سال دوم، شماره ۱ (پاییز ۱۳۸۳)

مدل‌های بازیابی اطلاعات

- تأثیر مدل‌های بازیابی اطلاعات بر میزان ربط
- ربط در مدل سنتی و مدل‌های تعاملی بازیابی اطلاعات
- بررسی اجمالی مزایا و معایب مدل‌های پاتخورهٔ ربط

اطلاع‌شناسی
INFORMOLOGY

چکیده

ربط یکی از مباحث مهم نظام‌های ذخیره و بازیابی است. هرچند که ربط امری نسبی به نظر می‌رسد، تلاش برای افزایش ربط یعنی تطابق بین پرسش کاربر و پاسخ نظام‌های بازیابی همچنان ادامه دارد. متخصصان این حوزه برای ارتقای میزان ربط پاسخ نظام‌های ذخیره و بازیابی، مدل‌هایی را طراحی و ارائه داده‌اند. این مدل‌ها به دو گروه عمدهٔ سنتی و مدرن تقسیم می‌شوند که هر گروه خود چندین مدل را دربر دارد. پیروی از هرکدام از مدل‌ها مزایا و محدودیت‌هایی را به دنبال دارد. در مقالهٔ حاضر روند شکل‌گیری و پیشرفت مدل‌های بازیابی و تأثیر هرکدام از آنها بر میزان ربط مورد توجه قرار گرفته و مزایا و معایب هرکدام از آنها تشریح شده است. سپس با استنتاج کلی از مدل‌های بازیابی و طرز عملکرد آنها، فصل مشترک و پیشبردی مدل‌های موجود در راستای ارتقای ربط مورد بررسی قرار گرفته است. در پایان طی نتیجه‌گیری از مطالب، پیشنهادهایی در مورد به‌کارگیری مدل‌های بازیابی در نظام‌های ذخیره و بازیابی با هدف افزایش ربط ارائه گردیده است.

کلیدواژه‌ها: بازیابی اطلاعات. مدل‌های بازیابی. ربط.

تأثیر مدل‌های بازیابی اطلاعات بر میزان ربط

محمد حسن‌زاده

تأثیر مدل‌های بازیابی اطلاعات بر میزان ربط

محمد حسن‌زاده[1]

مقدمه

نظام‌های اطلاعاتی در طول تاریخ پیدایش خود به دنبال این هدف بوده‌اند که میزان انطباق خواستهٔ کاربران و مدارک ارائه شده به عنوان نتایج جست‌وجو را افزایش دهند. این کار به‌طور معمول از طریق تطابق کلید واژه‌های عبارت جست‌وجوی کاربران و کلیدواژه‌های نمایندهٔ مدارک موجود در مخزن مدارک یا پایگاه داده‌های مراکز اطلاع‌رسانی انجام می‌شود. از آنجایی که خواسته‌های واقعی کاربران به‌طور کامل در قالب زبان مصنوعی ماشین و همچنین کلید واژه‌های منفرد و حتی عبارت‌های جست‌وجو تبدیل نمی شود، در نتیجه، طراحان نظام‌های اطلاعاتی و متخصصان حوزهٔ اطلاعات و اطلاع‌رسانی همیشه به دنبال شیوه‌های بهینه برای افزایش تطابق پرسش کاربران با مدارک موجود بوده‌اند. در این راستا، مدل‌های مختلفی برای بازیابی اطلاعات شکل گرفته است که بررسی آنها می‌تواند دانش ما را نسبت به مسئلهٔ ربط و میزان کارآیی هر کدام از مدل‌ها در این زمینه ارتقاء دهد. در این مقاله، مدل‌های بازیابی اطلاعات در قالب دو دسته مدل‌های بازیابی و مدل‌های مروری مورد بررسی قرار گرفته است که هرکدام از آنها چندین مدل خاص را دربر می‌گیرد. در این قسمت، ابتدا به تعریف و بررسی مفهوم مدل‌های بازیابی می‌پردازیم.

۱. دانشجوی دکتری علوم کتابداری و اطلاع‌رسانی دانشگاه فردوسی مشهد
Email: hzadehm@hotmail.com

مدل‌های بازیابی اطلاعات و سیر تحول آنها

مدل بازیابی اطلاعات عبارت از مجموعه‌ای از دیدگاه‌ها یا بازنمون‌های منطقی مربوط به مدارک موجود در یک مجموعه (مثلاً یک پایگاه) و نیازهای اطلاعاتی بیان شدهٔ کاربران است که تحت چارچوبی خاص با یکدیگر ارتباط یافته و در قالب نتایج جست‌وجو ارائه می‌گردند (بایزا ـ ییتس[1] و ریبیریو ـ نتو[2]، ۱۹۹۹، ص ۲۳).

همان‌گونه که از تعریف بالا برمی‌آید، چهار مؤلفهٔ اصلی در هر مدل بازیابی قابل شناسایی است:

۱. مجموعه‌ای از نمادهای منطقی یا بازنمون‌های مربوط به مدارک موجود در مجموعه‌ای که به شکل اصطلاح‌های نمایه‌ای در پایگاه ارائه می‌شود.

۲. مجموعه‌ای از نمادهای منطقی یا بازنمون‌های نیازهای اطلاعاتی کاربران که به‌منزلهٔ عبارت پرسش از سوی آنها به نظام ارائه می‌گردد.

۳. چارچوبی نظری و مفهومی که در قالب آن، بازنمون‌های مدارک، عبارت‌های پرسش و روابط آنها مدل‌سازی می‌گردد.

۴. الگوریتم رتبه‌بندی در پایگاه که مدارک را بر اساس میزان تطابق نتایج بازیابی با عبارت پرسش مرتب می‌سازد.

برای ایجاد یک مدل، ابتدا بازنمون‌های مدارک (اصطلاح‌های نمایه‌ای) و نیاز اطلاعاتی کاربر مورد توجه قرار می‌گیرد. سپس با در دست داشتن این بازنمون‌ها، چارچوب مدل‌سازی آنها مشخص می‌شود؛ که در قالب آن، عمل رتبه‌بندی نتایج بازیابی انجام پذیرد. هر کدام از مدل‌های بازیابی دارای چارچوب خاص خود هستند و از نظریه‌های متفاوتی برای اعمال رتبه‌بندی و ارائهٔ نتایج استفاده می‌کنند. مدل بازیابی اطلاعات بر اساس انجام وظایف بازیابی برای کاربران به دو دستهٔ عمدهٔ (مدل‌های مبتنی بر وارد کردن عبارت جست‌وجو (بازیابی) و مدل‌های مبتنی بر مرور آنچه نظام به‌صورت الفبایی، موضوعی، یا رده‌ای ارائه می‌دهد (مروری) تقسیم می‌شوند. هرکدام از این مدل‌ها خود دارای زیرمجموعه است. سه مدل کلاسیک بازیابی عبارتند از مدل بولی[3]، مدل برداری[4] و مدل احتمالی[5]. در مدل‌های بولی، مدارک و عبارت پرسش به‌عنوان مجموعه‌هایی از واژگان نمایه‌ای[6] در نظر گرفته می‌شوند؛ و به اصطلاح، از نظریهٔ مجموعه‌ها[7] الهام می‌گیرند (گودی‌وادا[8] و دیگران، ۱۹۹۷).

در مدل‌های برداری، مدارک و عبارت‌های پرسش به صورت فضای برداری چند بعدی

1. Baeza-Yates
2. Ribirio-Neto
3. Boolean model
4. Vector space model
5. Probabilistic model
6. Index Terms
7. Set Theory
8. Gudivada

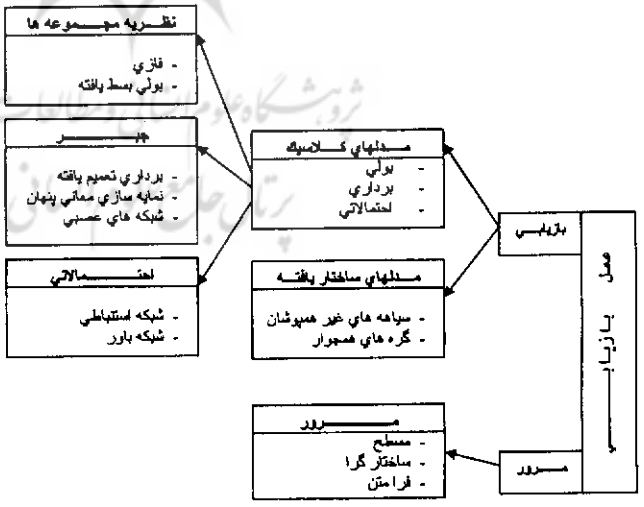
نمایش داده می‌شوند و به اصطلاح، مدل برداری نوعی مدل جبری[1] است. در مدل احتمالی چارچوب مدل‌سازی مدارک و عبارت‌های پرسش بر اساس نظریهٔ احتمالات انجام می‌پذیرد.

در طول زمان، جهت تکمیل کارکرد این مدل‌ها، چارچوب‌های مدل‌سازی جایگزین برای هرکدام از مدل‌های کلاسیک ارائه شده است. مدل‌های فازی[2] و بولی بسط یافته[3] به‌عنوان جایگزینی برای مدل‌های مبتنی بر نظریهٔ مجموعه‌ها ارائه گردید. مدل‌های جبری، با جایگزین‌هایی مانند مدل برداری تعمیم‌یافته[4]، نمایه‌سازی معانی پنهان[5] و شبکهٔ عصبی[6] به تکامل خود ادامه داده‌اند. برای مدل‌های مبتنی بر نظریهٔ احتمالات نیز جایگزین‌هایی مانند مدل‌های مبتنی بر شبکهٔ استنباطی[7] و شبکهٔ باور[8] ارائه گردیده است.

علاوه بر این مدل‌ها، مدل‌های ساختاریافته نیز به عنوان نوع دیگری از مدل‌های بازیابی در عرصهٔ بازیابی اطلاعات ظاهر شدند. این مدل‌ها علاوه بر محتوای متن، ساختار متن را نیز مورد توجه قرار می‌دهند. در این مقاله دو نوع از این مدل‌ها مورد بررسی قرار خواهد گرفت که عبارتند از: مدل‌های سیاهه‌های غیرهمپوشان و مدل گره‌های همجوار.

