



شبکه واژگان و کاربرد آن در گسترش پرس و جو

سارا خوش وطن^۱، وحیده رشادت^۲، محمدرضا فیضی درخشی^۳

^۱ دانشکده کامپیوتر، دانشگاه آزاد اسلامی واحد شبستر، شبستر، ایران
sarakhosvatan@yahoo.com

^۲ دانشکده کامپیوتر، دانشگاه آزاد اسلامی واحد شبستر، شبستر، ایران
vreshadat@yahoo.com

^۳ دانشکده کامپیوتر، دانشگاه تبریز، تبریز، ایران
mfeizi@tabrizu.ac.ir

چکیده

وجود انبوهی از صفحات متنی در وب جهان گستر موجب شده است تا نتایجی که موتورهای جستجو در اختیار کاربران قرار می‌دهند حاوی مقدار زیادی اطلاعات نامربوط باشد. این مساله به شدت کارایی موتورهای جستجو را زیر سوال برده است. برای حل این مشکل، موتورهای جستجو از ابزارهای مختلفی برای بهبود کارایی بازیابی اطلاعات استفاده می‌کنند. گسترش پرسش کاربر نمونه‌ای از این تلاش‌ها است. منابع مختلفی وجود دارند که می‌توان از آنها در گسترش پرسش استفاده کرد. از جمله این منابع هستان-شناسی شبکه واژگان است. روش‌های مبتنی بر هستان‌شناسی از روابط معنایی بین مفاهیم برای گسترش پرسش استفاده می‌کند. در این مقاله به اختصار با مفهوم هستان‌شناسی و نمونه‌ای از آن، یعنی هستان‌شناسی شبکه واژگان آشنا خواهیم شد. سپس کاربرد آن را در گسترش پرسش با بررسی چند روشی که از آن استفاده کرده‌اند بیان خواهیم کرد.

کلمات کلیدی: بازیابی اطلاعات، گسترش پرس و جو، شبکه واژگان

عملیات بازیابی به طور کارا تر انجام شود. یکی از روش‌های افزایش کارایی بازیابی گسترش پرس و جو است. برای گسترش پرس و جوی کاربران روش‌های مختلفی بیان شد که برای این منظور کلمات مناسب به منظور رسیدن به نتایج مرتبط و افزایش کارایی بازیابی، برای افزوده شدن به پرسش انتخاب می‌شوند. یکی از این روش‌ها هستان-شناسی است. روش‌های مبتنی بر هستان‌شناسی از روابط معنایی بین کلمات برای انتخاب کلمات بسط مناسب استفاده می‌کنند. در این مقاله به اختصار چالش‌های مربوط به گسترش بازیابی اطلاعات در وب بیان شده و در ادامه نیز انواع روش گسترش پرسش مبتنی بر شبکه واژگان که یکی از پرکاربردترین هستان‌شناسی‌ها است، بررسی خواهد شد.

۱- مقدمه

امروزه وب به‌عنوان بزرگ‌ترین منبع داده‌ها است که به سهولت قابل دسترسی است. از این رو طبیعی است که اطلاعات از آن استخراج شود و در این شرایط موتورهای جستجوی وب به یکی از پراستفاده‌ترین ابزارها در اینترنت تبدیل شده‌اند. اما رشد تصاعدی و آهنگ سریع تغییر و تحول در وب، بازیابی تمام اطلاعات مربوط را واقعا مشکل می‌سازد. در واقع، کاوش در وب شاید تنگنای اصلی موتورهای جستجوی وب است.

بازیابی اطلاعات عبارت است از تکنیک‌هایی جهت یافتن اسناد مرتبط با هر درخواست کاربر، به طوری که مرتبط‌ترین سند به درخواست کاربر در رتبه بالاتری در لیست خروجی قرار گیرد. در سیستم‌های بازیابی اطلاعات، اسناد به گونه‌ای پیش‌پردازش و بازنمایی می‌شوند که

