



دانشگاه شهید بهشتی

دانشکده مهندسی و علوم کامپیوتر

ارزیابی روش های پیشبینی تخصص آینده

پایان نامه کارشناسی ارشد مهندسی کامپیوتر
گرایش نرم افزار

نگارش

سینا محمدی

استاد راهنما

دکتر محمود نشاطی

تابستان ۱۴۰۳



دانشگاه شهید بهشتی
دانشکده مهندسی و علوم کامپیوتر

پایان نامه کارشناسی ارشد مهندسی کامپیوتر - گرایش نرم افزار
تحت عنوان:
ارزیابی روش های پیشبینی تخصص آینده

در تاریخ
پایان نامه دانشجو، سینا محمدی، توسط کمیته تخصصی داوران مورد بررسی و تصویب نهایی قرار
گرفت.

امضا	نام و نام خانوادگی	۱- استاد راهنما اول:
امضا	نام و نام خانوادگی	۲- استاد راهنما دوم:
امضا	نام و نام خانوادگی	۳- استاد مشاور:
امضا	نام و نام خانوادگی	۴- استاد داور (داخلی):
امضا	نام و نام خانوادگی	۵- استاد داور (خارجی):
امضا	نام و نام خانوادگی	۶- نماینده تحصیلات تکمیلی:

کلیه حقوق مادی مترتب بر نتایج مطالعات،
ابتکارات و نوآوری‌های ناشی از تحقیق موضوع
این پایان‌نامه متعلق به دانشگاه شهید بهشتی
می‌باشد.

به نام خدا

نام و نام خانوادگی: سینا محمدی

عنوان پایان نامه: ارزیابی روش های پیشبینی تخصص آینده

استاد راهنما: دکتر محمود نشاطی

اینجانب سینا محمدی تهیه کننده پایان نامه کارشناسی ارشد حاضر، خود را ملزم به حفظ امانت داری و قدردانی از زحمات سایر محققین و نویسندگان بنابر قانون Copyright می دانم. بدین وسیله اعلام می نمایم که مسئولیت کلیه مطالب درج شده با اینجانب می باشد و در صورت استفاده از اشکال، جداول و مطالب سایر منابع، بلافاصله مرجع آن ذکر شده و سایر مطالب از کار تحقیقاتی اینجانب استخراج گشته است و امانت داری را به صورت کامل رعایت نموده ام. در صورتی که خلاف این مطلب ثابت شود، مسئولیت کلیه عواقب قانونی با شخص اینجانب می باشد.

نام و نام خانوادگی: سینا محمدی

تاریخ و امضا:

چکیده

بازیابی تخصص یکی از موضوعات تحقیقاتی در مباحث مربوط به بازیابی اطلاعات میباشد. این موضوع با هدف یافتن نیروی متخصص یا تحلیل داده های کلان که بیشتر به شکل متن ها و نوشته ها ظاهر شده و ساختار مشخصی ندارد، کمک میکند تا سازمان ها و شرکت ها بتوانند نیروی متخصص مورد نیاز خود را به راحتی پیدا کنند. به طور کلی مباحث مربوط به بازیابی تخصص به چندین دسته تقسیم می شوند. دسته اول شامل استخراج داده و آماده سازی آن میباشد. که این داده ها برای مقاصد تحقیقاتی معمولاً از انجمن های آنلاین استخراج میشود. دسته دوم شامل دسته بندی افراد متخصص، نمره دهی به آن ها و به طور کلی یافتن بهترین کاندیدا از بین کاندیدا های موجود میباشد. دسته آخر عملیات های تحلیل داده و برخی متد های علوم داده را روی داده های استخراج شده از دسته دوم اعمال میکند و تحلیل های دقیق تری از متخصص ها ارائه می دهد. در این مطلب تلاش شده با تمرکز روی دسته سوم، داده های استخراج شده از کاربران را بر پایه زمان دسته بندی کرده و با اعمال یک سری از روش های آماری و احتمالاتی پیش بینی و تخمین حدودی از تخصصی که یک متخصص در آینده ممکن است کسب کند ارائه دهد. این پیشبینی به سازمان ها و شرکت ها کمک میکند تا بهترین کاندیدا را بر اساس برنامه ای که برای آینده شغلی فرد استخدام شده دارند و مسیری از یادگیری مهارت ها که برای آن فرد در نظر گرفته اند پیدا کرده و خطای کمتری در استخدام نیروی متخصص داشته باشند.