همان‌گونه که پیش‌تر اشاره شد، مدل‌های مروری نیز یکی از انواع عمده مدل‌های بازیابی است. این مدل‌ها با ظهور وب جهانی اهمیت زیادی یافته‌اند. مدل مروری مسطح، مدل مروری ساختارگرا، و مدل مروری فرامتنی به عنوان مدل‌های مروری رایج بررسی خواهد شد.

برای جمع‌بندی کلی این قسمت از مقاله، تقسیم‌بندی کلی مدل‌ها را در قالب شکل ۱ از کتاب بایزا ییتس و ریبیرو ـ نتو (۱۹۹۹، ص ۲۱) نقل می‌گردد:



شکل ۱. دسته‌بندی مدل‌های بازیابی اطلاعات

1. Algebraic
2. Fuzzy model
3. Extended boolean model
4. Generalized vector model
5. Latent semantics indexing
6. Neural network
7. Inference network
8. Belief network

همان‌گونه که در این شکل نیز دیده می‌شود، هر کدام از مدل‌های بازیابی دارای زیرمجموعه‌هایی هستند که به مرور زمان بسط یافته‌اند. بررسی ویژگی هر کدام از این مدل‌ها هرچند که به صورت کامل در قالب مقاله‌ای واحد مقدور نیست، می‌تواند ارتباط آن را با میزان ربط نتایج بازیابی نشان دهد.

مدل‌های بولی

مدل بولی مدلی بسیار ساده است که بر نظریهٔ مجموعه‌ها و جبر بولی مبتنی است. از آنجایی که مفهوم مجموعه از دیدگاه ریاضیات کاملاً مشخص و روشن است، درک مدل بولی نیز دشوار نیست. علاوه بر آن، عبارات جست‌وجو که به صورت عبارات بولی تشکیل می‌شود مفهوم روشنی دارد. با توجه به همین ویژگی ساده و قابل فهم، مدل بولی در سال‌های گذشته توجه بیشتری را به خود جلب کرده و توسط بسیاری از نظام‌های کتابشناختی تجاری مورد پذیرش و استفاده قرار گرفته است.

در مدل بولی، فرض بر این است که کلیدواژهٔ مورد نظر کاربر می‌تواند دو حالت داشته باشد: الف) در میان مدارک موجود در مجموعه (پایگاه) وجود دارد، ب) در میان مدارک موجود نیست. به عبارت دیگر، از قاعدهٔ ۰ ، ۱ (هست یا نیست) پیروی می‌کند که مبتنی بر منطق دودویی است. هر عبارت پرسش در مدل بولی می‌تواند سه عملگر را تحت پوشش قرار دهد: "و" ، "یا" و "به جز".

همان‌گونه که پیش‌تر ذکر شد، مدل بولی به لحاظ سادگی و قابل فهم بودن، توسط بسیاری از نظام‌های کتابشناختی مورد استفاده قرار گرفته است (این واقعیت در مورد نرم‌افزارهای کتابخانه‌ای فارسی نیز صادق است)؛ اما این مدل معایبی نیز دارد. به اعتقاد بایزا ییتس و ریبرو نتو (۱۹۹۹) مدل بولی دارای ضعف‌های عمده‌ای به شکل زیر است: الف) راهبرد بازیابی این مدل مبتنی بر ملاک تصمیم‌گیری دودویی است (به‌طور مثال، چنین فرض می‌شود که یک مدرک یا مرتبط است یا مرتبط نیست). این امر سبب می‌گردد که هیچ‌گونه رتبه‌بندی انجام نگیرد؛ بنابراین، نتایج بازیابی به صورت مطلوب حاصل نمی‌شود، ب) با اینکه عبارت‌های پرسش بولی به لحاظ صرفی از صراحت کافی برخوردارند، اما همیشه ترجمهٔ یک نیاز اطلاعاتی به یک عبارت پرسش بولی کار ساده‌ای نیست. در واقع، بیشتر کاربران نمی‌توانند پرسش خود را در قالب عبارت بولی بیان کنند و عبارت‌های پرسش که کاربران با استفاده از عملگرهای بولی تشکیل می‌دهند

اغلب بسیار ساده است. علاوه بر آن، می‌توان به معایبی مانند عدم تعیین محل قرارگرفتن واژهٔ مورد نظر در داخل متون بازیابی شده، عدم توجه به جهت عبارت‌ها (عدم تمایز بین تقدم و تأخر واژگان، یعنی رابطهٔ مفاهیم با هم)، عدم توجه به فاصلهٔ بین واژگان عبارت‌های پرسش، و مانند آن اشاره کرد. علی‌رغم این ضعف‌ها، مدل بولی، به‌دلیل سادگی آن، هنوز هم به عنوان یکی از مدل‌های غالب در عرصهٔ بازیابی اطلاعات (به‌ویژه پایگاه‌های تجاری) کاربرد فراوان دارد. با توجه به اینکه ربط، مفهومی نسبی است و نمی‌تواند از قواعد مطلق (بودن یا نبودن؛ یک یا صفر) پیروی کند، مدل‌های جایگزینی برای رفع نقایص مدل بولی طراحی شده است که در اینجا به معرفی و بررسی دو مدل مجموعهٔ فازی و مدل بولی بسط یافته پرداخته می‌شود.

الف. مدل مجموعهٔ فازی (مجموعهٔ نامعین)

این مدل از نظریهٔ مجموعهٔ فازی تبعیت می‌کند. این نظریه نخستین بار در سال ۱۹۹۳، توسط زاده[1] (۱۹۹۳) مطرح شد. این نظریه با مجموعه‌هایی سروکار دارد که مرزهای آنها کاملا مشخص نیست. آنچه در مدل فازی اهمیت دارد، مفهوم عضویت یک عنصر در یک مجموعه است. دامنهٔ عضویت از صفر به معنای عدم عضویت تا ۱ به معنای عضویت کامل، متغیر است. بنابراین، عضویت در مجموعهٔ فازی به هیچ عنوان دودویی نیست بلکه به شکل تدریجی و نسبی است. یعنی میزان عضویت یک عنصر در یک مجموعه می‌تواند درجهٔ خاصی داشته باشد، در حالیکه این عضویت در مدل بولی به صورت دودویی است و یک عنصر نمی‌تواند بیش از یکی از دو حالت (عضویت یا عدم‌عضویت) در یک مجموعه داشته باشد. سه عملگر رایج در مورد مجموعه‌های فازی عبارتند از: ضرب مجموعه‌ها، جمع دو یا چند مجموعهٔ فازی، و تفاضل دو یا چند مجموعهٔ فازی. به عبارت دیگر، در مدل فازی، ابتدا میزان عضویت یک کلیدواژه در یک مدرک تعیین می‌شود، سپس با اعمال عملگرهای سه‌گانه، ارتباط بین چند مجموعه یا مدرک نیز تعیین می‌گردد. این کار به رتبه‌بندی نتایج جست‌وجو می‌انجامد. با تکیه بر همین نظر، مدل فازی شکل‌گیری یک اصطلاحنامه را مورد توجه قرار می‌دهد تا رابطهٔ واژگان با یکدیگر تعریف شود. در نهایت، با استفاده از اصطلاحنامه و تطبیق عبارت پرسش کاربر با اصطلاحنامه و تعیین رابطهٔ واژگان موجود در پرسش کاربر و مدارک موجود در مجموعه، تعداد مدارک مرتبطی که به این روش بازیابی می‌شود بیشتر از مدارکی است که به صورت معمول در مدل بولی بازیابی می‌شد.

1. Zadeh

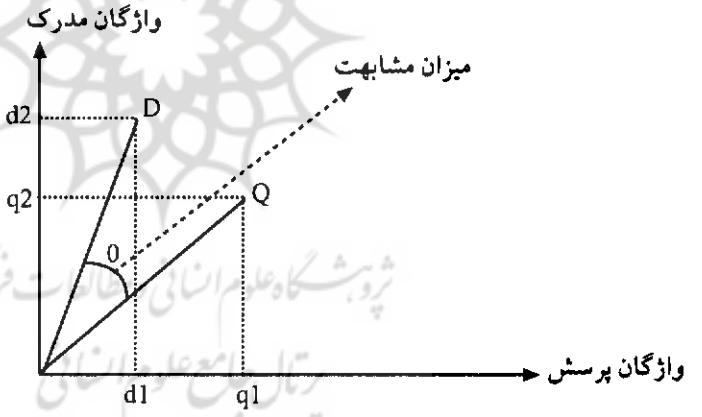
به‌طور کلی می‌توان مدل فازی را چنین توصیف کرد: این مدل با استفاده از یک ماتریس واژه به واژه، همبستگی بین یک مدرک و واژگان نمایه‌ای فازی آن را محاسبه می‌کند. علاوه بر آن، با استفاده از جمع جبری و ضرب جبری، درجهٔ کلی عضویت یک سند در یک مجموعهٔ فازی مربوط به پرسش کاربر را تعیین می‌کند. البته این مدل توان بهره‌گیری از بازخورد کاربر در مورد ربط نتایج بازیابی شده را نیز دارد. مدل‌های مربوط به مجموعهٔ فازی بیشتر در متون مربوط به نظریهٔ فازی مورد بحث قرار گرفته است و در میان متخصصان اطلاع‌رسانی هنوز پذیرش چندانی نیافته است؛ ولی به نظر می‌رسد که با گذشت زمان بر میزان پذیرش آن افزوده شود.

