

۱-۴ روشهای مبتنی بر اصطلاحنامه

یک اصطلاحنامه مجموعه‌ای مرجع است که در آن کلمات مرتبط یا هم معنی در یک گروه قرار گرفته‌اند.

به عبارت دیگر مجموعه‌ای کنترل شده از کلمات گزیده شده و نظام یافته است که بین آنها روابط معنایی و رده‌ای، یا سلسله‌مراتبی برقرار است. شمایی از یک اصطلاحنامه در شکل (۱-۲) آمده است.



شکل ۱-۲: شمایی از بخشی از یک اصطلاحنامه

اولین به‌کارگیری اصطلاحنامه‌ها در حوزه‌ی کاربردهای کامپیوتری به اوایل دهه‌ی ۱۹۵۰ بازمی‌گردد که یک اصطلاحنامه برای نمایه‌سازی در یک حوزه‌ی خاص در IBM طراحی شد. دیری نگذشت که روشهای مشابهی در شرکتهای دیگر نیز به کار گرفته شد و استفاده از چنین اصطلاحنامه‌هایی که در آن زمان به صورت دست‌ساز تولید می‌شدند به سرعت گسترش یافت.

در حوزه‌ی توسعه‌ی درخواست استفاده از اصطلاحنامه‌ها به این صورت است که پس از ورود درخواست توسط کاربر، واژه‌های درخواست، مورد بررسی قرار می‌گیرند و واژه‌های مترادف آنها براساس وزنشان انتخاب و به درخواست افزوده می‌شوند [۴ و ۵].

مسئله مهم در مورد اصطلاحنامه چگونگی انتخاب کلمات مناسب از بین واژه‌های همگروه کلمات درخواست است. این مسئله را وزن کلمات مشخص می‌کند. برای تعیین وزن کلمات، روشهای مختلفی

مورد استفاده قرار می‌گیرند که یکی از رایجترین آنها محاسبه‌ی میزان هم‌رخدادی کلمات است. مثلاً در مقاله‌ی [۶] وزن کلمات را در یک حیطه‌ی خاص، مقدار هم‌رخدادی آنها در آن حیطه تعیین می‌کند. هم‌رخدادی اشاره دارد به فراوانی رخداد همزمان کلمات در مجموعه‌ای از متون. این مفهوم بیانگر ارتباط معنایی کلمات است. ماتریس هم‌رخداد ماتریسی است که حاوی این مقدار برای هر دو کلمه از مجموعه واژگان می‌باشد.

در مقاله‌ی [۷] از چندین نوع اصطلاحنامه استفاده شده و برای تعیین وزن کلمات از جمع‌بندی وزن آنها در همه‌ی این اصطلاحنامه‌ها استفاده می‌گردد. نتایج بدست آمده نشان داده‌اند که این روش در مقایسه با استفاده از فقط یک نوع اصطلاحنامه، کارایی بیشتری دارد. در مقاله‌ی [۸] سعی شده است تا از رخداد همزمان کلمات در اسناد یکسان برای محاسبه‌ی شباهت، استفاده شده و یک اصطلاحنامه بر این مبنا به صورت اتوماتیک ساخته شود.

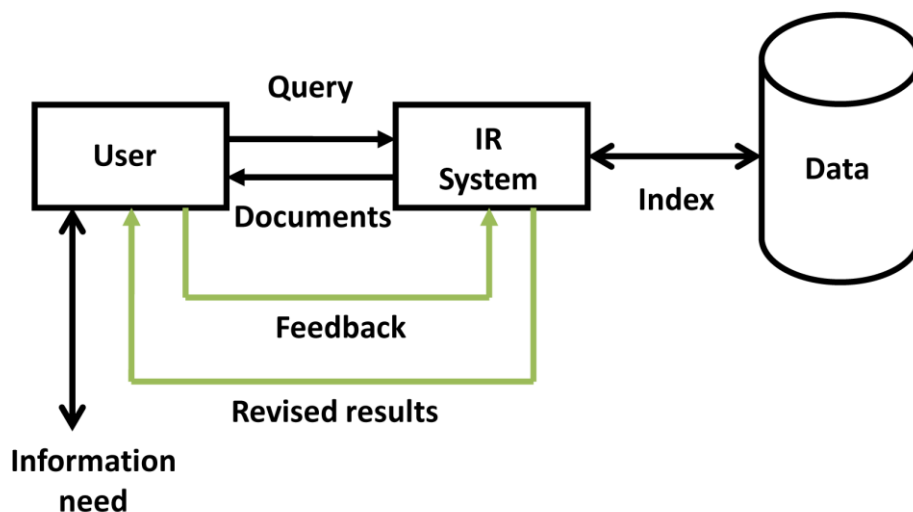
اما گذشته از تفاوت‌های وزن‌دهی و ساختاری به طور کلی، ساخت، نگهداری و به‌روزرسانی اصطلاحنامه‌ها بسیار زمانبر و پرهزینه است. به همین دلیل ساخت اتوماتیک اصطلاحنامه‌ها مورد توجه قرار گرفته است. یکی از روشهای خاص وب در این زمینه استفاده از گزارشات مربوط به درخواستهای کاربران^۱ است که نیازمند پردازش حجم وسیعی از داده‌ها است که خود با مسائل زیادی روبرو است. به طور کلی گفته می‌شود که روشهای مبتنی بر اصطلاحنامه در افزایش recall موفق عمل می‌کنند ولی می‌توانند precision را به شدت کاهش دهند به خصوص در زمانی که کلمات دارای ابهام باشند. مقادیر precision و recall، معیارهایی برای ارزیابی میزان ارتباط نتایج بازیابی شده با درخواست کاربر در سیستمهای بازیابی اطلاعات هستند که در بخش ۵-۲ تشریح خواهند شد. به این ترتیب با توجه به این