واژگان کلیدی: بازیابی تخصص، بازیابی اطلاعات، تحلیل زمانی

فهرست مطالب

۱	مقدمه	۱
۳	ادبیات تحقیق	۲
۴	۱.۲ مدل های احتمالاتی	۴
۴	۲.۲ خوشه بندی	۴
۵	۳.۲ مدل های احتمالاتی بر پایه پیش بینی آینده	۵
۵	۴.۲ یادگیری ماشین	۵
۶	۵.۲ بازیابی سند	۶
۶	۶.۲ سری های زمانی	۶
۷	۷.۲ معیار های ارزیابی	۷
۹	۳ کار های مرتبط	۹
۱۳	۴ نتیجه گیری و کار های آینده	۱۳
۱۵	مراجع	۱۵
۱۶	واژه نامه انگلیسی به فارسی	۱۶
۱۷	واژه نامه فارسی به انگلیسی	۱۷

فهرست تصاویر

فهرست جداول

فصل ۱

مقدمه

مسائل بازیابی تخصص به بررسی و تحلیل میزان تخصص افراد، حوزه های فعالیت آن ها و همچنین میزان عمیق بودن آن ها در هر حوزه تخصص می پردازد.

مطالعات پیشین در این حوزه به معرفی روش هایی میپردازد که با تحلیل متن هایی که هر کاربر در یک سری از وب سایت های مرتبط تولید کرده است . تخصص های آن کاربر را مشخص کرده و با تکنیک های خوشه بندی این مهارت ها را به دسته های مشخص تقسیم بندی میکند. از طرف دیگر مطالعاتی وجود دارد که میزان عمق دانش کاربران در هر حوزه تخصصی را مشخص میکند. و به این صورت مشخص میکند که کاربر از لحاظ میزان مهارت در حد یک کارآموز میباشد یا مهارت های عمیق تری دارد و میتواند آن را در سطوح بالاتر قرار داد.

در این مطلب تلاش شده در ادامه مطالعات پیشین روش هایی ارائه شود که از طریق آن با بررسی تغییرات تخصص افراد در حوزه های مختلف پیش بینی از تخصص هایی که فرد احتمالاً در آینده به سراغ آن ها خواهد رفت ارائه کند. این موضوع از چند جهت حائز اهمیت است. اول این که کسب و کار هایی که به دنبال جذب نیروی کارآموز و آموزش آن میباشند میتوانند با این تکنیک ها مسیر پیش بینی شده آینده آن فرد را پیدا کرده و از میان کاندید های موجود فردی را که مسیر یادگیری مهارت او تناسب بیشتری با نقشه راهی که آن کسب و کار برای آینده آن کارآموز در نظر گرفته مطابقت دارد جذب میکنند.

از جهت دیگر کسب و کار هایی که در نقشه راه آن ها بگونه ای است که در ماه ها و سال های آینده به سراغ یک تکنولوژی یا ابزار خاص خواهند رفت نیازی نیست که از ماه ها قبل فردی را با آن مهارت جذب کنند (این کار هزینه بالاتری برای کسب و کار دارد). آن ها با استفاده از این تکنیک فردی را که مهارت استفاده از ابزار هایی را دارد که کسب و کار در حال حاضر از آن استفاده میکند را دارد و پیشبینی مسیر یادگیری او به گونه ای است که در آینده در مهارتی که به کسب و کار اضافه میشود تخصص پیدا میکنند جذب میکنند.