ب. مدل بولی بسط یافته

از آنجایی که در مدل بولی هیچ شرطی برای وزن‌دهی واژگان و رتبه‌بندی مجموعهٔ پرسش وجود ندارد، در نتیجه حجم برونداد حاصل از جست‌وجو یا ناچیز است یا بسیار زیاد. با توجه به همین مشکلات، نظام‌های نوین بازیابی اطلاعات دیگر صرفاً بر پایهٔ مدل بولی شکل نمی‌گیرند. در حقیقت، بیشتر نظام‌های نوین به نوعی یکی از اشکال بازیابی برداری را اساس کار خود قرار می‌دهند. یکی از رویکردهای جایگزین که برای مدل بولی ارائه گردید، گسترش مدل بولی با استفاده از عملگرهای تطبیق نسبی و وزن‌دهی واژگانی بود. این راهبرد، امکان ترکیب فرمول‌بندی پرسش بولی و ویژگی‌های مدل برداری را فراهم می‌آورد. مدل بولی بسط یافته ابتدا توسط سالتون و همکاران (۱۹۸۳) معرفی شد. بر اساس این مدل، یک مدرک که کلیدواژهٔ x یا کلیدواژهٔ y را دربر می‌گیرد می‌تواند به اندازهٔ مدرکی که این کلید واژه‌ها را دربر ندارد نامرتبط باشد. بنابراین، صرف وجود یا عدم‌وجود یک کلیدواژه در یک مدرک نمی‌تواند بیانگر ربط یا عدم‌ربط یک مدرک با یک پرسش باشد. بلکه میزان حضور یک یا چند کلیدواژه و وزن آن در مدرک و همچنین در پرسش کاربر تعیین می‌کند که یک مدرک تا چه اندازه‌ای با یک عبارت پرسش کاربر ارتباط دارد. همین نکتهٔ کلیدی، ایدهٔ اصلی مدل بسط‌یافتهٔ بولی است. ویژگی عمدهٔ این مدل آن است که به‌طور همزمان می‌تواند از ویژگی‌های مدل بولی و برداری استفاده کند؛ و در واقع، مدلی دو رگه است که خصوصیات نظریهٔ مجموعه‌ها و مدل‌های جبری را دربر می‌گیرد. اکنون برخی از نظام‌های بازیابی اطلاعات از این مدل بهره می‌گیرند. بایستی دید که در آینده این مدل تا چه اندازه مقبولیت پیدا می‌کند.

مدل‌های برداری

با توجه به محدودیت‌های اوزان دودویی، لزوم ارائهٔ چارچوبی برای انطباق نسبی، بیش از پیش احساس‌می‌شد (سالتون و لسک[1]، ۱۹۶۸) این کار مستلزم وزن‌دهی غیردودویی به کلیدواژه‌های موجود در عبارت پرسش کاربران و مدارک موجود در مجموعه بود. مدل برداری دقیقاً در راستای همین هدف ارائه گردید. در این مدل، ابتدا کلیدواژه‌های موجود در هر کدام از مجموعه‌های عبارت پرسش و بازنمون مدارک موجود در پایگاه وزن‌دهی می‌شوند؛ سپس این وزن‌ها برای تعیین میزان مشابهت بین عبارت پرسش و مدارک مورد استفاده قرار می‌گیرند. سرانجام، مدارک بازیابی شده بر اساس درجهٔ مشابهت با عبارت پرسش رتبه‌بندی می‌شوند. در این مدل، چنانچه مدارک تا حدودی نیز با عبارت پرسش مشابهت داشته باشند بازیابی می‌شوند. نتایج ارائه شده توسط این مدل بسیار دقیق‌تر از نتایج مدل بولی است. این مدل قادر است با اخذ یک مقدار مشخص به عنوان آستانهٔ مشابهت مورد نظر کاربر، فقط مدارکی را بازیابی کند که مورد نظر کاربر باشد. شکل زیر یک فضای برداری را نشان می‌دهد:



وزن‌دهی در مدل برداری از طریق محاسبه TF (بسامد کلیدواژه‌ها در داخل یک مدرک)[2] صورت می‌پذیرد. این عامل از طریق فرمول تعداد بسامد ظهور واژگان نمایه‌ای تقسیم بر تعداد کل واژگان محاسبه می‌شود. این کار به IDF (تعیین مشابهت بین‌خوشه‌ای)[3] معروف است. برای اینکه درجهٔ تشابه میان مدارک مختلف در کل مجموعه مدارک نیز تعیین شود، مدل برداری از عاملی به نام بسامد معکوس مدرک[4] استفاده می‌کند تا طول مدرک بر نتیجه بازیابی تأثیر نداشته باشد. این کار در مدل برداری "تعیین عدم مشابهت بین‌خوشه‌ای"[5] نامیده می‌شود. به هراندازه که الگوریتم‌های مورد

1. Salton & Lesk
2. Term Frequency (TF)
3. Inter- cluster Similarity
4. Inverse Document Frequency = tf(t) = log (N/n(t))

استفاده برای خوشه‌بندی دارای دقت بیشتری باشند به همان اندازه کنترل این دو عامل (TF, IDF) موجب افزایش تعادل در تشخیص دقیق میزان مشابهت بین کلیدواژه‌های عبارت پرسش و مدارک موجود در مجموعه می‌شود.

مزایای مدل برداری عبارتند از: ۱) وزن‌دهی و ارائه برداری از اوزان مربوط به کلیدواژه‌های عبارت پرسش و مدارک موجود در مجموعه، موجب سادگی در تشخیص میزان تشابه میان این دو مجموعه شده و در نتیجه، عملکرد بازیابی ارتقاء می‌یابد. ۲) راهبرد مبتنی بر تطابق نسبی این اجازه را به نظام بازیابی اطلاعات می‌دهد تا مدارکی را بازیابی کند که برخی از شرایط عبارت پرسش را برآورده می‌سازد. بنابراین، کاربر می‌تواند به بیشترین منابع مرتبط دسترسی داشته باشد و آنچه را که با نیاز خود بیشتر منطبق باشد انتخاب نماید. ۳) رتبه‌بندی مدارک بازیابی شده بر اساس میزان تشابه آنها با عبارت پرسش، به کاربر کمک می‌کند تا در مورد نتایج ارائه شده قضاوت کند.

علاوه بر موارد مذکور، مدل برداری مدلی ساده و سریع است. با توجه به این مزایا، امروزه مدل برداری از مطلوبیت نسبتاً خوبی در عرصهٔ بازیابی اطلاعات برخوردار است. چنین تصور می‌شود که میزان ربط نتایج ارائه شده توسط این مدل بیشتر است. با گذشت زمان جایگزین‌هایی نیز برای مدل برداری ارائه شده است که کارکردهای این مدل را تکمیل کرده است. در این مقاله به سه مدل (برداری تعمیم یافته، نمایه‌سازی معانی پنهان، و شبکهٔ عصبی) پرداخته می‌شود.

الف. مدل فضای برداری تعمیم یافته

طراحان مدل فضای برداری تعمیم یافته (ونگ و زیاکرو[1]، ۱۹۸۵) معتقدند که در مدل‌های بازیابی سنتی (به صورت کلی) و در مدل برداری (به صورت خاص) کلیدواژه‌های نمایه به صورتی مستقل از هم در شکل‌دهی یک بردار مشارکت می‌کنند و مجموعه بردارهای به‌دست آمده از تک تک کلیدواژه‌ها، بردار یک مدرک یا بردار عبارت پرسش را تشکیل می‌دهد؛ و چه بسا، زمانی که دو بردار مربوط به دو کلیدواژه با یکدیگر ترکیب شوند نتیجه صفر باشد. بنابراین، مدل فضای برداری تعمیم‌یافته را پیشنهاد کردند. در این مدل، دو کلیدواژهٔ نمایه می‌تواند غیرمتعامد[2] (غیرعمود بر هم) باشند. به این معنی که بردارهای مربوط به کلیدواژه‌های نمایه به‌عنوان بردارهای متعامد، که اساس فضای برداری را تشکیل می‌دهند، در نظر گرفته نمی‌شوند؛ بلکه خود از اجزای کوچک‌تری تشکیل شده‌اند که ریشه در مجموعهٔ مورد نظر دارند.

1. Wong & Ziakro
2. Non - orthogonal

مدل فضای برداری تعمیم یافته، بر این باور است که ظهور مشترک واژگان نمایه‌ای در داخل مدرک موجود در مجموعه موجب همبستگی و ارتباط بین واژگان کلیدی می‌شود. از آنجایی که این ایده سال‌ها قبل از آنکه مدل فضای برداری تعمیم‌یافته مطرح شود شکل گرفته بود، ایده‌ای نو به شمار نمی‌رود؛ اما به‌کارگیری آن در قالب یک مدل بازیابی، چارچوب نظام‌مندی را در اختیار می‌گذارد که همبستگی میان کلیدواژه‌های نمایه‌ای را به خوبی نمایش می‌دهد. اما با توجه به اینکه این دو مدل در دهۀ ۱۹۸۰ ارائه شده، هنوز به خوبی روشن نیست که آیا در موقعیت‌های علمی نیز برتری‌های قابل توجهی خواهد داشت یا خیر. علاوه بر آن، این مدل بسیار پیچیده تر از مدل برداری است.

ب. مدل نمایه‌سازی معانی پنهان

خلاصه کردن محتوای اطلاعاتی یک مدرک یا عبارت پرسش به مجموعه‌ای از واژگان نمایه‌ای می‌تواند به دو دلیل موجب ضعف نتایج بازیابی شود. نخست، ممکن است بسیاری از مدارک نامرتبط در مجموعۀ پاسخ وارد شود. ثانیاً، مدارک مرتبطی وجود داشته باشد که با هیچ‌کدام از کلید واژه‌های موجود در عبارت پرسش منطبق نباشد و در نتیجه بازیابی نشود. کلیدواژه‌مدار بودن فرایند بازیابی اطلاعات مهم‌ترین دلیل این نارسایی‌هاست. ایدۀ موجود در یک متن بیشتر به مفاهیم مطرح در آن نزدیک است تا کلیدواژه‌هایی که برای توصیف آن به کار برده می‌شود. بنابراین، فرایند تطبیق مدارک با یک عبارت پرسش می‌تواند به جای اینکه بر اساس تطابق کلیدواژه‌های نمایه صورت پذیرد، مبتنی بر تطابق مفهومی باشد. در نتیجه، مدارکی که چه‌بسا با واژگان نمایه‌ای عبارت پرسش مطابقت نداشته باشند ولی از لحاظ مفهومی مرتبط باشند بازیابی می‌گردند. به‌طور مثال، یک مدرک می‌تواند به این دلیل بازیابی شود که با یک مدرک دیگر که به عبارت پرسش مربوط است اشتراک مفهومی دارد. نمایه‌سازی معانی پنهان دقیقا بر اساس همین ایده در سال ۱۹۸۸، شکل گرفت.