¹ - User query logs

مسئله و همچنین پرهزینه بودن ساخت و نگهداری، معمولاً از اصطلاحنامه‌ها در حیطه‌های خاص و مجموعه‌های محدودتر داده‌ای استفاده می‌شود.

۱-۲- روشهای مبتنی بر بازخورد ربط:

ایده‌ی اصلی در بازخورد ربط کمک گرفتن از کاربر در فرایند بازیابی اطلاعات است. به این نحو که نتایج اولیه را بر اساس قضاوت کاربر بهبود می‌بخشیم. فرایند بازخورد ربط را به طور کلی می‌توان به صورت زیر بیان کرد. این روند می‌تواند یک بار یا بیشتر تکرار شود چنانکه در شکل (۱ - ۳) نیز نشان داده شده است.



شکل ۱ - ۳: روند کلی بازخورد ربط

- کاربر درخواست اولیه‌ی خود را بر اساس اطلاعات مورد نیازش صادر می‌کند.
- نتایج اولیه بر اساس این درخواست بازیابی می‌شوند.

- نتایج حاصل توسط کاربر قضاوت شده و برخی به عنوان مرتبط و یا نامرتب انتخاب می گردند.
- بر اساس این انتخابها درخواست به گونه ای تغییر می یابد که توصیف دقیقتری از اطلاعات مورد نیاز کاربر باشد که با استفاده از اسناد انتخاب شده مشخص می گردند.
- نتایج نهایی بر اساس این درخواست جدید بازیابی می شوند.

۱-۲-۱- روش Rocchio

تاکنون روشهای مختلفی برای پیاده سازی بازخورد ربط پیشنهاد شده اند. الگوریتمی که در بسیاری موارد به عنوان یک استاندارد از آن یاد شده و برای ارزیابی سایر روشها مورد مقایسه قرار می گیرد، الگوریتم Rocchio است که در سال ۱۹۷۱ پیشنهاد شد و در سیستم SMART به کار گرفته شد. در این الگوریتم درخواست بهبود یافته به صورت فرمول (۱ - ۱) تعریف می شود [۹].

$$q_m = \alpha q_0 + \beta \frac{1}{|D_r|} \sum_{d_j \in D_r} d_j - \gamma \frac{1}{|D_{nr}|} \sum_{d_j \in D_{nr}} d_j \quad (1 - 1)$$

در این رابطه، q_0 همان درخواست اولیه کاربر، D_r و D_{nr} سندهای مرتبط و نامرتب انتخاب شده توسط کاربر و d_j نماینده یک سند بازیابی شده است. α ، β ، و γ نیز ثابتهای هستند که نقش وزن را برای هر یک از جملات رابطه بازی می کنند. به طور کلی سندهای مرتبط از اهمیت و تاثیر بیشتری برخوردارند و بنابراین، معولا β بزرگتر از γ است. در بسیاری از موارد γ برابر صفر است یعنی کافی است کاربران فقط اسناد مرتبط را مشخص کنند. به این ترتیب در ادامه ی بحث، فرض بر این است که γ برابر صفر باشد. در تئوری، بازخورد ربط قادر است هم Precision و هم Recall را افزایش دهد، اما در عمل ثابت شده است در مواردی که بهبود Recall مورد توجه باشد کارآمدتر است [۳].