علاوه بر دو مثال فوق میتوان مثال های دیگری نیز ارائه کرد که این راهکار فرآیند جذب و استخدام را بهینه میکند و هزینه نهایی آموزش نیروی جدید را برای کسب و کار ها کاهش میدهد.

فصل ۲

ادبیات تحقیق

۱.۲ مدل های احتمالاتی

مدل های احتمالاتی به منظور تحلیل و پیش بینی پدیده ها و داده ها با استفاده از اصول احتمال به کار می روند. این مدل ها از مفاهیم احتمالات برای مدل سازی عدم قطعیت و تغییرپذیری داده ها بهره می برند. از جمله کاربردهای مدل های احتمالاتی می توان به تشخیص الگوها، پیش بینی نتایج و تصمیم گیری تحت عدم قطعیت اشاره کرد. این مدل ها در زمینه های مختلف مانند یادگیری ماشین، پردازش زبان طبیعی و تحلیل داده کاربرد دارند. با استفاده از داده های تاریخی و توزیع های احتمالی، مدل های احتمالاتی می توانند نتایج آینده را با دقت بالاتری پیش بینی کنند.

۲.۲ خوشه بندی

خوشه بندی به عنوان یکی از تکنیک های مهم در حوزه بازیابی اطلاعات و به ویژه در بازیابی تخصص ها نقش بسزایی ایفا می کند. خوشه بندی فرآیندی است که طی آن، مجموعه ای از داده ها به گروه های مشابه و همگن تقسیم می شوند. این تکنیک با هدف دسته بندی داده ها به گونه ای انجام می شود که داده های موجود در هر خوشه بیشترین شباهت را به یکدیگر و کمترین شباهت را به داده های خوشه های دیگر داشته باشند. در حوزه بازیابی تخصص ها، خوشه بندی می تواند برای شناسایی و گروه بندی تخصص های مشابه استفاده شود. به طور مثال، می توان متخصصانی که در یک زمینه خاص فعالیت دارند را در یک خوشه قرار داد تا به این ترتیب، بازیابی اطلاعات مرتبط با تخصص ها به شکل کارآمدتر و دقیق تری انجام شود. این امر می تواند در سیستم های پیشنهاددهنده، جستجوهای تخصصی و تحلیل شبکه های اجتماعی به کار رود. روش های مختلفی برای انجام خوشه بندی وجود دارد که از جمله مهم ترین آن ها می توان به روش های مبتنی بر فاصله، و روش های سلسله مراتبی اشاره کرد. هر یک از این روش ها با توجه به ویژگی های داده ها و اهداف خوشه بندی، مزایا و محدودیت های خاص خود را دارند. در حوزه بازیابی اطلاعات، خوشه بندی علاوه بر دسته بندی تخصص ها، می تواند به بهبود دقت نتایج جستجو، کاهش پیچیدگی فضای جستجو و افزایش کارایی سیستم های بازیابی کمک کند. به عنوان مثال، با خوشه بندی مدارک و مقالات علمی، می توان به جستجوی موضوعات مرتبط پرداخت و نتایج دقیق تری را به کاربران ارائه داد.

در نهایت، می‌توان گفت که خوشه‌بندی به‌عنوان یکی از ابزارهای قدرتمند در تحلیل داده‌ها و بازیابی اطلاعات، نقش کلیدی در بهبود فرآیندهای بازیابی تخصص‌ها و مدیریت دانش ایفا می‌کند و می‌تواند به ارتقای سطح کیفی سیستم‌های اطلاعاتی منجر شود.