ایدۀ اصلی مدل نمایه‌سازی معانی پنهان این است که بردار مربوط به هر سند و عبارت پرسش در فضایی با ابعاد کوچک‌تر که با مفاهیم آنها مرتبط است رسم شود (فورناس[1] و دیگران، ۱۹۸۸). در این مدل اعتقاد بر این است که بازیابی در یک فضای کوچک‌تر ممکن است بهتر از بازیابی در فضای کلیدواژه‌های نمایه‌ای صورت پذیرد. این کار از طریق تجزیۀ مقادیر منفرد[2] انجام می‌شود. بدین‌ترتیب، هم ابعاد فضای اصطلاح -

1. Furans
2. Singular Values Decomposition

مدرک کاهش می‌یابد و هم مشکل اصلی نظام‌های بازیابی اطلاعات ــ یعنی مسائل مربوط به مترادف‌ها و چندمعنایی‌ها ــ تاحدودی حل می‌شود. نمایه‌سازی معانی پنهان سبب می‌شود تا اصطلاحات و مدارک در فضایی با ابعاد بزرگ‌تر و به‌طور کامل بازنمون شوند. در نتیجه روابط معنایی (پنهان) میان اصطلاحات و مدارک در طول جست‌وجو آشکار می‌گردد (اسپوری[1]، ۱۹۹۵).

نمایه‌سازی معانی پنهان برای تشخیص محتوای مفهومی به ساختار جمعی واژگان یک مدرک توجه می‌کند. این مدل، واژگان را شاخصی نامطمئن تلقی می‌کند که نمی‌تواند به‌طور کامل محتوای مدارک را نمایان سازد. در مدل نمایه‌سازی معانی پنهان فرض بر این است که تغییرپذیری گزینش واژگان، بازنمون ساختار معنایی مدارک را با مشکل روبه‌رو می‌سازد. با کاهش ابعاد فضای اصطلاح ـ مدرک، روابط معنایی موجود میان مدارک آشکار می‌شود. عوامل ایجاد اختلال[2] (تفاوت در کاربرد واژگان، واژگانی که در تشخیص تفاوت بین مدارک نقشی ندارند، و نظیر آن) از بین می‌رود. این مدل، با استفاده از الگوهای به‌کارگیری واژگان در مجموعهٔ مدارک، با کنار هم قرار دادن الگوهای مشابه در داخل فضای اصطلاح ـ مدرک، اجازه می‌دهد تا مدارکی که به لحاظ معنایی مرتبط هستند در کنار یکدیگر قرار گیرند؛ حتی اگر دارای واژگان مشترک نباشند، با تحلیل‌های آماری این الگوها نتیجهٔ لازم را به دست می‌آورند (دومیز[3]، ۱۹۹۵).

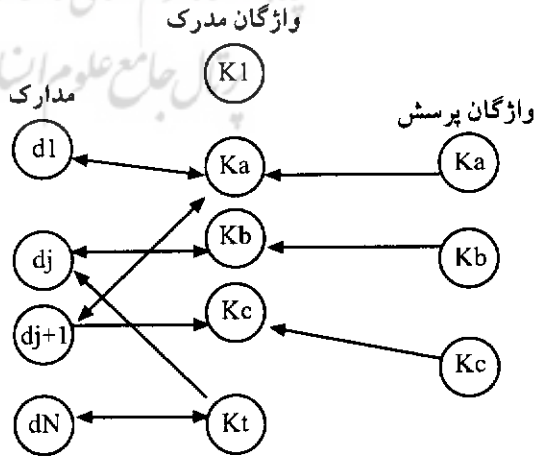
این مدل در مقایسه با مدل‌های دیگر، چندین ویژگی خاص دارد: الف) مدارک را در فضایی با ابعاد بزرگ‌تر بازنمون می‌کنند، ب) هر دوی اصطلاحات و مدارک، در یک فضای مشابه بازنمون می‌شوند، و ج) هیچ تلاشی برای تفسیر معنایی هر کدام از ابعاد صورت نمی‌گیرد؛ بلکه صرفاً فرض می‌شود که هر کدام از ابعاد، یک یا چند روابط معنایی را در فضای اصطلاح ـ مدرک بازنمایی کند. سرانجام، نمایه‌سازی معانی پنهان قادر است مجموعه‌های بزرگ را نیز بازیابی کرده و به شکل عملی به ذخیره و بازیابی آنها بپردازد. می‌توان چنین نتیجه گرفت که نمایه‌سازی معانی پنهان بهتر از مدل‌های دیگر عمل می‌کند. آزمایشی که توسط دومیز (۱۹۹۵) انجام شد، نشان داد که این شیوه ۳۰ درصد بیشتر از سایر مدل‌ها مدارک مرتبط را بازیابی می‌کند.

ج. مدل شبکهٔ عصبی

نظام بازیابی اطلاعات تلاش می‌کند کلیدواژه‌های موجود در عبارت پرسش را با کلیدواژه‌های بازنمون مدرک موجود در مجموعه تطابق دهد. این کار در مدل بولی

1. Spoerri
2. Noise
3. Dumaise

بر اساس منطق دودویی انجام می‌شد و در مدل سنتی برداری میزان تشابه میان بردارهای وزنی مربوط به هر کدام از کلیدواژه‌ها صورت می‌پذیرفت. با توجه به اینکه شبکهٔ عصبی به تطبیق‌گرهای خوب معروف است، به عنوان یکی از مدل‌های جایگزین برای مدل‌های برداری ارائه شده است. همان‌گونه که می‌دانیم، مغز انسان از میلیاردها نورون[1] تشکیل شده است. هر نورون را می‌توان به‌عنوان یک واحد پردازش ابتدایی در نظر گرفت که به محض فعال شدن از طریق علائم درونِدادی، عکس‌العمل نشان داده و علائم برونداد تولید می‌کند. این علائم و نشان‌های حاصله از طریق سیناپس‌ها[2] به سایر نورون‌ها منتقل می‌شود. آنها نیز علائم جدیدی را به عنوان برونداد تولید می‌کنند. این فرایند، که ممکن است در طول چندین لایه از نورون‌ها تکرار شود، به فرایند فعال‌سازی گسترده[3] معروف است. در نتیجه، اطلاعات درونِدادی می‌تواند پردازش (تجزیه و تحلیل و تفسیر) شود. یک شبکه عصبی شبیه تصویر شبکه‌ای به هم پیوسته نورون‌ها در مغز انسان است. گره‌های این شبکه به‌منزلهٔ واحدهای پردازش نورون‌ها است، و رابط میان نورون‌ها نقش سیناپس‌ها را ایفا می‌کنند. موقعیت هر نورون در این شبکه، بر اساس سطح فعالیت آن (موقعیت اولیهٔ نورون مذکور، علائمی که به عنوان درونِداد دریافت می‌کند) تعریف می‌شود. گرهٔ "الف"، با توجه به سطح فعالیت خود، ممکن است علامتی را به گرهٔ (نورون) همجوار خود یعنی گرهٔ "ب" ارسال کند. قدرت تحریک این علامت در گرهٔ "ب" به وزن منتسب به رابط (سیناپس) بین "الف" و "ب" بستگی دارد. شکل ۲ شبکهٔ عصبی بازیابی اطلاعات را نشان می‌دهد (ویلکینسون و هینگستون[4]، ۱۹۹۱).



۱. سلول عصبی
۲. رابط بین سلول‌های عصبی
3. Spread activation process
4. Wilkinson & Hingston

همان‌گونه که در شکل ۲ مشاهده می‌شود، شبکهٔ عصبی از سه لایه تشکیل شده است. لایهٔ واژگان پرسش، لایه واژگان مدرک، و لایهٔ خود مدرک. گره‌های واژگان پرسش از طریق ارسال علامت به گرهٔ واژگان مدرک، فرایند استنباط را آغاز می‌کند. به دنبال آن ممکن است گره‌های واژگان مدرک نیز خود علامت‌هایی را به گرهٔ مدرک ارسال کنند. اینجا مرحلهٔ نخست تکمیل می‌شود. یعنی علامت از گره‌های واژهٔ مدرک به گره‌های مدرک منتقل می‌شود. البته پس از مرحلهٔ اول شبکهٔ عصبی کار خود را متوقف نمی‌کند. در واقع، گره‌های مدرک به نوبهٔ خود ممکن است علامت‌های جدیدی را تولید کنند که به گره‌های واژگان مدرک باز گردانده می‌شود (می‌توان گفت که بین مدرک و واژگان آن ارتباط دوسویه وجود دارد). گره‌های واژگان مدرک ممکن است پس از دریافت محرک‌ها مجدداً پیام‌هایی زا به گره‌های مدرک ارسال کنند و این فرایند تکرار شود. در هر تکرار، علامت‌ها ضعیف‌تر می‌شوند و در نهایت، فرایند فعال‌سازی گسترده متوقف می‌گردد. این فرایند ممکن است یک مدرک را تا زمانی که دیگر هیچ واژهٔ پرسش را دربر نگیرد، فعال سازد. بنابراین، کل این فرایند را می‌توان به منزلهٔ ایجاد یک اصطلاحنامهٔ داخلی تلقی کرد. نظام پس از بررسی تمام علامت‌ها، با استفاده از وزن‌دهی، نتایج را رتبه‌بندی می‌کند. چنین به نظر می‌رسد که این مدل بتواند تا حد امکان مدارک مرتبط با عبارت پرسش را بازیابی کند، اما در عین حال به درستی معلوم نیست که در مجموعه‌های بزرگ‌تر نیز کارایی داشته باشد؛ زیرا به‌نظر می‌رسد بررسی تمام علامت‌ها کار وقت‌گیری باشد. تحقیقات بیشتر در این زمینه می‌تواند این موضوع را دقیق‌تر بررسی نماید.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مدل‌های احتمالی

مدل احتمالی در سال ۱۹۷۶ توسط رابرتسون و اسپارک[1] مطرح شد و بعدها به عنوان مدل بازیابی مستقل دودویی شهرت یافت. این مدل‌ها تلاش می‌کنند تا مسائل بازیابی را در چارچوبی احتمالی حل کنند. با این پیش‌فرض که وقتی کاربر پرسشی را ارائه می‌کند، مجموعه‌ای از مدارک وجود دارد که دقیقاً به آن پرسش مرتبط است؛ و این مجموعه، به مجموعهٔ آرمانی معروف است. با در دست داشتن ویژگی‌های این مجموعهٔ آرمانی دیگر مشکلی برای بازیابی اسناد مرتبط نخواهیم داشت. بنابراین، می‌توان فرایند پرسش را فرایند تشخیص ویژگی‌های یک مجموعهٔ آرمانی پاسخ نامگذاری کرد. مشکل اصلی اینجاست که ما نمی‌دانیم که این ویژگی‌ها چیست، و تنها چیزی که در اختیار