۲- مفاهیم پایه هستان‌شناسی

قیدها دسته‌بندی کردند. سپس کلمه‌های هر دسته را به گونه‌ای گروه‌بندی کردند که کلمه‌های مترادف و هم‌معنی در یک گروه قرار گرفتند. هر کدام از این گروه‌ها را یک مجموعه معنی^۲ می‌گویند که جزء اصلی تشکیل‌دهنده‌ی شبکه واژگان است. هر مجموعه‌ی معنی، معنی متفاوتی از کلمه را نشان می‌دهد. به عنوان مثال، مجموعه معنی "bicycle"، دو کلمه‌ی *bicycle* و *bike* هستند. یعنی هر دو کلمه به مفهوم *bicycle* اشاره می‌کنند. هر مجموعه‌ی معنی در شبکه واژگان دارای تعریف^۳ است. این تعریف شامل توضیح‌های کوتاهی در مورد معنی آن و مثال‌هایی از کاربرد آن است [3][11]. بین مجموعه‌های معنی با توجه به نوع آنها روابط معنایی مختلفی وجود دارد [1][8][11]. نسخه ۳،۰ جدیدترین نسخه شبکه واژگان است که در سال ۲۰۰۶ به روز شده است. این نسخه شامل ۱۵۵۲۸۷ کلمه است که در ۱۱۷۶۵۹ مجموعه معنی گروه‌بندی شده‌اند [11].

۳-۱- روابط موجود در شبکه واژگان

بین مجموعه‌های معنی در شبکه واژگان روابط معنایی مختلفی وجود دارد. این روابط با توجه به نوع مجموعه‌های معنی انواع مختلفی دارد. دو رابطه مهم در شبکه واژگان که به طور وسیعی در گسترش پرس-وجو مورد استفاده قرار می‌گیرند روابط *is-a* و *part-of* هستند. ابرمعنایی^۴: به پدر یک رابطه *is-a*، ابرمعنا گفته می‌شود. مثلاً *vehicle* در رابطه "vehicle is-a car". زیرمعنایی^۵: به فرزند یک رابطه *is-a*، زیرمعنا گفته می‌شود. مثلاً *car* در رابطه "vehicle is-a car". کل^۶: به پدر یک رابطه *part-of*، کل گفته می‌شود. مثلاً *car* در رابطه "car has-an engine".

جزء^۷: به فرزند یک رابطه *part-of*، جزء گفته می‌شود. مثلاً *engine* در رابطه "car has-an engine".

شبکه واژگان بین انواع گوناگون رابطه *part-of* تمایز قائل می‌شود. بنابراین امکان ایجاد ساختار سلسله مراتبی را فراهم می‌کند. انواع گوناگون رابطه *part-of* که در شبکه واژگان وجود دارند عبارتند از:

مفهوم عضویت: *professor is-a-member-of staff*
 مفهوم اجزاء تشکیل‌دهنده: *tears are-made-of water*
 مفاهیم دیگر: *China is-part-of Asia*

یک کلمه ممکن است چندین معنی داشته باشد بنابراین در چندین مجموعه معنی ظاهر خواهد شد. شبکه واژگان برای هر کلمه

مفاهیم یک دامنه، خصوصیات و روابط بین آنها است. هستان‌شناسی‌ها به صورت دستی و یا خودکار ایجاد می‌شوند و در زمینه‌های مختلفی از جمله بازیابی اطلاعات مورد استفاده قرار می‌گیرند. هستان‌شناسی روش ساخت‌یافته‌ای را برای نمایش دانش فراهم می‌کند. در واقع می‌توان آن را نوعی گراف در نظر گرفت که مفاهیم، گره‌های آن و روابط بین مفاهیم، یال‌های آن هستند. هر مفهوم دارای توصیف متنی است که آن را تعریف می‌کند. در هستان‌شناسی ممکن است چند کلمه به یک مفهوم اشاره کنند و یا حتی این امکان وجود دارد که یک کلمه به بیش از یک مفهوم اشاره کند [3]. مثلاً برای توصیف "bicycle" می‌توان از کلمه‌های *bike* و *bicycle* استفاده کرد. در مقابل، کلمه *bike* هم به مفهوم *bicycle* و هم به مفهوم *motorcycle* اشاره می‌کند [3].