۳.۲ مدل‌های احتمالاتی بر پایه پیش‌بینی آینده

این مدل‌ها بر پایه تئوری احتمال و آمار بنا شده‌اند و با استفاده از داده‌های گذشته، سعی در پیش‌بینی وقایع و روندهای آینده دارند. در بازیابی تخصص‌ها، مدل‌های احتمالاتی می‌توانند برای پیش‌بینی نیازهای آینده به تخصص‌های خاص، شناسایی تغییرات در الگوهای تخصص و پیشنهاد تخصص‌های مورد نیاز در بازار کار استفاده شوند. یکی از کاربردهای مهم مدل‌های احتمالاتی در بازیابی تخصص‌ها، تحلیل پیش‌بینانه بر اساس داده‌های تاریخی است. این مدل‌ها با تحلیل داده‌های مربوط به فعالیت‌های گذشته متخصصان و روندهای شغلی، می‌توانند پیش‌بینی‌هایی در مورد آینده این تخصص‌ها ارائه دهند. به‌عنوان مثال، با استفاده از مدل‌های احتمالاتی می‌توان تخصص‌هایی که احتمال رشد بالایی دارند را شناسایی کرده و به کاربران پیشنهاد داد. از جمله روش‌های معروف در این حوزه، مدل‌های مارکوف، زنجیره‌های زمانی و تحلیل رگرسیون می‌باشند. مدل‌های مارکوف با استفاده از احتمال انتقال از یک حالت به حالت دیگر، روندهای آینده را پیش‌بینی می‌کنند. زنجیره‌های زمانی نیز با تحلیل داده‌های سری زمانی و بررسی الگوهای تکرار شونده، به پیش‌بینی آینده می‌پردازند. تحلیل رگرسیون نیز با استفاده از متغیرهای مستقل و وابسته، روابط بین داده‌ها را مدل‌سازی کرده و پیش‌بینی‌های دقیقی ارائه می‌دهد.

این مدل‌ها نه تنها در پیش‌بینی روندهای آینده، بلکه در بهبود سیستم‌های پیشنهاددهنده و جستجو نیز کاربرد دارند. به‌طور مثال، با پیش‌بینی نیازهای آینده به تخصص‌های خاص، سیستم‌های بازیابی می‌توانند به کاربران خود تخصص‌های مورد نیاز در آینده را پیشنهاد دهند و آن‌ها را برای بازار کار آینده آماده کنند. در مجموع، مدل‌های احتمالاتی با توانایی پیش‌بینی و تحلیل داده‌ها، نقش کلیدی در بهبود کارایی سیستم‌های بازیابی اطلاعات و تخصص‌ها ایفا می‌کنند و به بهینه‌سازی فرآیندهای تصمیم‌گیری و مدیریت دانش کمک می‌نمایند.

۴.۲ یادگیری ماشین

یادگیری ماشین به عنوان یکی از شاخه‌های پیشرفته و نوین علوم کامپیوتر، در سال‌های اخیر مورد توجه فراوانی قرار گرفته است. این فناوری با استفاده از الگوریتم‌ها و مدل‌های ریاضی قادر است به تحلیل و تفسیر داده‌ها پرداخته و الگوهای موجود در آنها را شناسایی کند. یکی از کاربردهای مهم یادگیری ماشین، پیش‌بینی آینده بر اساس داده‌های گذشته و حال است. در زمینه بازیابی اطلاعات، یادگیری ماشین می‌تواند نقش مؤثری در بهبود فرآیند بازیابی تخصص‌ها ایفا کند. بازیابی تخصص‌ها، که زیرمجموعه‌ای از بازیابی اطلاعات محسوب می‌شود، به معنای شناسایی و استخراج دانش و مهارت‌های افراد متخصص در یک حوزه خاص است. با استفاده از تکنیک‌های یادگیری ماشین، می‌توان الگوها و روابط پنهان میان داده‌های موجود را شناسایی و از آنها برای پیش‌بینی تخصص‌ها و مهارت‌های مورد نیاز در آینده استفاده کرد. مدل‌های یادگیری ماشین قادرند با تحلیل داده‌های مربوط به سوابق کاری، مقالات علمی، پروژه‌های تحقیقاتی و سایر منابع مرتبط، به شناسایی الگوهای تخصصی بپردازند. این مدل‌ها می‌توانند بر اساس داده‌های جمع‌آوری شده، نیازهای آتی بازار کار را پیش‌بینی کرده و به سازمان‌ها کمک کنند تا برای آینده آماده باشند. به عنوان مثال، با تحلیل داده‌های موجود، می‌توان پیش‌بینی کرد که کدام تخصص‌ها در سال‌های آتی بیشتر مورد نیاز خواهند بود و چه مهارت‌هایی باید تقویت شوند. به طور خلاصه، یادگیری ماشین با توانایی پیش‌بینی آینده و تحلیل داده‌های موجود، می‌تواند به عنوان ابزاری قدرتمند در بهبود فرآیند بازیابی تخصص‌ها و افزایش کارایی سازمان‌ها و نهادها عمل کند. این فناوری نه تنها به شناسایی تخصص‌های مورد نیاز کمک می‌کند، بلکه می‌تواند به توسعه و بهبود مهارت‌های فعلی نیز یاری رساند.