1. Robertson & Spark

داریم واژگان کلیدی است و بایستی از معانی آنها برای تعیین این ویژگی‌ها استفاده کنیم. از آنجایی که این ویژگی‌ها به هنگام پرسش شناسایی نمی‌شوند، بایستی ابتدا این ویژگی‌ها را حدس بزنیم. این حدس اولیه به ما اجازه می‌دهد تا یک مجموعهٔ پاسخ آرمانی را با استفاده از احتمال، به شکل ابتدایی توصیف کنیم؛ در نتیجه، نخستین مجموعه از اسناد بازیابی می‌گردد. تعامل بعدی با کاربر، توصیف مجموعهٔ پرسش آرمانی را تکمیل می‌کند. در طی این تعامل، کاربر مدارک بازیابی شده را مرور و انتخاب می‌کند که کدام یک مرتبط و کدام یک نامرتبط هستند (می‌توان تنها مدارک رتبه‌های بالاتر را بررسی کرد). سپس، نظام از این اطلاعات برای بهبود توصیف مجموعهٔ آرمانی پاسخ استفاده می‌کند. با تکرار این فرایند به دفعات مکرر انتظار می‌رود که چنین توصیفی تکامل یافته و به توصیف واقعی مجموعهٔ آرمانی پاسخ نزدیک‌تر شود. بنابراین، بایستی همیشه در ابتدای کار نیاز به حدس‌زدن توصیف مجموعهٔ آرمانی پاسخ را در ذهن خود داشته باشیم. علاوه بر این، نیاز است تا تلاش آگاهانه‌ای نیز برای شکل‌دهی چنین توصیفی در قالب احتمال صورت گیرد.

به‌طور خلاصه می‌توان گفت که مدل احتمالی، پرسش کاربر را اخذ می‌کند، احتمال ربط و عدم ربط مدارک موجود در مجموعه به پرسش کاربر را حدس می‌زند، و با اخذ بازخورد از کاربر نتایج نهایی را ارائه می‌دهد. لازم به ذکر است که متغیرهای وزن واژگان نمایه‌ای، همگی دودویی هستند.

مزیتی که، به لحاظ نظری، برای مدل احتمالی می‌توان بیان کرد این است که مدارک بر اساس احتمال ربط آنها به پرسش کاربر به شکل از بالا به پایین رتبه‌بندی می‌شوند. در مقابل، معایبی نیز وجود دارد که مهم‌ترین آنها عبارتند از: ۱) نیاز به حدس در جداسازی اولیهٔ مجموعه‌های مرتبط و نامرتبط، ۲) عدم توجه به بسامد واژگان نمایه‌ای درون یک مدرک، و ۳) پذیرش فرضیهٔ استقلال برای واژه‌های نمایه‌ای. البته هنوز به روشنی معلوم نیست که آیا استقلال کلیدواژه‌ها واقعاً یک نقص است یا خیر. برای غلبه بر این نارسایی‌ها، مدل‌های جایگزین احتمال به نام مدل شبکهٔ استنباط و شبکهٔ باور ارائه گردید.

الف. مدل شبکه استنباطی

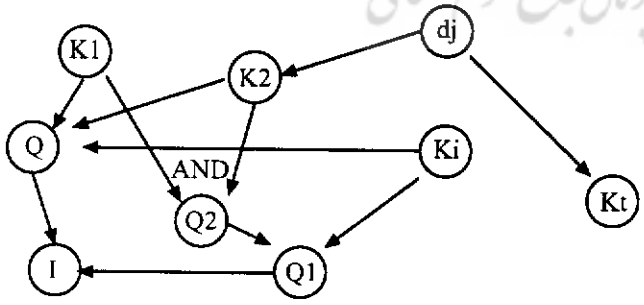
دو مکتب فکری سنتی در حوزهٔ احتمالات عبارتند از دیدگاه بسامدگرا[1] و دیدگاه معرفت‌گرا[2]. دیدگاه بسامدگرا، احتمال را مفهومی آماری درنظر می‌گیرد که به قوانین

1. Frequentist view
2. Epistemological view

شانس (احتمال) مرتبط است. دیدگاه معرفت‌گرا، احتمال را به‌منزلهٔ درجه‌ای از باور تفسیر می‌کند که خاص‌گرایی آن ممکن است آزمایش آماری را منتفی کند. این رویکرد به این دلیل اهمیت دارد که ما در زندگی روزمرهٔ خود به کرّات به احتمالات اشاره می‌کنیم، بدون اینکه تعریف روشنی از آزمایش‌های آماری که احتمالات را تبیین می‌کنند داشته باشیم.

مدل شبکهٔ استنباطی، از دیدگاهی معرفت‌شناختی مسئلهٔ بازیابی اطلاعات را مورد توجه قرار می‌دهد (تورتل و کرافت[1]، ۱۹۹۱). این مدل، با متغیرهای تصادفی واژگان نمایه‌ای، مدارک، و پرسش کاربر کار خود را آغاز می‌کند. یک متغیر تصادفی مربوط به مدرکی معین، رخداد مشاهدهٔ آن سند را بازنمایی می‌کند (به‌طور مثال، این مدل چنین فرض می‌کند که مدارک در هنگام کاوش برای یافتن مدارک مرتبط مشاهده می‌شوند). مشاهدهٔ مدرک معین حکایت از باور و پذیرش متغیرهای تصادفی مربوط به کلید واژه‌های نمایه‌ای آن دارد. بنابراین، مشاهدهٔ یک مدرک به عنوان علت افزایش اطمینان و پذیرش متغیرهای مربوط به واژگان نمایه‌ای آن عمل می‌کند.

متغیرهای اصطلاح نمایه‌ای و مدرک به صورت گره در شبکه نشان داده می‌شوند. رابطها از گرهٔ یک مدرک به گره‌های واژگان آن متصل می‌شود که مشاهدهٔ مدرک موجب باور و پذیرش گره‌های واژگان آن می‌شود. متغیر تصادفی مربوط به عبارت پرسش کاربر، رخدادی را شکل می‌دهد که درخواست اطلاعات کاربر که در قالب عبارت پرسش بیان شده بود، پاسخ داده شده است. باور و پذیرش این گره (عبارت پرسش) تابعی از باورها و پذیرش گره‌های مربوط به واژگان عبارات پرسش است. بنابراین، رابطها از گره‌های واژگان نمایه‌ای به گره‌های مربوط به واژگان عبارت پرسش متصل می‌شود. شکل ۳ شبکهٔ استنباطی بازیابی اطلاعات را به تصویر می‌کشد.



شکل ۳. شبکهٔ استنباطی بازیابی اطلاعات

1. Turtle & Croft

همان‌گونه که مشاهده می‌شود، ki و k2، k1 به عنوان کلیدواژه‌های مرتبط به عبارت پرسش Q یا زیرمجموعهٔ آن (Q1، Q2) تشخیص داده شده است. بنابراین، با توجه به میزان ارتباط هر کدام از آنها با عبارت پرسش، نتایج نهایی ارائه می‌شود. رتبه‌بندی نتایج در مدل شبکهٔ استنباطی بر اساس احتمال اجتماع عبارت پرسش و کلیدواژه‌های مدرک محاسبه می‌شود که نوعی استنباط به‌شمار می‌آید.

ب. مدل شبکه باور

مدل شبکه باور در سال ۱۹۹۶ توسط ریبریو ـ نتو و مانتز معرفی شد. این مدل نیز مانند مدل شبکهٔ استنباطی بر تفسیر معرفت‌شناختی احتمالات استوار است؛ اما به لحاظ پذیرش یک فضای نمونهٔ کاملاً تعریف شده، از مدل شبکهٔ استنباطی متمایز می‌گردد. در نتیجه، توپولوژی (ریخت‌شناسی) شبکه‌ای مبتنی بر احتمال را ارائه می‌کند که میان واحدهای پرسش و خود مدرک تمایز قائل است. این مورد، تفاوت اصلی بین این دو مدل جایگزین برای مدل احتمالی کلاسیک است. در مدل شبکه باور، پرسش کاربر به عنوان گره‌ای از شبکه درنظر گرفته می‌شود که به یک متغیر تصادفی دودویی مرتبط می‌شود. این متغیر در صورتی که پرسش به‌طور کامل فضای کلیدواژه را پوشش دهد، معادل "۱" محسوب می‌شود. بنابراین، با استفاده از احتمال، درجهٔ همپوشانی فضای مفهومی توسط عبارت پرسش محاسبه می‌گردد. یک مدرک را نیز می‌توان به‌عنوان گره‌ای از شبکه در نظر گرفت و با استفاده از احتمال این پوشش، درجهٔ پوشش فضای مفهومی مدرک را محاسبه کرد. با توجه به مطالب بالا، پرسش کاربر و مدارک موجود در مجموعه به صورت زیرمجموعه‌هایی از واژگان نمایه‌ای شکل می‌گیرند. هرکدام از این زیرمجموعه‌ها به منزلهٔ یک مفهوم است که در فضای خاصی تجسم می‌یابد. علاوه بر آن، پرسش‌های کاربران و بازنمون مدارک به صورت شفاف مدل‌سازی می‌شود. این کار به تعریف ریخت‌شناسی شبکهٔ باور می‌انجامد. در این مدل، رتبه‌بندی نتایج براساس رابطهٔ تطابق مفهومی انجام گرفته و درجهٔ پوشش مفهوم موجود در پرسش از طریق مفهوم موجود در مدرک محاسبه می‌شود. انتظارمی‌رود که این مدل بتواند نارسایی‌های موجود در مدل احتمالی کلاسیک را برطرف سازد. اکنون پس از بررسی مدل‌های کلاسیک (بولی، برداری، و احتمالی) و جایگزین‌های آنها می‌توان به بررسی مدل‌های ساختار یافته پرداخت.