هستان‌شناسی‌ها را بر مبنای نوع اطلاعات آنها به دو دسته تقسیم می‌کنند: یک گروه هستان‌شناسی‌هایی هستند که برای یک دامنه خاص طراحی و ساخته شده‌اند. این هستان‌شناسی‌ها کلمه‌ها را بر مبنای دامنه‌ای که برای آن ساخته شده‌اند معنا و تفسیر می‌کنند. دسته‌ی دوم هستان‌شناسی‌هایی هستند که متعلق به دامنه خاصی نیستند و به آنها هستان‌شناسی‌های عمومی می‌گویند. محدوده‌ی دانشی که آنها ارائه می‌کنند وسیع‌تر از هستان‌شناسی‌های دسته‌ی اول است. هستان‌شناسی‌های *OpenCyc*، *GFO*، *Dublin Core*، *DOLCE*، *SUMO*، شبکه مفهومی و شبکه واژگان از این نوع هستند [10].

۳-۱-۲- روابط بین مفاهیم در هستان‌شناسی

در هستان‌شناسی‌های مختلف روابط گوناگونی بین مفاهیم وجود دارد. مانند *similar*، *see-also*، *pertains* و *domain*. اما دو رابطه‌ای که در اغلب هستان‌شناسی‌ها وجود دارند و به طور وسیعی در گسترش پرس‌وجو مورد استفاده قرار می‌گیرند عبارتند از: *is-a* و *part-of*. این دو رابطه از اهمیت زیادی برخوردار هستند و امکان ایجاد ساختار سلسله مراتبی را فراهم می‌کنند [3].

۳- شبکه واژگان

شبکه واژگان پایگاه‌داده‌ای از لغت‌های انگلیسی است که ساخت آن توسط آقای میلر^۱ و همکارانش به صورت دستی در آزمایشگاه علوم-شناختی دانشگاه پرینستون آمریکا از سال ۱۹۸۲ شروع شد. آنها کلمه‌های زبان انگلیسی را در چهار دسته اسم‌ها، فعل‌ها، صفت‌ها و

² Synset

³ Gloss

⁴ Hypernyms

⁵ Hyponyms

⁶ Holonyms

⁷ Meronyms

¹ George Miller

زمانی اتفاق می‌افتد که پرس‌وجو مبهم باشد. مثلاً پرس‌وجوی "apple" مبهم است. زیرا apple می‌تواند هم به معنی نوعی میوه و هم به معنی نام شرکت به کار رود. زمانی که کاربر این پرس‌وجو را مطرح می‌کند ممکن است سیستم آن را در جهت "apple as fruit" گسترش دهد در حالی که منظور کاربر "apple computer" است. یکی دیگر از مشکلات گسترش پرس‌وجو outweighing است. outweighing حالت خاصی از انحراف پرس‌وجو است که در آن کلمه‌های گسترش به کل کلمه‌های پرس‌وجو مرتبط نیستند و فقط به کلمه منفردی از آن مرتبطند. برای غلبه بر این مشکلات راه‌حل‌های مختلفی پیشنهاد شده است [3].

۵- نمونه‌هایی از گسترش پرسش به استفاده از شبکه واژگان

الن ورهی^{۱۳} جزء اولین کسانی بود که در سال ۱۹۹۳ از هستان‌شناسی در گسترش پرس‌وجو استفاده کرد. او در مقاله خود الگوی کلی برای گسترش پرس‌وجو با استفاده از هستان‌شناسی ارائه داد که از دو مرحله تشکیل شده است:

- در مرحله اول کلمه‌های موجود در پرس‌وجو رفع ابهام می‌شوند به طوری که هر کلمه در پرس‌وجو منطبق با یک مفهوم در هستان‌شناسی باشد.
- در مرحله دوم کلمه‌های مرتبط با کلمه‌های پرس‌وجو از شبکه واژگان استخراج و به عنوان کلمه‌های گسترش به آن اضافه می‌شوند [3].