۵.۲ بازیابی سند

بازیابی سند به عنوان بخشی اساسی از علم اطلاعات، نقش مهمی در پژوهش‌های ادبیات موضوعی ایفا می‌کند. این فرآیند به معنای جستجو، انتخاب و بازیابی مستندات و اسناد مرتبط با یک موضوع خاص است. با استفاده از روش‌های پیشرفته بازیابی اطلاعات، می‌توان الگوها و روابط پنهان در بین متون ادبی را شناسایی کرده و به تحلیل دقیق‌تری از آنها پرداخت. بازیابی سند با استفاده از فنون یادگیری ماشین و پردازش زبان طبیعی،

می‌تواند به پیش‌بینی موضوعات و مسائل آینده در ادبیات موضوعی کمک کند. این روش‌ها با تجزیه و تحلیل متون موجود، می‌توانند الگوهایی را کشف کنند که در آینده به عنوان موضوعات جدید یا جنبه‌های نو ادبیات مطرح خواهند شد. به طور خلاصه، بازیابی سند با بهره‌گیری از رویکردهای پیشرفته علم اطلاعات و همچنین فنون یادگیری ماشین، می‌تواند در پژوهش‌های ادبیات موضوعی به بهبود فرآیند انتخاب و تحلیل متون کمک کند و به پیش‌بینی آینده این حوزه نیز کمک نماید.

۶.۲ سری‌های زمانی

سری‌های زمانی به عنوان یک روش تحلیلی در علوم داده‌ها، برای مدل‌سازی و پیش‌بینی رفتار آینده داده‌ها استفاده می‌شوند. این روش بر مبنای مشاهدات متوالی در طول زمان استوار است که می‌تواند شامل داده‌های مالی، آب و هوا، فروش و غیره باشد. هدف اصلی تحلیل سری‌های زمانی، شناسایی الگوها و روندهای موجود در داده‌ها و استفاده از آنها برای پیش‌بینی مقادیر آینده است. در تحلیل سری‌های زمانی، مفاهیم اساسی شامل روند، فصلی بودن و نویز هستند. روند به تغییرات بلندمدت در داده‌ها اشاره دارد، در حالی که فصلی بودن به الگوهای تکراری در بازه‌های زمانی مشخص اطلاق می‌شود. نویز نیز به نوسانات تصادفی و غیرقابل پیش‌بینی در داده‌ها گفته می‌شود. یکی از روش‌های متداول برای مدل‌سازی سری‌های زمانی، مدل‌های خودهمبستگی و میانگین متحرک است. در این مدل‌ها، مقادیر آینده به صورت تابعی از مقادیر گذشته و خطاهای پیش‌بینی قبلی تخمین زده می‌شوند. به عنوان مثال، مدل‌های خودهمبستگی مقادیر آینده را بر اساس همبستگی مقادیر گذشته پیش‌بینی می‌کنند، در حالی که مدل‌های میانگین متحرک از خطاهای پیش‌بینی‌های قبلی برای اصلاح مقادیر آینده استفاده می‌کنند. روش دیگر برای تحلیل سری‌های زمانی، تحلیل طیفی است که در آن سری‌های زمانی به اجزای فرکانسی تجزیه می‌شوند. این روش برای شناسایی دوره‌های مختلف نوسانات در داده‌ها مفید است. در حوزه بازیابی تخصص‌ها، تحلیل سری‌های زمانی می‌تواند به پیش‌بینی نیازهای آتی بازار کار و تخصص‌های مورد نیاز کمک کند. با تحلیل داده‌های تاریخی مربوط به استخدام‌ها، پروژه‌ها و مهارت‌های مورد تقاضا، می‌توان روندهای تغییرات در بازار کار را شناسایی کرده و برای برنامه‌ریزی‌های آتی استفاده کرد. به طور خلاصه، تحلیل