مدل‌های ساختاریافته[1]

فرض کنید که یک کاربر با حافظهٔ دیداری قوی به دنبال مدرکی است که در یکی از صفحات آن، واژهٔ "رضایتمندی کاربران" به صورت ایتالیک در داخل متنی قرار گرفته است و پس از آن متن، جدولی وجود دارد که در عنوان آن، واژهٔ "اعضای هیأت علمی" وجود دارد. با استفاده از مدل‌های کلاسیک بازیابی اطلاعات، پرسش را می‌توان به این‌گونه ("رضایتمندی کاربران" + "اعضای هیات علمی") نوشت. با واردکردن این پرسش، تمامی مدارکی که این دو عبارت را دارا هستند بازیابی می‌شود. چنانچه ساختار متن را در نظر گرفته باشیم کاربر می‌تواند پرسش خود را در قالب عبارت بهتری بیان کند: "رضایتمندی کاربران" + "اعضای هیات علمی".

در عبارت پرسش بالا، خواسته‌های دیداری کاربر نیز لحاظ شده است. علاوه بر آن، می‌توان با استفاده از رابط کاربر پیشرفته انجام این کاوش را ساده‌تر کرد. در این مثال، با استفاده از ویژگی‌های ساختاری و بیان آن می‌توان میزان ربط را افزایش داد. مدل‌هایی که اطلاعات مربوط به محتوای متن را با اطلاعات مربوط به ساختار مدرک ترکیب می‌کنند به مدل بازیابی متنی ساختار یافته معروف هستند. در مثال بالا، یک مدل ساختار یافته کلیهٔ مدارکی را که شرایط تعیین شده را داشته باشند مورد جست‌وجو و بازیابی قرار می‌دهد. در نگاه نخست، چنین به نظر می‌رسد که این مدل بیشتر کار بازیابی داده را انجام می‌دهد تا بازیابی اطلاعات؛ زیرا با توجه به تطابق ساختار و کلیدواژه‌ها در ساختاری خاص، آنها را بازیابی می‌کند. اما به تدریج با استفاده از معیار همجواری، نوعی رتبه‌بندی نیز صورت می‌پذیرد که تاحدودی این مسئله را حل می‌کند. در این قسمت به دو مدل بازیابی متون ساختاریافته اشاره می‌شود که عبارتند از: مدل مبتنی بر سیاهه‌های غیرهمپوشان[2] و مدل مبتنی بر گره‌های همجوار[3].

الف. مدل مبتنی بر سیاهه‌های غیرهمپوشان

بورکوفسکی[4] (۱۹۹۲) پیشنهاد می‌کند که کل متن هر مدرک به نقاط متنی غیرهمپوشان تقسیم و به شکل یک سیاهه گردآوری شود. از آنجایی که راه‌های مختلفی برای تقسیم یک متن به مناطق غیرهمپوشان وجود دارد، سیاهه‌های مختلفی می‌توان ایجاد کرد. به‌طور مثال، می‌توانیم سیاهه‌ای از همه بخش‌های یک مدرک را داشته باشیم، در یک سیاههٔ دیگر همه فصل‌های مدرک را گردآوری کنیم و مجدداً در سیاهه‌ای دیگر،

1. Structured Model
2. Non- -overlapping lists
3. Proximated node
4. Burkowski

همهٔ زیر بخش‌های مدرک را جای دهیم. این سیاهه‌ها در قالب ساختارهای جداگانه و متفاوت نگهداری می‌شوند. با اینکه ممکن است نقاط مختلف یک متن در سیاهه‌ای یکسان هیچ نوع همپوشانی نداشته باشند، اما نقاط متنی از سیاهه‌های مختلف ممکن است با همدیگر همپوشانی داشته باشند. برای ایجاد امکان کاوش و بازیابی واژگان کلیدی و نقاط متنی، فایلی مقلوب تهیه می‌شود که هر جزء ساختاری به عنوان یک مدخل در داخل آن نمایه قرارمی‌گیرد. به همراه هر مدخل، سیاههٔ دیگری برای نقاط متنی وجود دارد که بسامد ظهور آنها را در ساختار کلی نشان‌می‌دهد. در این مدل، از آنجایی که نقاط متنی غیر همپوشان هستند، انواع پرسش‌هایی که می‌تواند مطرح شود نیز ساده‌اند. از این رو، با استفاده از عملگرهای بولی می‌توان مدارک مورد نظر را بازیابی کرد.

ب. مدل مبتنی بر گره‌های همجوار

ناوارو[1] و بایزا یتس (۱۹۹۷) مدلی را پیشنهاد کردند که فرد را قادر می‌سازد تا ساختارهای نمایه‌ای سلسله‌مراتبی مستقلی را برای متن یک مدرک تعریف کند. هرکدام از این ساختارهای نمایه‌ای دارای سلسله مراتب معینی است که از فصل‌ها، بخش‌ها، پاراگراف‌ها، صفحات، و سطرها تشکیل شده، و هرکدام از آنها یک گره محسوب می‌شود. هر کدام از این گره‌ها یک یا چند نقطهٔ متنی را به دنبال دارند. علاوه بر آن، ممکن است دو سلسله مراتب متفاوت به نقاط متنی واحدی ارجاع دهند.

با در دست داشتن عبارت پرسش کاربر، که به سلسله مراتب متفاوتی مربوط می‌شود، پاسخ نهایی متشکل از گره‌هایی است که همگی فقط مربوط به یکی از آنهاست. بنابراین، یک پاسخ نمی‌تواند متشکل از گره‌هایی باشد که به دو سلسله مراتب مختلف وابسته است. در جوار این ساختار نمایه‌ای، فایلی مقلوب نیز شکل می‌گیرد که محل قرار گرفتن واژگان را نشان می‌دهد. به هنگام کاوش می‌توان با تعریف عبارت و مشخص کردن اجزای ساختاری و ترکیبی از آنها به جست‌وجو پرداخت. در این حالت، مدل می‌تواند میان گویایی و کارآیی یک عبارت پرسش تعادل ایجاد کند. گویایی عبارت پرسش، هرچند هم محدود باشد، می‌تواند سبب شود که پردازش مؤثر تا اندازه‌ای انجام پذیرد. یعنی ابتدا اجزایی که با عبارت مورد نظر پرسش مطابقت دارند جست‌وجو می‌شوند؛ سپس ارزیابی می‌شود که کدام یک از این اجزاء ــ به لحاظ ساختاری ــ با پرسش ارتباط دارد. می‌توان چنین نتیجه گرفت که در این مدل سرعت بازیابی افزایش پیدا می‌کند، اما با

1. Navarro

توجه به اینکه در مرحلهٔ بعد گره‌های همجوار مورد ارزیابی و جست‌وجو قرار می‌گیرند ممکن است پاسخ نهایی با محدودیت روبه‌رو شود.

مدل‌هایی که تاکنون درباره آنها بحث شد همگی مبتنی بر عبارت پرسش هستند. کاربر پرسشی را مطرح می‌کند و نظام بازیابی پس از اخذ پرسش کاربر با استفاده از الگوریتم‌های خاص خود به تطبیق عبارت پرسش با مدارک موجود می‌پردازد؛ سپس، (در مدل‌های برداری و احتمالی) با وزن‌دهی به واژگان، آنها را بر اساس میزان ربط رتبه‌بندی می‌کند و نتایج نهایی را برای کاربر نمایش می‌دهد. کاربر در فرایندهای پشت صحنه دخالت ندارد. عبارت پرسش بیانگر نیاز اطلاعاتی کاربر به‌شمار می‌رود و میزان ربط نیز از این زاویه تعیین می‌شود؛ چه بسا با نیاز واقعی کاربر فاصله زیادی داشته باشد. مدل‌هایی نیز وجود دارد که از عبارت پرسش استفاده نمی‌کنند بلکه مبتنی بر مرور هستند. در این قسمت به بررسی مدل‌های مروری پرداخته می‌شود.

مدل‌های مروری

گاه ممکن است کاربران مایل نباشند و یا نتوانند یک عبارت پرسش تشکیل دهند. به جای آن، علاقه‌مندند که در میان مدارک موجود و یا بازنمون آنها به وارسی بپردازند. درست مثل اینکه در یک کتابخانه، کاربران به جای جست‌وجو در نظام بازیابی کتابخانه، مایل به مرور کتاب‌های موجود در قفسه باشند. در این وضعیت، می‌گوییم که کاربر به جای "جست‌وجو" می‌خواهد به "مرور" بپردازد. کاربر می‌تواند از هر دو طریق به هدف خود دست یابد. در اینجا به بررسی سه مدل مروری (مسطح[1]، ساختاریافته[2]، و فرامتنی[3]) می‌پردازیم.

الف. مرور مسطح و مرور ساختار یافته

این مدل بر این ایده استوار است که کاربر در فضایی مرتبط با مدرک، که دارای ساختار مسطحی است، به مرور می‌پردازد. به‌طور مثال، اطلاعات مدارک می‌توانند به عنوان نقاط (دو بعدی) یک طرح (در قالب یک چارچوب) و یا به عنوان عناصر (یک بعدی) یک سیاهه (مانند اطلاعات کتاب‌شناختی منابع به صورت الفبایی) نمایش داده شوند. کاربر با مرور اجمالی قسمت‌های مختلف مدرک به دنبال اطلاعات مورد نظر خود می‌گردد. وی همچنین ممکن است در میان مدارک همجوار نیز به دنبال کلیدواژه‌های مورد علاقهٔ خود بگردد. بنابراین، کاربر در یک قالب مسطح به مرور منابع

1. Flat
2. Structured
3. Hypertext

می‌پردازد. در یک صفحهٔ وب، کاربر یک مدرک را با استفاده از حرکت ماوس (موشواره) یا کلیدهای جهت‌نما مرور می‌کند. از آنجایی که این صفحات معمولاً دارای بخش‌های مختلف نیستند، کاربر پس از مرور آنها ممکن است نداند که در کجای مدرک قرار دارد. در نتیجه، برای اینکه مرور به راحتی انجام پذیرد مدارک را در قالب ساختارهای مشخص سازماندهی می‌کنند. این ساختار می‌تواند دارای سلسله‌مراتب خاصی باشد. راهنماها (دایرکتوری)های مورد استفاده در وب (گوگل[1]، یاهو[2]، دیموز[3]، و نظیر آن) از این جمله‌اند. در این صورت، از مدل مرور ساختار یافته استفاده می‌کنیم. یک مدرک نیز می‌تواند از ساختار خاصی (مانند فصل‌ها، بخش‌ها و قسمت‌های دیگر) پیروی کند تا کاربر با مرور این ساختار به متن نهایی برسد، در این صورت از مدل مسطح پیروی کرده است. نظام‌هایی که از این مدل استفاده می‌کنند معمولاً سابقهٔ عملیات مرور را نیز می‌توانند نگهداری کنند تا قسمت‌های مشاهده شدهٔ پیشین را نشان دهد. در نتیجه کاربر می‌تواند از طریق مرور (بدون جست‌وجو) به مدرک یا مدارک مورد نظر خود دست یابد. این مدل‌ها چنانچه به عنوان مکمل و در کنار سایر مدل‌ها ــ به عنوان امکانات جانبی ــ مورد استفاده قرار گیرد؛ کاربر می‌تواند در برخی مواقع که نمی‌تواند نیاز خود را در قالب عبارت پرسش بیان کند، با مرور موجودی به اطلاعات مورد نیاز خود دست یابد. یکی دیگر از مواردی که مدل مروری می‌تواند مکمل سایر مدل‌ها باشد استفاده از اصطلاحنامه برای انتخاب و یا مستندسازی کلیدواژه‌های عبارت پرسش است. در این صورت، کاربر می‌تواند با مرور اصطلاحنامه به کلیدواژهٔ مورد نظر خود برسد و از آن طریق به اطلاعات مرتبط و مورد نیاز دست پیدا کند. در این گونه مواقع، مدل مروری ساختار یافته می‌تواند مفید واقع شود.