ورهی در هر دو روش پیشنهادی خود در سال‌های ۱۹۹۳ و ۱۹۹۴ از مدل فضای برداری گسترش‌یافته فوکس^{۱۴} برای بازیابی اطلاعات استفاده کرده است. این مدل از این ایده ناشی می‌شود که برای هر بردار می‌توان زیربردارهایی در نظر گرفت. هر کدام از این زیربردارها بر مبنای اطلاعاتی که از روی آنها ساخته می‌شوند دارای یک ctype هستند. به عنوان مثال، یک بردار می‌تواند از ریشه‌ی کلمه‌هایی که در شبکه واژگان وجود ندارند و یا از مترادف‌های کلمه‌های رفع ابهام شده و یا از ریشه‌ی کلمه‌های رفع ابهام شده ساخته شود. اگر قرار باشد در یک مجموعه از این مدل استفاده شود ابتدا زیربردارهای بردارهای اسناد و پرس‌وجو بر مبنای موارد بالا و یا موارد دیگر ساخته می‌شوند. سپس میزان مشابهت بین سند و پرس‌وجو با جمع مشابهت‌های زیربردارهای مختلف آنها محاسبه می‌شود. اسنادی که دارای امتیاز

یک شمارنده‌ای در نظر می‌گیرد. این شمارنده تعداد مجموعه‌های معنی که کلمه در آنها ظاهر شده است را نشان می‌دهد. از شبکه واژگان برای رفع ابهام معنی کلمه‌ها^۸، بازیابی اطلاعات، طبقه‌بندی و خلاصه‌سازی خودکار متن‌ها استفاده می‌شود [11].

۴- گسترش پرس‌وجو

گسترش پرس‌وجو فرآیندی است که در آن یک سری کلمه‌های اضافی با هدف بهبود نتایج به پرس‌وجوی کاربر اضافه می‌شوند [3]. یکی از موضوعات کلیدی در گسترش پرس‌وجو توانایی کلمه‌های اضافه شده در متمایز کردن اسناد است. کلمه‌های مناسب برای گسترش کلمه-هایی هستند که بتوانند اسناد مرتبط بیشتری را بازیابی کنند و از بازیابی اسناد غیرمرتبط جلوگیری کنند [7].

گسترش پرس‌وجو از طریق افزودن کلمه‌های وابسته به کلمه‌های پرس‌وجو انجام می‌شود. این کلمه‌ها ممکن است مترادف‌ها^۹، ریشه‌ی کلمه‌های^{۱۰} پرس‌وجو، کلمه‌هایی که با کلمه‌های پرس‌وجو وقوع همزمان دارند و یا به آنها نزدیک هستند باشند. اما این معیارها تنها معیارهای موجود نیستند [4]. یک روش احتمالی که از معیارهای دیگری غیر از ارتباط با کلمه‌های پرس‌وجو استفاده می‌کند ممکن است نسبت به روش‌هایی که برای یافتن ارتباط با کلمه‌های پرس‌وجو تلاش می‌کنند بهتر عمل کند [3].

یکی از اهداف گسترش پرس‌وجو غلبه بر مشکل عدم تطبیق کلمه‌ها^{۱۱} است [2]. این مشکل زمانی به وجود می‌آید که در مجموعه-ی اسناد و پرس‌وجوی کاربر از کلمه‌های متفاوتی برای اشاره به یک مفهوم واحد استفاده شود. مثلاً کاربر پرس‌وجوی «ایانه» را مطرح می‌کند ولی در مجموعه‌ی اسناد از کلمه‌ی «کامپیوتر» استفاده شده است. بنابراین اسنادی که از نظر معنایی با پرس‌وجوی کاربر مرتبط هستند به دلیل اینکه کلمه‌ی پرس‌وجوی اصلی در آنها وجود ندارد بازیابی نمی‌شوند. اگر این مشکل مورد توجه قرار نگیرد باعث کاهش چشمگیری در کارایی بازیابی می‌شود [5]. بنابراین با کمک فرایند گسترش پرس‌وجو می‌توان اسناد مرتبطی که شامل کلمه‌های پرس‌وجو نیستند را نیز بازیابی کرد.

۴-۱- مشکلات گسترش پرس‌وجو

گسترش پرس‌وجو با مشکلات متعددی روبرو است. یکی از مهمترین آنها انحراف پرس‌وجو^{۱۲} است. انحراف پرس‌وجو یعنی پرس‌وجوی کاربر در جهتی بسط داده می‌شود که مورد نظر او نیست. انحراف معمولاً

⁸ Word Sense Disambiguous

⁹ Synonyms

¹⁰ Stemming

¹¹ Word mismatch

¹² Query drifting

¹³ Elen Voorhees

¹⁴ Fox

بنابراین هنگام بازیابی اسناد بدون دانستن معنی دقیق کلمه، سیستم تمام اسنادی که شامل آن کلمه است را بازیابی می‌کند و این عامل منجر به کاهش دقت می‌شود. برای حل این مشکل باید رفع ابهام انجام شود.

در [6] روش جدیدی بر پایه رفع ابهام پیشنهاد شده است. در این روش رفع ابهام با ایجاد شبکه معنایی برای هر معنی کلمه در شبکه واژگان و اشتراک‌گیری و امتیازدهی انجام می‌شود. به این ترتیب که بعد از حذف کلمات توقف، $Q = \{w_1, w_2, \dots, w_n\}$ به عنوان پرسش اولیه در نظر گرفته می‌شود. برای هر کلمه‌ای که در پرسش است مجموعه‌های معنی آن کلمه به صورت زیر از شبکه واژگان استخراج می‌شود:

$$S(w_k) = \{S^k | S^k \in \text{Synset}_{\text{WordNet}}(w_k), w_k \in Q\} \quad (1)$$

w_k -k امین کلمه پرسش و S^k ، j-امین مجموعه معنی مربوط به -k امین کلمه پرسش است و پیکربندی‌های معنایی ممکن برای Q به این صورت در نظر گرفته می‌شود: $C_k = (S_{x_1}^k, S_{x_2}^k, \dots, S_{x_n}^k)$ که x_k اندیسی بین یک و تعداد معنایی ممکن برای w_k است.

برای هر پیکربندی C_k به این صورت عمل می‌شود:

۱. برای هر معنی یک شبکه معنایی ایجاد می‌شود.
۲. بین شبکه‌های معنایی اشتراک‌گیری انجام می‌شود.
۳. برای هر پیکربندی ایجاد شده امتیازی در نظر گرفته می‌شود.

در نهایت پیکربندی با بالاترین امتیاز به عنوان معنی مورد نظر انتخاب می‌شود.

در ایجاد شبکه معنایی بیشتر روابط از شبکه واژگان استخراج می‌شود به جز روابط topic, gloss, domain که از SemCor استخراج می‌شود که یک پیکره خلاصه‌نویسی شده است که در آن در هر جمله معنی هر کلمه از فهرست معنایی موجود در شبکه واژگان استخراج می‌شود. gloss از تعریف مفاهیم موجود در شبکه واژگان ایجاد می‌شود و رابطه domain نیز از برچسب‌های دامنه موجود در شبکه واژگان استخراج می‌شود. شکل (۲-۳) شبکه معنایی کلمه bus را نشان می‌دهد.

بالاتری باشند ارتباط بیشتری با پرس‌وجو داشته و به عنوان نتیجه بازیابی در اختیار کاربر قرار می‌گیرند [8].

ورهی در روش پیشنهادی خود در سال ۱۹۹۳، بعد از رفع ابهام کلمه‌های پرس‌وجو از مدل فضای بردار گسترش‌یافته آقای فوکس برای بازیابی اسناد مرتبط استفاده کرده است. روش رفع ابهام پیشنهادی او بر پایه‌ی این فرضیه است که کلمه‌هایی که با همدیگر در یک متن اتفاق می‌افتند با وجود این که هر کدام به طور جداگانه معنایی متعددی دارند اما با همدیگر معنی مناسبی را به وجود می‌آورند. مثلاً هر کدام از کلمه‌های hit (ضربه زدن)، glove (دستکش)، bat (چوب کریکت) و base (هر کدام از چهار پایگاه بیس‌بال) معنی خاص خود را دارند. اما اگر این چهار کلمه در یک مجموعه قرار بگیرند {base, bat, glove, hit} معنی baseball را به وجود می‌آورند. نتایج تحلیل‌ها نشان می‌دهد که بردارهای مبتنی بر معنی که از فرایند رفع ابهام حاصل می‌شوند نسبت به بردار اصلی بهبودی ندارند و بردارهای اصلی در کل بهتر از بردارهای مبتنی بر معنی هستند. علت این تنزل این است که رفع ابهام پرس‌وجوهای کوتاه دشوار است چون وقتی متن کوتاه است نه تنها فرایند رفع ابهام را انجام نمی‌دهد بلکه ممکن است معنی نادرست را نیز انتخاب کند [8].

ورهی روش دیگری را برای گسترش پرس‌وجو در سال ۱۹۹۴ پیشنهاد کرد. در این روش او برای گسترش پرس‌وجو از روابط لغوی - معنایی موجود در شبکه واژگان نظیر:

- گسترش پرس‌وجو فقط با کلمه‌های مترادف
- گسترش پرس‌وجو با مترادف‌ها و همه‌ی فرزندان در رابطه is-a
- گسترش پرس‌وجو با مترادف‌ها و پدر و همه فرزندان در رابطه is-a
- گسترش پرس‌وجو با مترادف‌ها و هر مجموعه‌ی معنی که مستقیماً به مجموعه‌ی معنی داده شده متصل است.

استفاده کرده است. برای بازیابی اطلاعات، همانند روش قبل از مدل فضای برداری گسترش‌یافته آقای فوکس استفاده کرده است. نتایج آزمایش‌ها نشان می‌دهد که روش‌های مورد استفاده در گسترش پرس‌وجو تاثیر قابل توجهی در کارایی بازیابی ندارند. در نتیجه روابط لغوی - معنایی بهبود قابل توجهی در کارایی بازیابی ایجاد نمی‌کنند. اگر کلمه‌های گسترش وزن بالاتری نسبت به کلمه‌های اصلی بگیرند نتیجه بدتر می‌شود. این روش برای پرس‌وجوهای طولانی مناسب نیست. اما برای پرس‌وجوهای کوتاه و ناقص مناسب است و کارایی پرس‌وجوهای کوتاه را افزایش می‌دهد [9].

یک کلمه بستگی به متنی که در آن قرار دارد ممکن است معنایی متعددی داشته باشد. پرس‌وجو ممکن است شامل کلمه مبهم باشد و

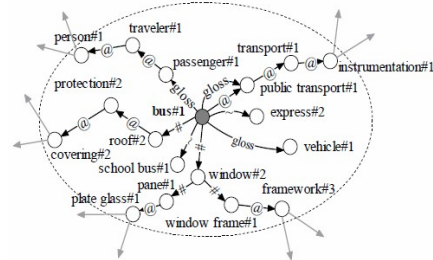
مورد نظر انتخاب می‌شود و سپس از این پیکربندی کلماتی که مجموعه معنی‌شان حداقل یک نود مشترک (یعنی $k=1$) با دیگر مجموعه معنی‌های آن پیکربندی دارد بسط داده می‌شوند که در روش بسط با کلمات تفسیر با وجود اینکه در بعضی موارد نتایج بد بوده ولی در کل بهبود حاصل شده است. در بسط با گره‌های مشترک نتایج بسیار بد بوده و باعث کاهش کارایی شده است. با وجود اینکه آزمایش‌های انجام شده در این مقاله محدود بوده و روی موتور جستجوی گوگل انجام شده است که هر پرسش حداکثر ده کلمه می‌تواند داشته باشد ولی در کل نتایج امید بخش بوده است.

۶- نتیجه‌گیری

در حال حاضر هستان‌شناسی‌های مختلفی وجود دارند که هر کدام با هدف خاصی ایجاد شده‌اند و در زمینه‌های مختلفی مورد استفاده قرار می‌گیرند. در این مقاله سعی کردیم تا دید کلی از هستان‌شناسی ارائه بدهیم. شبکه واژگان به عنوان یک هستان‌شناسی عمومی کاربردهای مختلفی دارد. در این مقاله ضمن معرفی این هستان‌شناسی، یکی از کاربردهای آن را که گسترش پرسش بود معرفی کردیم. آشنایی با هستان‌شناسی‌های مختلفی که وجود دارند و به وجود می‌آیند، اهداف ایجادشان و کاربردهای آنها زمینه‌هایی هستند که جای بسی مطالعه دارند.

مراجع

- [1] رفعتی، ک.، موقر رحیم‌آبادی، ع.، ۱۳۸۶، *گسترش پرسش جستجو به کمک روابط معنایی کلمه‌ها*، سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران، جزیره کیش، خلیج فارس، ایران.
- [2] Abdelali, A., Cowie, J., Soleiman, H., 2007, *Improving query precision using semantic expansion*, Information Processing and Management: an International Journal, Volume 43 Issue 3, pp. 705-716.
- [3] Andreou, A., 2005, *Ontologies and Query expansion*, School of Informatics, University of Edinburgh.
- [4] Hazra, I., Aditi, Sh., 2009, *Thesaurus and query expansion*, International Journal of Computer science & Information Technology (IJCSIT), Vol 1, No 2, November 2009.
- [5] Jiang, Z., Yu, Z., 2010, *A New Technology of Query Expansion Based on New User Interest Model*, IEEE.
- [6] Navigli, R., Velardi, P., 2003, *An analysis of ontology-based query expansion strategies*, workshop on adaptive text extraction and mining (ATEM 2003), In 14th European conference on machine learning (ECML 2003), September 22–26.
- [7] Song, M., Song, I., Allen, R., 2007, *Integration of association rules and ontologies for semantic query expansion*, Published in: Journal Date & Knowledge Engineering, Volume 63, Issue 1.
- [8] Voorhees, E., 1993, *Using Word-Net to disambiguate word senses for text retrieval*, Proceedings of the 16th Annual



شکل (۱): شبکه معنایی برای کلمه bus [6]

بعد از رفع ابهام به کمک شبکه معنایی^{۱۵} و انتخاب پیکربندی‌های مناسب، برای گسترش پرسش علاوه بر روابط ترادف و پدر بودن از تعریف معنی کلمه نیز استفاده شده است. به طور کلی برای گسترش پرسش از پنج روش استفاده شده است. که این پنج روش به این شرح است:

- گسترش با مجموعه معنی: کلمات بسط از مجموعه‌های معنی بدست آمده از الگوریتم رفع ابهام پیشنهادی انتخاب می‌شوند.
- گسترش با رابطه پدری^{۱۶}: کلمات بسط از مجموعه معنی پدری انتخاب می‌شوند که در رابطه مستقیم با مجموعه معنی کلمه مورد نظر در شبکه واژگان قرار دارند.
- گسترش با مجموعه معنی تفسیر^{۱۷}: کلمات بسط از مجموعه معنی‌های کلمات موجود در تفسیرشان انتخاب می‌شوند.
- گسترش با کلمات تفسیر: کلمات بسط از کلمات موجود در تفسیرشان انتخاب می‌شوند.
- گسترش با گره‌های مشترک: کلمات بسط از گره‌های مشترک پیکربندی مورد نظر انتخاب می‌شوند.

به طور کلی دواستراتژی برای گسترش پرسش در نظر گرفته شده است. گسترش پرسش با استفاده از کلمات تک‌معنایی موجود در پرسش تأثیر مهمی در بهبود بازایی اطلاعات دارد. پس در روش اول صرفاً کلمات تک‌معنایی موجود در پرسش بسط داده می‌شوند. در روش دوم مجموعه معنی کلمات با توجه به الگوریتم رفع ابهام پیشنهادی از شبکه واژگان انتخاب می‌شود. این مجموعه معنی‌ها که حداقل k نود مشترک با مجموعه معنی‌های دیگر آن پیکربندی دارند برای بسط در نظر گرفته می‌شود. در روش اول، در همه نتایج بهبود حاصل می‌شود مخصوصاً بسط با کلمات تفسیر، بهبود ۲۶،۸۸٪ به وجود می‌آورد و در روش دوم، با توجه به روش رفع ابهام، پیکربندی

¹⁵ Semantic Network

¹⁶ Hyperonymy

¹⁷ gloss



نهمین همایش منطقه ای کامپیوتر و سیستم‌های هوشمند
باشگاه پژوهشگران جوان واحد علوم و تحقیقات آذربایجان شرقی
اسفند ۱۳۹۰



international ACM SIGIR Conference on Research and Development in information Retrieval (SIGIR '93), Pittsburgh, New York: ACM Press, pp. 171-180.

[9] Voorhees, E., 1994, *Query expansion using lexical-semantic relations*, Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in information Retrieval, Dublin, Ireland, New York:Springer-Verlag, pp. 61-69.

[10] <http://en.wikipedia.org/wiki/Ontology>

[11] <http://en.wikipedia.org/wiki/Wordnet>