سری‌های زمانی با شناسایی الگوها و روندهای موجود در داده‌ها، ابزاری قدرتمند برای پیش‌بینی آینده و تصمیم‌گیری‌های مبتنی بر داده فراهم می‌کند. این روش به ویژه در حوزه‌هایی که نیاز به پیش‌بینی دقیق و برنامه‌ریزی بلندمدت دارند، کاربرد دارد.

۷.۲ معیارهای ارزیابی

ارزیابی مدل‌های بازایی اطلاعات در حوزه علوم کامپیوتر با استفاده از معیارهای مختلفی مانند دقت، بازآوری و میانگین دقت متوسط (MAP) انجام می‌شود که هر یک از این معیارها جنبه‌های متفاوتی از عملکرد مدل را اندازه‌گیری می‌کنند.

دقت (Precision) به معنای نسبت تعداد مدارک مرتبط بازایی شده به کل مدارک بازایی شده است. به بیان دیگر، دقت نشان می‌دهد که از میان مدارکی که توسط مدل بازایی شده‌اند، چند درصد از آن‌ها واقعاً مرتبط با پرسش کاربر بوده‌اند. به عنوان مثال، اگر مدل ۱۰ مدرک را بازایی کند و از این تعداد، ۷ مدرک مرتبط باشند، دقت مدل ۷۰ درصد خواهد بود. دقت بالا نشان می‌دهد که مدل قادر است نتایج دقیق‌تری را به کاربر ارائه دهد و خطای نوع اول (False Positive) کمتر است.

بازآوری (Recall) معیار دیگری است که تمرکز آن بر میزان بازایی مدارک مرتبط است. بازآوری به معنای نسبت تعداد مدارک مرتبط بازایی شده به کل مدارک مرتبط موجود در مجموعه است. به عبارت دیگر، این معیار نشان می‌دهد که مدل تا چه حد توانسته است تمامی مدارک مرتبط با پرسش کاربر را بازایی کند. برای مثال، اگر در مجموعه ۲۰ مدرک مرتبط وجود داشته باشد و مدل ۱۵ مورد از آن‌ها را بازایی کند، بازآوری ۷۵ درصد خواهد بود. بازآوری بالا نشان‌دهنده توانایی مدل در یافتن اکثر یا تمامی مدارک مرتبط است و نشان می‌دهد که خطای نوع دوم (False Negative) کمتر است.

میانگین دقت متوسط (MAP) یکی از جامع‌ترین معیارها برای ارزیابی مدل‌های بازایی اطلاعات است که هر دو جنبه دقت و بازآوری را در نظر می‌گیرد. MAP به این صورت محاسبه می‌شود که ابتدا برای هر پرسش دقت در هر رتبه‌ای که یک مدرک مرتبط بازایی می‌شود محاسبه می‌گردد. سپس دقت‌های میانگین برای هر پرسش محاسبه شده و در نهایت میانگین این دقت‌ها برای تمامی پرسش‌ها به عنوان MAP محاسبه می‌شود.

MAP می‌تواند به عنوان یک معیار کلی برای مقایسه عملکرد مدل‌های مختلف مورد استفاده قرار گیرد، زیرا هم دقت در رتبه‌های بالاتر و هم توانایی مدل در بازیابی مدارک مرتبط در طول تمامی رتبه‌ها را می‌سنجد. به دلیل این ویژگی‌ها، MAP اغلب در مقالات علمی و ارزیابی سیستم‌های بازیابی اطلاعات به کار می‌رود. با استفاده از این معیارها، محققان و مهندسان می‌توانند مدل‌های بازیابی اطلاعات را از جنبه‌های مختلف ارزیابی کرده و نقاط قوت و ضعف هر مدل را شناسایی کنند. این امر به بهبود سیستم‌های بازیابی اطلاعات و ارائه نتایج دقیق‌تر و کامل‌تر به کاربران کمک شایانی می‌کند. به طور کلی، درک صحیح از این معیارها و استفاده مناسب از آن‌ها می‌تواند تأثیر قابل توجهی در طراحی و بهینه‌سازی سیستم‌های بازیابی اطلاعات داشته باشد.

فصل ۳

کارهای مرتبط

به طور کلی کارهایی که در حوزه بازیابی تخصص انجام شده را میتوان به چهار دسته کلی تقسیم کرد این چهار دسته شامل موارد زیر میباشد.

- شناسایی منابع داده
- استخراج داده
- طبقه بندی تخصص ها
- خودکار سازی فرآیند استخراج و طبقه بندی

در این مطلب نتیجه کارهای پیشین در دسته سوم یعنی طبقه بندی تخصص ها مورد استفاده قرار گرفته و با استفاده از روش های خوشه بندی که در کارهای گذشته معرفی شده، حوزه های مهارتی آینده افراد شناسایی میشود. در این بخش به توضیح کارهای پیشین انجام شده در بخش سوم پرداخته میشود.

طبقه بندی تخصص ها

در این بخش از پایان نامه، مروری بر تحقیقات پیشین در زمینه بازنمایی تخصص انجام شده است. این حوزه با استفاده از تکنیک های مختلفی برای استخراج و نمایش تخصص افراد از داده های موجود، توسعه یافته است. بر اساس مرور انجام شده، این تحقیقات به سه جنبه اصلی تقسیم می شوند: روش های بازنمایی، پشتیبانی زمانی، و پشتیبانی معنایی.

روش های بازنمایی تخصص

روش های مختلفی برای بازنمایی تخصص در متون علمی به کار گرفته شده اند که هر کدام بر اساس نوع داده ها و اهداف خاص تحقیقاتی طراحی شده اند. این روش ها را می توان به پنج دسته اصلی تقسیم کرد:

- **فراوانی اصطلاحات:** این روش بر مبنای تعداد دفعات تکرار اصطلاحات در متون علمی استوار است. در این رویکرد، هر واژه به عنوان یک بُعد در نظر گرفته می شود و بر اساس تعداد تکرار آن در متن، میزان تخصص فرد یا ارتباط متن با یک موضوع خاص ارزیابی می شود.

- **مدل‌های زبانی:** این مدل‌ها با استفاده از تحلیل احتمالی توزیع واژه‌ها در متون، یک مدل زبانی ایجاد می‌کنند که می‌تواند برای ارزیابی میزان شباهت متون جدید با متون قبلی به کار رود.
- **مدل‌های موضوعی:** این مدل‌ها توزیع احتمالی موضوعات در متون را تحلیل می‌کنند. یکی از معروف‌ترین روش‌ها در این دسته، Latent Dirichlet Allocation (LDA) است که برای شناسایی موضوعات اصلی در یک مجموعه داده به کار می‌رود.
- **گراف:** در این روش، داده‌های تخصص به صورت یک گراف نمایش داده می‌شوند. سپس با استفاده از الگوریتم‌های خاص مانند PageRank یا Random Walk، روابط و رتبه‌بندی تخصص‌ها تحلیل می‌شود.
- **روش‌های سفارشی:** این دسته شامل رویکردهای خاصی است که در آن‌ها از تکنیک‌های جایگزین برای بازنمایی تخصص استفاده می‌شود. به عنوان مثال، برخی مطالعات از داده‌های باز پیوندی (LOD) برای ساخت پروفایل‌های پژوهشگران استفاده می‌کنند.

پشتیبانی زمانی

تخصص افراد ممکن است با گذر زمان تغییر کند و این تغییرات می‌تواند به دلیل تغییر علاقه‌مندی‌ها یا تغییرات در حوزه‌های تحقیقاتی باشد. بنابراین، در تحلیل تخصص، باید زمان نیز به عنوان یک عامل مهم در نظر گرفته شود. تحقیقات در این زمینه به دو دسته اصلی تقسیم می‌شوند:

- **برش‌های زمانی:** در این رویکرد، تغییرات تخصص در طول زمان به صورت برش‌های زمانی مانند سال بررسی می‌شود. این روش به تحلیل تغییرات تدریجی در تخصص و پیش‌بینی علاقه‌مندی‌های آینده کمک می‌کند.
- **پیوسته:** این روش‌ها نیازی به برش‌های زمانی مشخص ندارند و تغییرات تخصص را به صورت پیوسته در طول زمان بررسی می‌کنند. این رویکردها با استفاده از برچسب‌های زمانی مرتبط با هر داده، تغییرات در تخصص افراد را تحلیل می‌کنند.

پشتیبانی معنایی

پشتیبانی معنایی به معنی استفاده از ساختارهای معنایی برای بهبود تحلیل و بازنمایی تخصص است. در حالی که بسیاری از مطالعات موجود از پشتیبانی معنایی استفاده نمی‌کنند، برخی از تحقیقات از هستی‌شناسی‌ها (Ontology) و پایگاه‌های دانش (Knowledge Base) برای این منظور بهره می‌برند. این ابزارها به تحلیل بهتر مفاهیم موجود در متون و ارتباطات معنایی بین آن‌ها کمک می‌کنند.

درخت موضوعی

بازیابی اطلاعات

تنتیب

بازیابی اطلاعات

فصل ۴

نتیجه‌گیری و کارهای آینده

ایده و محدوده کاری آینده

در این موضوع با تحلیل داده های خوشه بندی شده ابتدا یک نمایه از هر کاربر که در فرم های اینترنتی مرتبط فعالیت داشته، ساخته میشود. این نمایه توصیف میکند که چند درصد فعالیت کاری هر کاربر به یک موضوع خاص اختصاص داده شده است. سپس یک روش ارزیابی برای امتیازدهی به میزان دقت روش پیشبینی معرفی میشود

نتیجه گیری

کتاب‌نامه

[1] K. Balog, Y. Fang, M. de Rijke, P. Serdyukov, and L. Si, "Expertise retrieval," *Foundations and Trends® in Information Retrieval*, vol.6, no.2-3, pp.127-256, 2012.

[۲] مصطفی واحدی، "درختان پوشای کمینه دورنگی مسطح"، مجله فارسی نمونه، جلد ۱، صفحات ۲۲-۳۰،

آبان ۱۳۸۷.

واژه‌نامه انگلیسی به فارسی

Word کلمه
Word 2 کلمه ۲
Word 3 کلمه ۳

واژه‌نامه فارسی به انگلیسی

word 1 کلمه ۱
word 2 کلمه ۲
word 3 کلمه ۳

Abstract:

This is Abstract in English.

Keywords: Word1, Word2, Word3



Shahid Beheshti University

Faculty of Computer Science & Engineering

Semantic Matching Methods on Object Retrieval

By

Arash Dargahi Nobari

A THESIS SUBMITTED
FOR THE DEGREE OF
MASTER OF SCIENCE

Supervisor :

Dr. Mahmood Neshati

2019