پرتال جامع علوم انسانی

ب. مدل فرامتنی

هر متنی که نوشته می‌شود معمولاً چنین تصور می‌شود که به صورت ترتیبی، از بالا به پایین، خوانده خواهد شد و خواننده فقط از این طریق می‌تواند به اطلاعات مرتبط با نیاز خود دست یابد. اما گاه مشاهده می‌شود که خواننده نمی‌تواند با مطالعه (مرور) از بالا به پایین به محتوای اطلاعات دست پیدا کند؛ زیرا وی فقط به بخشی از کتاب یا مقاله علاقه‌مند است. در منابع چاپی و حتی الکترونیکی شیوهٔ تنظیم و ترتیب مطالب این امکان را از خواننده سلب می‌کند. بنابراین، برای حل این مسئله به سازماندهی

1. Google (www.google.com/dirhp)
2. Yahoo (www.yahoo.com)
3. Dmoz (www.dmoz.org)

متفاوتی مانند طراحی فرامتن نیاز است.

فرامتن، ساختار راهبری تعاملی سطح بالاست که فرد را قادر می‌سازد تا بتواند بر روی صفحهٔ کامپیوتر متن مورد نظر را مرور کند بدون آنکه مجبور باشد از سلسله‌مراتب خاصی پیروی کند. پیوندهای موجود می‌تواند فرد را به نقطهٔ خاصی از متن رهنمون گردد. حتی خواننده می‌تواند در صورت نیاز به منابع خارج از متن متصل شود. برخی مواقع تعداد پیوندهای فرامتن آنقدر زیاد است که کاربر در میان آنها مقصد خود را گم می‌کند. برای این کار لازم است هر مدرکی دارای نقشه‌ای باشد که وضعیت گره‌ها و پیوندها را در کل مدرک نشان دهد تا کاربر بتواند بر اساس آنها وضعیت خود را در میان مدرک مشخص کند. به هر ترتیب، ساختار فرامتنی یکی از شیوه‌های مروری است که می‌تواند کاربران را در رسیدن به مطالب مرتبط یاری دهد. مدل فرامتنی از زبان‌های نشانه‌گذاری فرامتنی بهره می‌گیرد. با برقراری پیوند میان قسمت‌های مختلف یک مدرک این امکان فراهم می‌آید تا کاربر بتواند از هر قسمت متن یا مدرک به قسمت دیگر آن منتقل شود. همچنین کاربر با رؤیت عنوان پیوندهای فرامتنی هر مدرک می‌تواند در اسرع وقت به مطلب دلخواه خود دست یابد. علاوه بر آن، مدارک مرتبط نیز می‌توانند به یکدیگر پیوند داشته باشند. در نتیجه، کاربر با دست یافتن به یک مدرک می‌تواند به سایر مدارک مرتبط نیز منتقل شود؛ این امر به افزایش ربط می‌انجامد. از ویژگی‌های مدل فرامتنی آن است که می‌توان در ارائهٔ نتایج نیز از آن بهره برد و نتایج ارائه شده برای یک عبارت پرسش را به مدارک مرتبط دیگر پیوند داد. در عین حال، باید توجه داشت که مرور پیوندها در جست‌وجوی فرامتنی همیشه نمی‌تواند به دستیابی اطلاعات مرتبط بینجامد؛ زیرا واژه، عبارت، یا نشانه‌ای که برای فرامتن و در اشاره به اطلاعات موجود در بخش دیگر یک مدرک و یا اطلاعات سایر مدارک به کار برده می‌شود ممکن است روشن و گویا نباشد. این امر می‌تواند فتح باب پژوهشی جدید و جالب در عرصهٔ بازیابی اطلاعات و افزایش ربط به شمار آید.

استنتاج از مرور مدل‌های بازیابی اطلاعات

با مرور انواع مدل‌های بازیابی می‌توان پی‌برد که عنصر ربط کانون تمرکز همه این مدل‌هاست و همهٔ آنها به نوعی به دنبال این هستند که مدارک بازیابی شده تا حد امکان از ربط بالایی برخوردار باشند. یعنی ربط عنصر کلیدی مدل‌های بازیابی اطلاعات است.

برای ارتقای ربط شیوه‌های مختلفی به کار گرفته می‌شود که عبارتند از: وزن‌دهی، رتبه‌بندی، توجه به ساختار درونی محتوای مدرک، توجه به معانی پنهان، اخذ بازخورد از کاربر، و ایجاد ارتباط دو سویه بین مدرک و واژگان آن. هرکدام از این مدل‌های بازیابی به یک یا چند مورد از مؤلفه‌های مذکور توجه داشته‌اند. به‌طور مثال، مدل‌های برداری، وزن‌دهی را کانون تمرکز الگوریتم‌های مورد استفاده خود قرار داده‌اند و مدل‌های جایگزین آنها تلاش کرده‌اند تا معانی پنهان مدارک را مورد توجه قرار دهند. مدل‌های جایگزین بولی (مجموعهٔ فازی و بولی بسط یافته) مفهوم عضویت را از دودویی به نسبی و تدریجی تغییر داده‌اند تا بتوانند آنها را رتبه‌بندی کرده و ایجاد ارتباط دو سویه بین مدرک و واژگان مدرک را مد نظر قرار دهند. می‌توان گفت که مدل‌های بازیابی اطلاعات در مسیر پیشرفت خود از نظر ارتقای میزان ربط و بازیابی مدارک مرتبط‌تر تلاش کرده‌اند تا:

۱. مفهوم عضویت (یک واژه در یک مدرک یا یک مدرک در مجموعهٔ مدارک) را از دودویی (قطعیت) به عضویت نسبی و تدریجی (مدل مجموعه فازی) تغییر دهند.

۲. نتایج ارائه شده را بر اساس وزن دهی و تشکیل بردارهای دوگانه (بردار مدارک و بردار پرسش) رتبه‌بندی نمایند تا کاربر بتواند از درجهٔ ارتباط آنها به عبارت پرسش خود آگاهی پیدا کند (مدل‌های برداری و شبکهٔ عصبی).

۳. به جای تطابق ظاهری واژگان عبارت پرسش و مدرک، معانی پنهان در عبارات پرسش و مدارک موجود را مورد توجه قرار داده‌اند (مدل نمایه‌سازی معانی پنهان).

۴. ساختار داخلی متن مدرک موجود را به عنوان مؤلفه‌ای یاری‌دهنده برای افزایش ربط و حتی تطبیق تصورات ذهنی کاربر و بازیابی مدارک مرتبط مورد استفاده قرار داده‌اند (مدل‌های ساختاریافته).

علاوه بر افزایش ربط، عناصر دیگری نیز مورد توجه مدل‌های بازیابی اطلاعات قرار گرفته است که از آنها می‌توان به کاربرمدار بودن تعبیر کرد. هرچند که رابط کاربر در نظام‌های اطلاعاتی می‌تواند به صورتی طراحی شود که بتواند رضایت کاربر را جلب کند؛ اما کاربرمدار بودن مدل‌های بازیابی اطلاعات خود، موضوعی است که می‌توان از بررسی مدل‌های بازیابی استنتاج کرد. شیوه‌هایی که برای حصول این هدف در برخی از مدل‌ها رعایت شده است عبارتند از: اخذ بازخورد از کاربر، آگاهی کاربر از فرایند تطبیق عبارات پرسش و واژگان نمایه‌ای مدرک، قابلیت مرور، و امکان مرور فرامتنی.

جدول ۱. ماتریس تطبیقی قابلیت‌های مدل‌های بازیابی اطلاعات

قابلیت‌ها / مدل‌ها	وزن‌دهی عضویت	نحوه واژگان با یکدیگر	ارتباط نتایج	رتبه‌بندی	اخذ بازخورد از کاربر	تعامل مستقیم
بولی	–	دودویی	مستقل	–	–	–
فازی	+	فازی	غیرمستقل	+	+	–
بولی بسط یافته	+	نسبی	غیرمستقل	+	–	–
برداری	+	نسبی	مستقل	+	–	–
برداری تعمیم یافته	+	نسبی	غیرمستقل	+	–	–
نمایه‌سازی معانی پنهان	+	نسبی	غیرمستقل	+	–	–
شبکه‌های عصبی	+	نسبی	غیرمستقل	+	–	–
احتمالاتی	+	دودویی	مستقل	+	+	نسبتاً وجود ندارد
شبکه استنباطی	+	مطرح نیست	غیرمستقل	+	–	–
شبکه باور	+	مطرح نیست	غیرمستقل	+	–	–
سیاهه‌های غیرهمپوشان	–	دودویی	مستقل	–	–	–
گره‌های همجوار	–	دودویی	غیرمستقل	–	–	–
مروری مسطح	مطرح نیست	مطرح‌نیست	غیرمستقل	–	+	+
مروری ساختار یافته	مطرح نیست	مطرح‌نیست	غیرمستقل	–	+	+
مروری فرامتنی	مطرح نیست	مطرح‌نیست	غیرمستقل	–	+	+

همان‌گونه که در جدول ۱ مشاهده می‌شود در میان مدل‌های کلاسیک و مدل‌های جایگزین آنها مدل احتمالی، اخذ بازخورد از کاربر را مورد توجه قرار داده است؛ و همان‌گونه که پیش‌تر اشاره شد، پس از بازیابی مدارک احتمالی، قضاوت کاربر در مورد ربط آنها اساس بازیابی مدارک مرتبط قرار می‌گیرد. اما سایر مدل‌ها تا حدودی همهٔ مراحل را به صورت خودکار همراه با محاسبهٔ بسامد واژگان و یا ساختار محتوایی مدارک و عبارت پرسش انجام می‌دهند (البته همان‌گونه که در جدول ۱ مشاهده می‌شود در برخی از مدل‌ها مانند مدل‌های مروری با توجه به اینکه خودکاربر به مرور می‌پردازد ممکن است بازخورد مستقیم انجام پذیرد ولی در مدل‌های بازیابی عمومیت ندارد).

مدل‌های مروری تنها مدل‌هایی هستند که کاربر مستقیماً مدارک موجود را بررسی و گزینش می‌کند. آنچه که در این میان اهمیت دارد این است که با افزایش بسیار زیاد حجم مدارک، مدل‌های مروری نمی‌توانند پاسخگو باشند؛ زیرا کاربر وقت لازم برای مرور همهٔ متون و حتی فهرست مندرجات یا کلیدواژه‌های آنها را ندارد. بنابراین، مدل‌ها نمی‌توانند از یک رویکرد (بازیابی یا مرور) پیروی کنند. همچنین نمی‌توانند صرفاً یک یا چند مؤلفه (وزن‌دهی، رتبه‌بندی، نحوه عضویت، توجه به ساختار و معانی پنهان و نظیر آن) را رعایت کنند. بلکه مدل‌هایی موفق خواهند بود که همهٔ این موارد را در کنار هم و به محض نیاز استفاده کنند. همان‌گونه که بلکین و دیگران (۱۹۹۳) اشاره کرده‌اند، لازم است ترکیبی از رویکردهای تطبیق کامل و تطبیق نسبی مورد استفاده قرار گیرد. گسترش محیط وب و فن‌آوری‌های مربوط به آن و گسترش به‌کارگیری هوش مصنوعی در بازیابی اطلاعات از جمله عواملی است که می‌تواند امکان دستیابی به یک مدل مطلوب بازیابی اطلاعات را فراهم آورد. می‌توان گفت که رویکرد دانش‌مدار بازیابی اطلاعات در حال شکل‌گیری است.

بازیابی اطلاعات با استفاده از رویکرد دانش‌مدار

مدل‌هایی که تاکنون مورد بحث قرار گرفت، به جز در برخی موارد (مانند مدل فازی و نمایه‌سازی معانی پنهان)، همگی مدارک را بر اساس حضور کلیدواژه‌های عبارت پرسش بازیابی می‌کنند. در این رویکردها ممکن است بسیاری از مدارک مرتبط صرفاً به‌سبب عدم تشابه کلیدواژه‌های مدرک با کلیدواژه‌های عبارت پرسش بازیابی نشوند. بنابراین، به رویکرد دیگری نیاز است تا بتوان بر اساس آن مدارک مرتبط را حتی در صورت عدم حضور کلیدواژهٔ عبارت پرسش بازیابی کرد. از این رویکرد به رویکرد دانش‌مدار بازیابی اطلاعات تعبیر می‌کنند (لنکستر و وارنر[1]، ۱۹۹۳). در این رویکرد سعی می‌شود با استفاده از تحلیل‌های شکلی، صرفی، و نحوی مدارک میزان ربط آنها با عبارت پرسش تعیین شود. با استفاده از تحلیل شکلی ریشه، پیشوند و پسوند واژگان تعیین می‌شود؛ با استفاده از تحلیل صرفی، شکل جمله‌بندی و موقعیت واژگان در جمله‌ها شناسایی و تفسیر می‌شود؛ و سرانجام، با استفاده از تحلیل نحوی، مفهوم واژه در قالب جملات و حتی مترادف‌ها و متضادها و نظیر آن شناسایی می‌گردد تا در حد امکان بتوان مدارک مرتبط‌تر را بازیابی کرد. این رویکرد مستلزم دانش زبان‌شناسی و

1. Lancaster & Warner

به‌کارگیری هوش مصنوعی در کار بازیابی اطلاعات است. متخصصان علوم اطلاع‌رسانی می‌توانند در این زمینه با کمک گرفتن از علوم دیگر به شکل‌گیری نظام‌های بازیابی اطلاعات دانش‌مدار کمک کنند. همهٔ اینها مستلزم تحقیقات وسیع و دامنه‌دار در این زمینه است.

سخن پایانی

گرچه ربط امری نسبی است، به‌کارگیری مدل‌های کارآمد می‌تواند تاحدودی مسئلهٔ ربط ظاهری یا تطابق فیزیکی میان واژگان عبارت پرسش و مدارک موجود در مجموعه را حل نماید. شکل‌گیری و روند رو به رشد مدل‌های بازیابی به وضوح این ادعا را تأیید می‌کند. هرچند که حوزهٔ بازیابی اطلاعات حوزهٔ پژوهشی گسترده‌ای است، در ایران تحقیقات اندکی در این زمینه انجام شده است. نرم‌افزارهای کتابشناختی، کتابخانه‌ای، و تجاری موجود در ایران بیشتر از مدل‌های ابتدایی (بولی) استفاده می‌کنند (این ادعا حداقل در مورد نرم‌افزارهای کتابخانه‌ای موجود صادق است). گرایش به سوی استفاده از مدل‌های پیشرفته که ترکیبی از چند مدل (بولی، برداری، ساختاریافته و مروری) را به‌کار می‌گیرند می‌تواند به کاربران کمک کند تا به اطلاعات مرتبط که زمینه‌ساز تصمیم‌گیری آگاهانه است دست پیدا کنند. این امر محقق نمی‌شود، مگر با تلاش مستمر و ژرف متخصصان حوزه‌های علوم اطلاع‌رسانی و علوم رایانه تا بتوانند نظریه و عمل را در راستای ارائهٔ راه‌حل‌های مناسب برای مسائل بازیابی اطلاعات در هم آمیخته و در برآورده ساختن نیازهای اطلاعاتی کاربران پیشقدم باشند. استادان و دانشجویان، به‌ویژه در دوره‌های تحصیلات تکمیلی، می‌توانند حوزه بازیابی اطلاعات را به عنوان یک حوزهٔ تحقیقاتی جذاب و کارساز، به عنوان حوزهٔ تخصصی خود، پیش گرفته و در این حوزهٔ به تولید دانش تخصصی و بومی بپردازند. در چنین وضعیتی می‌توان به موفقیت نظام‌های بازیابی اطلاعات در افزایش ربط، یعنی پاسخگویی مؤثر به نیازهای اطلاعاتی، امیدوار بود. سخن پایان اینکه نباید از یاد برد که تحول در حوزهٔ بازیابی اطلاعات همچنان ادامه خواهد داشت.

مآخذ:

Baeza-Yates, R. & Riberio-Neto, B. (1999). *Modern information retrieval*. New

York. ACM Press.

Belkin, N. et al. (1993) "The effect of multiple query representations on information retrieval systems performance,". In *proceeding of 93rd annual international ACM SIGIR conference on research and development in information retrieval*. Pittsburgh.

Burkowski, F. (1992). "An algebra for hierarchically organized text- dominated databases". *Information Processing & Management*. Vol. 28, No. 3, pp. 333-48.

Dumaise, S. (1995). "Improving the retrieval of information from external sources". *Behavior Research Methods, Instruments and Computers*. Vol. 23, No. 2. pp. 210-236.

Furnas, G. et al. (1988). "Information retrieval using a singular value decomposition". In *proceeding of 11th annual international ACM. SIGIR conferences on research and development in information retrieval*. pp. 405-480.

Gudivada, V. et al. (1997). "Information retrieval on the World Wide Web". *IEEE Internet Computing*. Oct- Nov. pp. 58-68.

Lancaster, F. & Warner, A. (1993). "Information retrieval today". *Information Resources*, Arlington VA.

Navarro, G. & Baeza-Yates, R.(1997). "Proximal nodes: A model to quary document databases. by content and structure". *ACM Transactions on Office and Information Systems*. Vol. 15, No. 4, pp. 405.

Riberio-Neto, B. & Muntz, R. (1996). "A belief network for information retrieval". in *proceeding of 9th annual international ACM SIGIR conference on research and development in information retrieval*. Zurich, Switzerland. pp. 253-260.

Robertson, S. & Spark, K.(1976). "Relevance weighting of search terms". *Journal of American Society for Information Science*. Vol. 27, No.3, pp. 146.

Salton, G. & Lesk, M.E. (1968). "Computer evaluation of indexing and text

processing". *Journal of ACM*. Vol. 15, No. 1, pp. 1- 36.

Salton, G. et al. (1983). "Extended boolean information retrieval". *Communications of the ACM*. Vol. 11, pp. 1032-36.

Spoerri, A.(1995). Infocrystal: A visual tool for information retrieval (PhD thesis). Department of Civil and Environmental Engineering. [online]. Available: http://www.scils.rutgers.edu/~aspoerri/InfoCrystal/

Turtle, & Croft, L.(1991). Evaluation of an inference network - based retrieval model. *ACM Transactions on Information Systems*. Vol. 9, No. 3, pp. 187-222.

Wilkinson, R. & Hingston, P. (1991). "Using the cosine measure in a neural network for document retrieval". In *proceeding of ACM. SIGIR conferences on research and development information retrieval*. Chicago. pp. 202-210.

Wong, S. Ziakro, V. & Wong, P.(1985). "Generalized vector space model in information retrieval". In *proceeding of 8th ACM. SIGIR conference on research and development in information retrieval*. New York. pp. 18-25.

Zadeh, L. A. (1993). Fuzzy sets. In Dubois, S, Prade, H. & Yager, R. (Eds.). *Readings in fuzzy sets for intelligent systems*. Morgan Kaufmann.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی