

# A New Expert Finding Model Based on Term Correlation Matrix

Ehsan Pournoor<sup>1</sup> | Jalal Rezaenoor<sup>2</sup>

1. MSc in Information Technology Engineering; University of Qom  
e.pournoor@hotmail.com
2. [Corresponding Author] PhD in Industrial Engineering;  
Assistant Professor; University of Qom j.rezaee@qom.ac.ir

Iranian Journal of  
**Information  
Processing &  
Management**

Iranian Research Institute  
for Information Science and Technology  
(IranDoc)  
ISSN 2251-8223  
eISSN 2251-8231  
Indexed by SCOPUS, ISC, & LISTA  
Vol. 30 | No. 4 | pp. 1147-1171  
Summer 2015  
<https://doi.org/10.35050/JIPM010.2015.023>



**Abstract:** Due to enormous volume of unstructured information available on the Web and inside organizations, finding an answer to the knowledge needed in a short time is difficult. For this reason, besides search engines which do not consider users individual characteristics, recommender systems were created which use user's previous activities and other individual characteristics to help users find needed knowledge. Using recommender systems is increasing every day. By introducing expert people instead of recommending information to users, expert finder systems have provided this facility so that users ask their questions from experts. Having relation with experts not only causes information transition, but also transferring experiences and inceptions causes knowledge transition. In this paper we used university professors' academic resume as expert people profile and then proposed a new expert finding model that recommends experts to users query. We used Term Correlation Matrix, Vector Space Model and PageRank algorithm and proposed a new hybrid model which outperforms conventional methods. This model can be used in internet environment, organizations and universities that experts have resume dataset.

**Keywords:** Recommender Systems; Online Forums; Expert Finding; Vector Space Model

# مدل جدید پیشنهادگر فرد خبره با استفاده از ماتریس همبستگی کلمات

احسان پرنور<sup>۱</sup> | جلال رضایی نور<sup>۲</sup>

۱. کارشناسی ارشد مهندسی فناوری اطلاعات؛ دانشگاه قم e.pournoor@hotmail.com

۲. [پدیدآور رابط] دکتری مهندسی صنایع؛ استادیار؛ دانشکده فنی و مهندسی؛ دانشگاه قم j.rezaee@qom.ac.ir

## مقاله پژوهشی

دریافت: ۱۳۹۳/۱۰/۰۳

پذیرش: ۱۳۹۳/۱۱/۱۳

**چکیده:** با توجه به حجم عظیم اطلاعات بدون ساختار داخل سازمان‌ها و همچنین وب، یافتن پاسخ به نیاز دانشی در زمانی اندک، دشوار می‌باشد. به این منظور، علاوه بر موتورهای جستجو که در آنها ویژگی‌های فردی کاربران در نظر گرفته نمی‌شوند، سیستم‌های پیشنهادگر به‌وجود آمده‌اند و بر اساس رفتارهای پیشین کاربران سعی در کمک به دسترسی به دانش مورد نیاز آنها دارند. استفاده از این سیستم‌ها در وبسایت‌ها و شبکه‌های اجتماعی در حال گسترش است. سیستم‌های پیشنهادگر خبره نیز با پیشنهاد افراد خبره دارای دانش به‌جای ارائه اطلاعات، این امکان را فراهم کرده‌اند که کاربران سؤال‌های خود را از افراد خبره بپرسند. ارتباط با متخصصان باعث می‌شود که علاوه بر انتقال مطلب، با انتقال تجارب و دانسته‌ها، بینش و فهم نیز در مورد موضوع منتقل شود. ما در این مقاله از رزومه‌های اساتید دانشگاه به‌عنوان پروفایل افراد خبره استفاده کرده و مدل پیشنهادگر خبره جدیدی ارائه داده‌ایم که پرسش کاربر را دریافت کرده و افراد خبره را در زمینه پرسش او پیشنهاد می‌کند. ما با ترکیب ماتریس همبستگی کلمات، مدل فضای برداری و الگوریتم پیچ‌رنک مدلی ارائه کرده‌ایم که در مقایسه با روش‌های مرجع دقت بالاتری در شناخت افراد خبره دارد. این مدل در سیستم‌های دانشگاهی، داخل سازمانی و همچنین در محیط وب که متخصصان دارای اطلاعات رزومه باشند، کاربرد دارد.

**کلیدواژه‌ها:** پیشنهادگر خبره؛ فروم‌های آنلاین؛ مدل فضای برداری؛ ماتریس همبستگی

## داده‌های اطلاعات

فصلنامه علمی پژوهشی  
پژوهشگاه علوم و فناوری اطلاعات ایران  
(ایرانداک)

شاپا (چاپی) ۸۲۲۳-۲۲۵۱

شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱

نمایه در SCOPUS، JISC، LISTA و

jipm.irandoc.ac.ir

دوره ۳۰ | شماره ۴ | صص ۱۱۴۷-۱۱۷۱

تابستان ۱۳۹۴

<https://doi.org/10.35050/IJPM010.2015.023>



## ۱. مقدمه

با گسترش اینترنت و جوامع مجازی، روزبه‌روز بر حجم اطلاعات افزوده شده و دسترسی به اطلاعات مفید و مرتبط به کاری دشوار تبدیل شده است. در چنین شرایطی، وجود سیستمی که بتواند از میان حجم عظیمی از اطلاعات، اطلاعات مرتبط و مورد نیاز را بنا به نیاز و ویژگی‌های هر کاربر به وی پیشنهاد کند، به موضوعی بسیار مهم تبدیل شده است. سیستم‌های پیشنهادگر با بررسی رفتارهای پیشین کاربران در وبسایت‌ها، موضوعات مورد علاقه آنها را تشخیص داده و با پیشنهاد فیلم، موسیقی، کالا و ... راه پیدا کردن اطلاعات و کالاهای مورد نیاز را آسان می‌کنند. سیستم‌های حل مسئله از جمله سیستم‌های پیشنهادگر دانش هستند که دسترسی به دانش مورد نیاز را فراهم می‌کنند. وبسایت‌های پرسش و پاسخ، فروم‌های آنلاین و شبکه‌های اجتماعی روزبه‌روز رایج‌تر می‌شوند و مردم نیازهای اطلاعاتی خود را با استفاده از اینترنت و متخصصان دنیای مجازی رفع می‌کنند. با پیشرفت اینترنت و گسترش آن، اطلاعات به‌صورت آنلاین در دسترس همگان بوده و افراد، دانش خود را به اشتراک گذاشته و از دانش دیگران استفاده می‌کنند. جوامع مجازی<sup>۱</sup> نوع خاصی از شبکه‌های اجتماعی هستند که بیشتر در زمینه اشتراک اشتراک دانش و جستجوی نیاز اطلاعاتی مورد استفاده قرار می‌گیرند. این جوامع از افرادی تشکیل شده‌اند که در جهت افزایش درک و دانش خود در موضوعی خاص، در این جوامع عضو شده و با اشتراک دانش خود و دانش دیگران، دانسته‌های خود را افزایش می‌دهند. این افراد با مشارکت خود، منابع عظیم دانشی را فراهم می‌کنند که مورد استفاده همگان قرار می‌گیرد. امروزه، این جوامع علاوه بر مطالب متنی از رسانه‌های دیگری چون صوت، تصویر و عکس برای انتقال دانش استفاده می‌کنند. لذا، انتقال دانش تجربی بسیار راحت‌تر و صحیح‌تر از گذشته شده است. تعداد این جوامع روزبه‌روز افزایش پیدا می‌کند. علاوه بر اینترنت، اشتراک دانش در درون سازمان‌ها نیز به موضوعی بسیار مهم تبدیل شده است. سیستم‌های اشتراک دانش درون سازمانی، این امکان را فراهم کرده‌اند که کارکنان یک سازمان نیازهای اطلاعاتی یکدیگر را رفع نمایند. بیشتر شرکت‌های بزرگ مثل دل<sup>۲</sup>، مایکروسافت و آمازون از سیستم‌های اشتراک دانش و فروم‌های

1. virtual communities  
2. Dell

مباحثه‌ای برای افزایش میزان اطلاعات کارمندان خود استفاده می‌کنند (Wang et al. 2013). در جوامع مجازی، اشتراک دانش معمولاً به سه حالت انجام می‌شود: در حالت اول، کاربر سؤال خود را ثبت کرده و منتظر می‌ماند تا کاربران به آن پاسخ دهند. با اینکه مزیت این روش آن است که افراد دارای دانش به نیاز اطلاعاتی پاسخ‌گو هستند، ولی این روش معایبی نیز دارد. از جمله اینکه، ممکن است شخص پرس‌وجوکننده مدت زیادی منتظر بماند تا کسی به نیاز دانشی او پاسخ دهد و نیز ممکن است که سؤال کاربر بدون جواب باقی بماند. در حالت دوم، حجم بزرگی از اسناد و اطلاعات پالایش شده در بانک‌های اطلاعاتی ذخیره شده است و به محض پرس‌وجوی کاربر، اسناد و اطلاعات مرتبط با آن نمایش داده می‌شود. این حالت مشابه عملکرد موتورهای جستجو می‌باشد و دارای مزیت سرعت پاسخ‌گویی است. اما، چون سؤال کاربران و محتویات دانشی متخصصان به زبان طبیعی نوشته می‌شود، تشخیص محتوای مرتبط با سؤال کاربر از روی کلمات مشکل می‌باشد. حالت سوم، حالتی است که در آن ارتباط بین جستجوکننده و فراهم‌کننده دانش برقرار شود. این روش اساس کار سیستم‌های پیشنهادگر خبره می‌باشد که در آن خبره‌های دارای دانش به کاربران معرفی می‌شوند تا سؤال خود را از آنها بپرسند. تفاوت روش اول با سوم در این است که به جای اینکه سؤال کاربر به عموم کاربران عرضه شود، تنها به متخصصانی که توسط سیستم به‌عنوان افراد خبره در زمینه سؤال کاربر شناخته می‌شوند، ارائه می‌گردد که این به‌نوبه خود باعث افزایش سرعت و دقت پاسخ‌دهی می‌شود و با فراهم‌شدن تعاملات اجتماعی، اشتراک دانش آسان می‌گردد. نیاز اصلی این روش آن است که اطلاعات دانش، همان‌گونه که در اختیار پاسخ‌دهنده است، منتقل شود.

در این مقاله، ما مدلی جدید برای سیستم‌های پیشنهادگر خبره ارائه کرده‌ایم که با تحلیل داده‌های متنی رزومه اساتید دانشگاه، زمینه‌های تخصص آنها را پیش‌بینی کرده و به محض دریافت پرسش کاربران سیستم، خبره‌های دارای دانش در زمینه پرسش را تشخیص می‌دهد. مدل ارائه شده با استفاده از ماتریس همبستگی کلمات، گراف مشابهت اسناد (رزومه‌ها) را ایجاد کرده و نیز با استفاده از الگوریتم پیچ‌رنک<sup>1</sup> و مدل فضای

1. PageRank

برداری، متخصصان دارای دانش لازم را برای هر پرس و جو پیشنهاد می‌کند. این مدل در محیط‌های تحت وب، دانشگاهی و درون سازمانی که افراد متخصص دارای اطلاعات رزومه هستند، می‌تواند مورد استفاده قرار گیرد و خیره‌ها را در زمینه پرسش‌های کاربران پیشنهاد کند.

## ۲. پیشینه پژوهش

داده‌های خام پس از پردازش به اطلاعات تبدیل می‌شوند و معنا می‌یابند و زمانی که اطلاعات، تجربه، و بینش در مورد یک موضوع در اختیار شخصی قرار می‌گیرد، گفته می‌شود که شخص در آن موضوع دارای دانش است (Filemon & Uriarte 2008). بنابراین، دانش مفهومی جامع‌تر از اطلاعات بوده و انتقال آن از شخصی به شخص دیگر مشکل‌تر از انتقال اطلاعات است. انتقال فهم و بینش در مورد یک موضوع با استفاده از فقط متن کاری مشکل می‌باشد. اما زمانی که شخص آموخته‌های خود را به شخص دیگری می‌آموزد، با انتقال دانش و تجربه خود و با استفاده از گفتار، علایم، و نوشتار بهتر می‌تواند مفاهیم را منتقل سازد. از این رو، استفاده از سیستم‌های خبره یاب به جای سیستم‌های بازیابی متن بسیار مفیدتر خواهد بود، زیرا این سیستم‌ها به جای پاسخ‌گویی به سؤالات، افراد متخصص را جهت پاسخ‌گویی به نیاز دانشی کاربران معرفی می‌کنند. در سال‌های اخیر، با توجه به اهمیت سیستم‌های پیشنهادگر خبره، پژوهشگران زیادی جذب این حوزه شده‌اند. بیشتر مردم برای یافتن آنچه که به آن نیاز دارند، از موتورهای جستجو استفاده می‌کنند. موتورهای جستجو با پردازش آفلاین<sup>۱</sup> اطلاعات، صفحات وب را ساختار بندی و دسته‌بندی می‌کنند و به محض دریافت پرس و جوی کاربران با توجه به کلمات مشابه مورد استفاده در اسناد و نیز در پرس و جوی کاربر، صفحات وب را رتبه‌بندی کرده و به کاربر نمایش می‌دهند. موتورهای جستجو به عنوان سیستم‌های بازیابی اطلاعات، یافتن اطلاعات مورد نیاز را بسیار آسان می‌کنند، اما جزء سیستم‌های پیشنهادگر دانش به حساب نمی‌آیند، زیرا به ویژگی‌های افراد، پرسش‌های پیشین آنها، محل زندگی و رفتارهای آنها توجه نمی‌کنند و صرفاً مطالب مرتبط با پرسش کاربر را بازیابی می‌کنند و پیشنهادی در این زمینه ندارند. کاربران برای یافتن نیاز اطلاعاتی خود بایستی صفحات را مرور کرده و

1. offline

مرتبط‌ترین سند را پیدا کنند و این کار امری ناخوشایند، زمان‌بر و خسته‌کننده می‌باشد. آنها به سیستمی نیاز دارند که پرس‌وجوی خود را ثبت کنند و پاسخ خود را به‌صورت مختصر و مفید دریافت نمایند (Wang et al. 2013). سیستم‌های حل مسئله به‌وجود آمده‌اند تا به جای بازیابی تعداد زیاد اسناد، با ارائه پاسخ دقیق‌تر و کوتاه‌تر به کاربران کمک کنند که در زمان کمتری به دانش مورد نیاز خود دست یابند. این سیستم‌ها پاسخ‌ها را یا با تحلیل دقیق داده‌ها پیدا کرده و یا آنها را با استفاده از افراد خیره فراهم می‌کنند.

فروم‌ها از جمله سیستم‌های اشتراک دانش و حل مسئله هستند که کاربران با پُست‌ها و مباحث خود محتوای آنها را تشکیل می‌دهند. با توجه به موضوعات فروم‌ها، کاربران دانسته‌ها و سؤالات خود را به زبان خود و بدون استفاده از کلمات خاص به‌صورت زنجیروار پُست می‌کنند و محتوای ارزشمندی تولید می‌کنند که می‌تواند برای حل مسائل دیگران مورد استفاده قرار گیرد. علاوه بر اینترنت، در داخل سازمان‌ها نیز از فروم‌های خبری و مباحثه‌ای استفاده می‌شود (Updegrove et Pendergast 2006; Lichtner et al. 2009). اما، همچنان مشکل یافتن پاسخ درست وجود دارد. کاربرد بایستی اطلاعات فروم را مطالعه کرده و بهترین پاسخ را انتخاب کند که فرایندی زمان‌بر است. همچنین، در صورتی که شخص علاقه‌مندی برای پاسخ‌دهی وجود نداشته باشد، سؤالات بدون جواب باقی خواهد ماند. در مقاله لیو<sup>۲</sup> و همکاران یک مکانیزم پیشنهادگر اجتماعی برای اشتراک دانش در فروم‌های آنلاین ارائه شده است (Liu et al. 2012).

وب‌سایت‌های پاسخ به سؤال<sup>۳</sup> به‌عنوان جزئی از سیستم‌های حل مسئله<sup>۴</sup>، به این منظور به‌وجود آمده‌اند که به نیاز دانشی کاربران پاسخ دهند. این سیستم‌ها مطالب موجود در وب را کاوش کرده و اطلاعات را پالایش می‌کنند و جواب‌های کوتاهی را به کاربران ارائه می‌دهند. این سیستم‌ها روزبه‌روز رایج‌تر می‌شوند و دلیل آن هم این است که نه تنها کاربران سؤال خود را به زبان طبیعی خود می‌پرسند، آنها حتی می‌توانند سؤالات متفرقه مطرح نمایند (Liu et al. 2013)، اما توانایی پاسخ‌گویی در تمام زمینه‌ها را ندارند. پژوهشگران زیادی تلاش کرده‌اند تا با ایجاد سیستم‌های پاسخ به سؤال یافتن پاسخ صحیح

- 
1. forum
  2. Liu
  3. question answering systems
  4. problem solving systems

را در زمان اندک فراهم کنند (Chen et al. 2006; Zhao et al. 2009; Zhang et al. 2006; Zhao et al. 2009). وبسایت‌های پاسخ به سؤال، فروم‌های آنلاین و سیستم‌های خبره‌یاب از جمله سیستم‌های حل مسئله می‌باشند.

## ۱-۲. سیستم‌های خبره‌یاب

سیستم‌های خبره‌یاب سیستم‌هایی هستند که وظیفه آنها پیشنهاد متخصصان در حوزه یک موضوع می‌باشد. این سیستم‌ها پاسخ‌ها و پرسش‌های پیشین کاربران را ذخیره و پالایش کرده و حوزه تخصصی آنها را پیش‌بینی می‌کنند و به محض دریافت پرسش جدید، افراد خبره را که در این زمینه دارای تخصص می‌باشند، مشخص کرده و برای پاسخ‌گویی انتخاب می‌کنند. این کار باعث می‌شود که سؤالات به جای عرضه به عموم کاربران، به افراد دارای دانش ارائه شود و این به نوبه خود باعث افزایش سرعت و دقت پاسخ‌دهی می‌شود. علاوه بر این، ارتباط با متخصصان باعث می‌شود که خبره‌ها دانش، بینش، و تجربه خود را انتقال دهند و در این مسیر حتی از ارتباطات فیزیکی و غیرمجازی نیز استفاده کنند و یا در صورت عدم توانایی در پاسخ‌گویی، افراد خبره دیگری را معرفی نمایند. خبره‌یابی سنتی بیشتر در داخل سازمان‌هایی وجود دارد که در آنها اسناد و مخازن دانشی موجود است و در آنها دانش به صورت اسناد ساختاریافته ثبت شده و کیفیت اطلاعات بالا می‌باشد. برای مثال سیستم‌های آنسرگاردن<sup>۱</sup> (Ackerman & Malone 1990) و روش دمویر<sup>۲</sup> (Ackerman & McDonald 1996) و روش دمویر<sup>۲</sup> (Yimam & Kobsa 2003) برای یافتن متخصصان درون سازمان ارائه شده‌اند. در مقایسه با خبره‌یاب‌های داخل سازمان، در فروم‌های دانشی آنلاین<sup>۳</sup> ساختار دانش جهانی وجود ندارد. افراد به اندازه‌ای که در این فروم‌ها شرکت دارند و به اندازه‌ای که نیاز دانشی خود را حل کنند، به پرسش و پاسخ پرداخته‌اند. دانشی که در فروم‌های آنلاین وجود دارد، عمدتاً کیفیت بالایی ندارد و این به نوبه خود عملکرد سیستم‌های خبره‌یاب را دچار مشکل می‌سازد. سیستم‌های خبره‌یاب از روش‌های زیر برای پیدا کردن متخصصان استفاده می‌کنند: روش محتوایی، روش تحلیل شبکه و روش‌های ترکیبی.

1. AnswerGarden  
2. Demoir  
3. online

## ۲-۲. خبره‌یابی بر اساس تحلیل محتوا

در این تکنیک برای پیدا کردن متخصص، از روش‌های بازیابی اطلاعات برای یافتن مشابهت بین اسناد و پرس‌وجو استفاده می‌شود. خبره‌ها بر اساس مشابهتی که بین نوشتجات آنها وجود دارد، شناخته می‌شوند. برخی از پژوهشگران از روش پروفایلینگ<sup>۱</sup> برای یافتن متخصصان خود استفاده کرده‌اند. سیستم‌های یلوپیجز<sup>۲</sup>، گورو<sup>۳</sup>، ایتی‌ایت‌اولز<sup>۴</sup> از جمله سیستم‌هایی هستند که برای خبره‌یابی از پروفایلینگ استفاده می‌کنند (Wang et al. 2013). به این ترتیب که، برای هر متخصص پروفایلی شامل اطلاعات شخصی و مطالب مورد علاقه او ایجاد می‌گردد. این کار با کنترل صفحات مشاهده‌شده فرد متخصص جمع‌آوری شده و پس از آن، از این اطلاعات به عنوان حوزه تخصصی و برای محاسبه مشابهت حوزه دانشی وی استفاده می‌گردد. این پروفایل‌ها بر اساس کلماتی که در آنها استفاده شده، توصیف می‌شوند. برخی دیگر از پیشنهادگرهای خبره از نوشتجات متخصصان در فروم‌ها و سؤالات و پاسخ‌های آنها استفاده می‌کنند و با استفاده از کلمات به کار رفته در این نوشته‌ها حوزه تخصصی فرد را تشخیص می‌دهند. سیستم‌های خبره‌یاب زیادی وجود دارند که با استفاده از روش‌های بازیابی اطلاعات این مطالب را کاوش کرده و به صورت اتوماتیک تخصص افراد را مشخص می‌سازند (Krulwich & Burkey 1996; Streeter & Lochbaum 1988). از این مطالعات چنین استنباط می‌شود که درجه تخصص یک فرد به اسناد منتشرشده از او بستگی دارد (Balog et al. 2007; Balog et al. 2006). همچنین، پاسخ‌های افراد را که به عنوان پاسخ‌های منتخب و درست انتخاب شده‌اند، جمع‌آوری کرده و از تاریخچه پاسخ‌های منتخب برای محاسبه میزان مشابهت استفاده می‌کنند. در این مطالعات مدل‌هایی برای تعیین خبره‌ها با استفاده از مشابهت پرس‌وجو و اسناد خبره ارائه کرده‌اند. همچنین، در مقاله کرولیچ و همکاران، سیستم محتویاب<sup>۵</sup> که از تاریخچه پیغام‌های افراد برای شناسایی متخصصان استفاده می‌شود، توسعه داده شده است (Krulwich & Burkey 1996).

1. profiling
2. yellowpages.com
3. guru.com
4. 88owls.com
5. ContactFinder



معروف ترین مدل بازبایی اطلاعات که در مجموع نتایج بهتری نسبت به سایر مدل‌ها دارد، مدل فضای برداری است (Ricardo & Berthire 2011). در این مدل تمام اسناد و همچنین پرس و جوهای کاربران به عنوان برداری در فضای کلمات در نظر گرفته می‌شوند. برای مثال اگر  $n$  کلمه مجزا در کل مجموعه ما وجود داشته باشد، یک فضای  $n$  بُعدی خواهیم داشت که تمامی اسناد و همچنین پرس و جوی کاربران بر اساس سهمی که از این کلمات دارند، برداری در این فضا خواهند بود. مدل فضای برداری با استفاده از سنجش کسینوس زاویه بین بردار هر سند و بردار پرس و جو میزان مشابهت سند و پرس و جو را ارزیابی می‌کند و برای سنجش این مشابهت از وزن کلمات استفاده می‌شود. میزان مشابهت سند و پرس و جو با استفاده از رابطه (۱) مشخص می‌گردد:

$$sim(d_j, q) = \cos(\vec{d}_j, \vec{q}) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} \quad (1)$$

در معادله بالا،  $\vec{d}_j$  نشان دهنده بردار سند زام و  $\vec{q}$  نمایانگر بردار پرس و جو می‌باشد.

### ۳-۲. خیره‌یابی بر اساس روش‌های تحلیل شبکه

اخیراً برای بهبود الگوریتم‌های خیره‌یابی از روش‌های تحلیل گراف استفاده شده است. این مطالعات نشان داده‌اند که در محیط‌هایی مثل شبکه‌های اجتماعی که اشخاص با یکدیگر دارای ارتباطات اجتماعی هستند، استفاده از الگوریتم‌های بر اساس گراف نتایج بهتری داشته‌اند. آنها ادعا می‌کنند که انتقال دانش فردی از ارتباطات شخصی و غیررسمی ناشی می‌شود (Tobias & Finke 2006; Manning & Schutze 1999; Cooke et al. 2008). برخی از مطالعات از ایده زنجیره مارکو که ایده جدیدی در حوزه خیره‌یابی است، استفاده کرده‌اند (Lin et al. 2008). این مطالعات، خیره‌یابی را در شبکه‌های اجتماعی بررسی کرده و از اطلاعات ارتباطی مثل ارتباطات دوستی، ارتباطات پیگیری مطالب، پرس و جو کننده- پاسخ‌دهنده، فرستنده- گیرنده ایمیل و در محیط‌های اجتماعی آکادمیک مثل ارتباطات داشتن مقاله مشترک و هم‌رشته‌بودن برای یافتن تخصص فرد از روی تخصص رابطین آنها استفاده کرده‌اند. مطالعات نشان می‌دهد که استفاده از اطلاعات پیرامونی افراد در فعالیت‌های جستجوی دانش بسیار مهم است. در مقاله رومنی و همکاران بیان شده است

که ارتباط مثبتی بین تعداد پیام‌هایی که یک شخص و دیگران ارسال شده است و میزان شایستگی او وجود دارد (Romney et al. 1986). مطالعات سیستم‌های اطلاعاتی نشان می‌دهد که اطلاعات پیرامونی افراد و نه اطلاعات محتوایی و اسنادی آنها در فعالیت‌های جستجوی دانش بسیار مؤثر است. در روش‌های تحلیل گراف عمده‌تاً از الگوریتم‌های پیچ‌رنک و هیتس به‌عنوان روش‌های مرجع استفاده شده است. در مقاله ژانگ و همکاران شبکه ارتباطات فروم مباحثه‌ای بر اساس لینک‌های بین پرسش‌کننده و پاسخ‌دهنده مورد ارزیابی قرار گرفته است. آنها الگوریتم اکسپرتیس‌رنک<sup>۱</sup> را بر اساس الگوریتم پیچ‌رنک ارائه کرده و بر اساس اینکه چه کسی به چه کسی کمک کرده است، خبره‌یابی را انجام داده‌اند. آنها بیان کرده‌اند که الگوریتم پیچ‌رنک در مقایسه با سایر روش‌ها نتایج بهتری در بازیابی دارد (Zhang et al. 2007). ژرژیک و آگیشتن با استفاده از ایجاد لینک ارتباطات مابین فرستنده و گیرنده در شبکه ارسال ایمیل و مابین پرسش‌کننده و پاسخ‌دهنده در فروم یاهو، شبکه اجتماعی تشکیل داده و با تعیین شناسه<sup>۲</sup> برای هر فرد، میزان خبرگی او را با استفاده از الگوریتم هیتس<sup>۳</sup> مشخص کرده‌اند (Jurczyk & Agichtein 2007).

پیچ‌رنک، روشی مبتنی بر ساختار دموکراتیک وب است که از شبکه گسترده لینک‌ها جهت تعیین مرتبه و امتیاز یک صفحه معین استفاده می‌کند (Lawrence et al. 1999). در این الگوریتم از ساختار لینک بین صفحات وب برای تشخیص میزان اهمیت صفحات وب استفاده شده است. با استفاده از لینک‌های موجود بین صفحات وب، گرافی تشکیل می‌شود که در آن نودهایی<sup>۴</sup> که دارای تعداد لینک ورودی زیادتری هستند، به‌عنوان نودهای بااهمیت شناسایی می‌شوند. همچنین، اسنادی که لینک ورودی از یک نود بااهمیت دارند، دارای اهمیت بیشتری هستند. با استفاده از این ایده، الگوریتم پیچ‌رنک با محاسبه تعداد لینک‌های ورودی برای هر صفحه و وزن‌دهی صفحات، درجه اهمیت صفحات وب را رتبه‌بندی می‌کند. اگر  $u$  را یک صفحه وب فرض کنیم و  $F_u$  مجموعه صفحاتی که  $u$  به آنها لینک دارد و  $B_u$  مجموعه صفحاتی باشند که به  $u$  لینک دارند، در این صورت تعداد لینک‌های خروجی از  $u$  برابر  $|F_u|$  خواهد بود. اگر  $c$  نیز به‌عنوان فاکتور

1. ExpertiseRank  
2. authority  
3. HITS  
4. nodes

نرمال‌سازی فرض شود، در این صورت رتبه صفحه  $u$  با استفاده از رابطه (۲) تعیین می‌گردد:

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v} \quad (2)$$

## ۲-۴. خبره‌یابی بر اساس روش‌های ترکیبی

در مطالعات اخیر از روش‌های ترکیبی برای خبره‌یابی استفاده شده است. در این مطالعات، هم از روش تحلیل محتوایی و هم از روش‌های تحلیل شبکه استفاده شده است (Mattox et al. 1999; Kautz et al. 1997; Zhang et al. 2007). از این روش، بیشتر در شبکه‌های اجتماعی که افراد، هم دارای پروفایل و اطلاعات مکتوب و هم دارای روابط اجتماعی می‌باشند، استفاده شده است. استفاده از روش‌های ترکیبی نتایج بهتری را نسبت به روش‌های پیشین داشته است. لیو و همکاران با استفاده از تاریخچه نوشتجات کاربران و اطلاعات شخصی آنها و با ایجاد پروفایل دانش برای هر کاربر و تحلیل محتوایی آن، امتیاز هر خبره را مشخص کرده و سپس با تحلیل لینک‌های پرسش و پاسخ و با استفاده از ارتباط پرس‌وجو و پرس‌وجوهای پیشین خبره‌یابی را انجام می‌دهند (Liu et al. 2013). آنها همچنین، از پاسخ‌های منتخب افراد به‌عنوان بهترین پاسخ استفاده کرده و شهرت افراد را محاسبه کرده‌اند. همچنین، وانگ و همکاران مدل اکسپرت‌رنک<sup>۱</sup> را بر اساس الگوریتم پیچ‌رنک ارائه نموده و در آن از ترکیب تحلیل محتوایی و تحلیل شبکه استفاده کرده‌اند و نتایج آنها از هر دو روش محتوایی مجرد یا تحلیل شبکه مجرد بهتر بوده است (Wang et al. 2013). در مقاله‌ای دیگر سردیکو و همکارانش الگوریتمی را ارائه نمودند که در آن از هر دو روش بر پایه سند و ارتباطات غیرمستقیم سند- نویسنده استفاده کرده‌اند (Serdyukov et al. 2007). آنها همچنین از اطلاعات دیگری مثل لینک صفحات وب<sup>۲</sup>، ساختار و بازخورد کاربران استفاده کرده‌اند. ارزیابی‌ها نشان می‌دهند که روش آنها از الگوریتم‌هایی که تنها اسناد را در نظر می‌گرفتند، بهتر بوده است. کمپل و همکاران و دامور دو الگوریتم بر پایه هیتس را در حوزه ارتباطات ایمیل ارائه کردند. الگوریتم از هر

1. ExpertRank  
2. Hyperlink

دو محتوا و الگوهای ارتباطی آنها استفاده کرده است تا متخصصان را رتبه‌بندی کند. ارزیابی‌ها نشان می‌دهد که نتایج آنها نسبت به روش مبتنی بر سند دقیق‌تر بوده است (D'Amore 2004; Campbell et al. 2003). فو و همکاران تلاش مشابهی برای ترکیب روش محتوایی و بر پایه شبکه اجتماعی برای خبره‌یابی انجام دادند (Fu et al. 2007). آنها یک شبکه اجتماعی از ارتباطات ایمیل و رویارویی افراد در صفحات وب ایجاد کرده و بهبود عملکرد را در مقایسه با روش‌های پایه نشان دادند. در مقاله ژانگ و همکاران نیز از یک روش ترکیبی برای خبره‌یابی استفاده شده است (Zhang et al. 2005). آنها از مقالات علمی برای تشخیص خبره‌ها استفاده کرده‌اند. آنها همچنین، از روش‌های متن‌کاوی و تحلیل محتوا برای رتبه‌بندی اولیه استفاده کرده و پس از آن با استفاده از الگوریتمی مشابه، پیچ‌رنک لینک ارتباطات را مورد تحلیل قرار داده‌اند.

#### جدول ۱. مقایسه بین روش‌های متفاوت در سیستم‌های خبره‌یابی

مزایا	معایب	روش
استفاده از داده‌های پروفایل و تخصصی افراد و نوشتجات آنها	عدم استفاده از ارتباطات اجتماعی و ویژگی‌های رابطه‌ای	تحلیل محتوا (استفاده از مدل‌های باز یابی اطلاعات)
عدم استفاده از داده‌های تخصصی افراد مثل نوشتجات و داده‌های پروفایل آنها	استفاده از ارتباطات اجتماعی مابین متخصصان و تشخیص میزان خبرگی از طریق رفتارهای اجتماعی و اطلاعات پیرامونی افراد	تحلیل شبکه‌ای (استفاده از الگوریتم‌های تحلیل گراف)
استفاده از اطلاعات اجتماعی، بازخورد کاربران، داده‌های متنی و محتوایی افراد	-----	روش ترکیبی (استفاده مشترک از مدل‌های باز یابی اطلاعات و الگوریتم‌های تحلیل گراف)

#### ۳. روش پژوهش

برای شبیه‌سازی مدل، ما از داده‌های رزومه علمی اساتید دانشگاهی به‌عنوان اسناد

پروفایل متخصصان استفاده کرده‌ایم. این رزومه‌ها توسط تیم متن‌کاوی متشکل از دانشجویان کارشناسی و کارشناسی ارشد دانشگاه و از طریق اینترنت گردآوری شده است. در مرحله‌ی نمایه‌سازی اسناد، ابتدا کلمات زاید موجود (از قبیل حروف اضافه و افعال کمکی) در این رزومه‌ها را حذف نموده‌ایم و سپس تمامی کلمات با استفاده از الگوریتم پورتر<sup>۱</sup> به‌عنوان معروف‌ترین ریشه‌یاب زبان انگلیسی، مورد ریشه‌یابی قرار گرفته و پس از آن با استفاده از روش وزن‌دهی Tf-Idf وزن کلمات در اسناد مشخص شده است. در مدل ارائه‌شده از مفاهیم ماتریس همبستگی کلمات، مدل فضای برداری، و الگوریتم پیچ‌رنک استفاده شده است. این سه مفهوم در سیستم‌های بازیابی اطلاعات پیش از این مورد استفاده قرار گرفته‌اند. پیش از این، الگوریتم پیچ‌رنک در مواردی مورد استفاده قرار گرفته است که کاربران سیستم دارای ارتباطات اجتماعی مانند دوستی و یا ارسال پیام باشند. اما در فضای کاری این مقاله چنین ارتباطاتی موجود نمی‌باشد. این است که ایده‌ی استفاده‌شده در این مقاله استخراج ارتباطات معنایی بین رزومه‌ها با استفاده از ماتریس همبستگی کلمات و سپس تحلیل این ارتباطات با الگوریتم پیچ‌رنک و ترکیب نتایج آن با مدل فضای برداری بوده است.

### ۳-۱. ماتریس همبستگی کلمات<sup>۲</sup>

در بازیابی اطلاعات به روش فازی از ماتریس همبستگی کلمات برای بهبود پرس‌وجوی کاربر استفاده می‌شود. ماتریس همبستگی کلمات، ماتریسی است مربعی که ابعاد آن را تمام کلمات کلیدی موجود در مجموعه‌ی اسناد تشکیل می‌دهند. سپس بر اساس میزان حضور هم‌زمان کلمات در اسناد مختلف، میزان همبستگی آنها مشخص می‌گردد. فرض کنید مجموعه‌ی کلمات موجود در مجموعه‌ی اسناد، مجموعه‌ی  $\{t_1, t_2, t_3, \dots, t_n\}$  باشد، در این صورت ماتریس همبستگی کلمات به‌صورت زیر خواهد بود:

1. Porter  
2. Term Correlation Matrix

$$\begin{matrix} & t_1 & \cdots & t_n \\ \begin{matrix} t_1 \\ \vdots \\ t_n \end{matrix} & \begin{pmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & 1 & \vdots \\ c_{n,1} & \cdots & a_{n,n} \end{pmatrix} & & \end{matrix}$$

شکل ۱. ماتریس همستگی کلمات

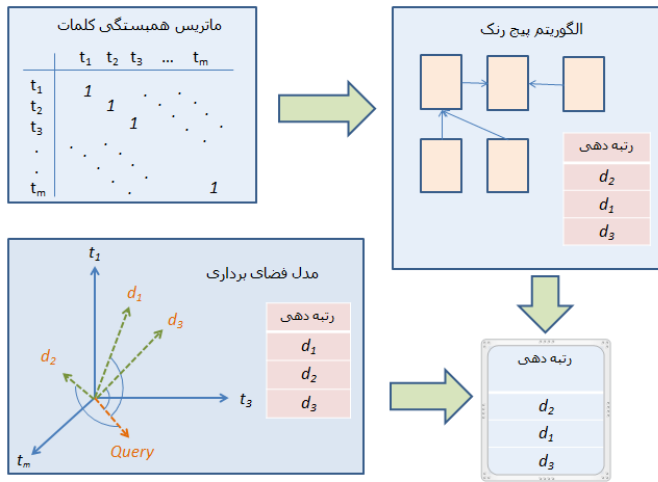
در این ماتریس مقادیر  $C_{ij}$  نشان‌دهنده میزان همستگی بین دو کلمه  $t_i$  و  $t_j$  است و مقدار آن با استفاده از رابطه (۳) محاسبه می‌شود:

$$C_{i,l} = \frac{n_{i,l}}{n_i + n_l - n_{i,l}} \quad (3)$$

در رابطه بالا،  $n_i$  نشان‌دهنده تعداد اسنادی است که کلمه  $t_i$  در آنها حضور دارد و به همین ترتیب،  $n_l$  تعداد اسنادی را مشخص می‌کند که کلمه  $t_l$  در آنها وجود داشته است. همچنین،  $n_{i,l}$  نشان‌دهنده تعداد اسنادی است که دو کلمه  $t_i$  و  $t_l$  با هم در آنها حضور داشته‌اند. پس از محاسبه مقادیر همستگی، به ازای هر دو کلمه موجود در مجموعه، ماتریس همستگی نشان‌دهنده میزان همستگی کلمات به هم خواهد بود.

### ۳-۲. مدل ارائه شده

ما در این مقاله با ترکیب روش‌های محتوایی و روش تحلیل شبکه و با استفاده از ماتریس همستگی کلمات، مدلی جدید برای یافتن خبره‌ها ارائه کرده‌ایم. شکل ۲، بیانگر معماری مدل ارائه شده است. در این مدل با استفاده از ماتریس همستگی کلمات، رابطه معنایی مابین اسناد را ایجاد می‌کنیم. سپس با استفاده از الگوریتم پیچ‌رنک، درجه اهمیت اسناد مشخص شده و بر اساس این درجه اهمیت، اسناد رزومه به صورت نزولی مرتب می‌شوند. در مرحله آخر رتبه حاصل از مدل فضای برداری و امتیاز اهمیت هر سند با هم ترکیب خطی شده و رتبه نهایی هر سند را مشخص می‌کند.



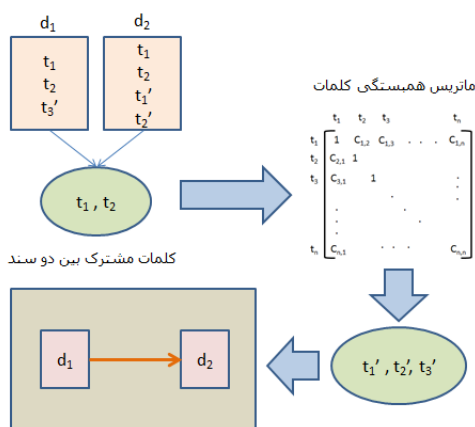
شکل ۲. مدل خبره‌یابی ارائه‌شده

در این مدل نحوهٔ ایجاد ارتباطات معنایی اسناد رزومه با استفاده از ماتریس همبستگی کلمات، ایده‌ای جدید می‌باشد که پیش از این مورد استفاده نبوده است. مدل ارائه‌شده متشکل از سه مرحله می‌باشد:

#### مرحلهٔ اول: استخراج ارتباطات اسناد

در مدل ارائه‌شده برای مشخص کردن ارتباط بین اسناد از کلمات مشترک موجود در آنها استفاده شده است. فرض کنید دو سند (رزومه)  $d_1$  و  $d_2$  دارای کلمات مشترکی باشند که در هر دو سند تکرار شده‌اند. هر چه میزان کلمات مشترک در بین دو سند زیاد باشد، نشان‌دهندهٔ میزان ارتباط معنایی بین این دو سند (ارتباط هم‌تخصصی بین دو فرد) است. فرض اتخاذ شده بر این پایه است که متخصصانی که در یک حوزه فعالیت می‌کنند، دارای رزومهٔ اطلاعات مشابه هستند و در اسناد رزومهٔ آنها از کلمات مشابهی استفاده شده است. با مشخص شدن کلمات مشترک و با استفاده از ماتریس همبستگی کلمات، به ازای کلمات مشترک، تعدادی کلمات همبسته با کلمات مشترک (کلماتی که میزان همبستگی آنها با کلمات مشترک زیاد است) مشخص می‌شوند. برای مثال، فرض کنید دو کلمه  $t_1, t_2$  کلمه‌های مشترک بین دو رزومهٔ  $d_1$  و  $d_2$  باشند که در هر دو سند وجود دارند. در این

صورت، این دو سند به دلیل وجود کلمات مشترک با هم ارتباط معنایی دارند. فرض کنید با استفاده از ماتریس همستگی کلمات، کلمه‌های  $d_1'$ ،  $d_2'$ ،  $d_3'$  به عنوان کلمه‌های همبسته به  $t_1$  و  $t_2$  پیدا شوند. در این مرحله هر کدام از اسناد که شامل تعداد زیادی از کلمات همبسته استخراج شده از ماتریس همستگی باشند، به عنوان سند با اهمیت تر شناسایی می شود. ارتباط بین دو سند با استفاده از لینکی که جهت آن به سمت سند مهم تر است، نشان داده می شود. شکل ۳، مراحل استخراج ارتباطات هم تخصصی افراد را نشان می دهد.



شکل ۳. روش استخراج ارتباطات معنایی اسناد

### مرحله دوم: رتبه‌دهی متخصصان با استفاده از الگوریتم پیچرنک

پس از اینکه ارتباطات بین دو سند موجود در مجموعه مورد بررسی قرار گرفت، گراف جهت دار حاصل از ارتباطات معنایی آنها تشکیل می شود. گراف حاصل از مرحله پیش با استفاده از الگوریتم پیچرنک مورد پالایش قرار گرفته و رتبه اسناد مشخص می شود. الگوریتم پیچرنک از تعداد لینک‌های ورودی اسناد استفاده کرده و بر اساس آن رتبه‌بندی اسناد را انجام می دهد. سندی در این الگوریتم دارای رتبه بالاتری خواهد بود که شامل کلمات بیشتر و مرتبط تری باشد. به بیان دیگر، سندی که در یک حوزه جامع باشد و بیشتر لغت‌های مورد استفاده در یک حوزه را دربرگیرد، دارای اهمیت بیشتر و رتبه آن توسط الگوریتم پیچرنک بالاتر خواهد بود.



### مرحله سوم: ترکیب با مدل فضای برداری

در این مرحله پرس و جوی کاربر بدون توجه به نتایج پیش و با استفاده از مدل فضای برداری مورد پالایش قرار گرفته و مشابهت آن با تمامی اسناد موجود در مجموعه محاسبه می شود. با توجه به حضور وزنی کلمات در اسناد و در پرس و جوی کاربر در مدل فضای برداری، مشابهت هر سند با پرسش کاربر سنجیده می شود. در این سنجش اسنادی مرتبط تر خواهند بود که شامل کلمات استفاده شده با نسبت وزنی بالایی باشند و اگر سندی هیچ کدام از کلمات پرس و جو را نداشته یا حتی اگر کلمات هم معنی آنها را داشته باشد، به عنوان یک سند نامرتبط تلقی خواهد شد. همچنین، دو مرحله قبل مستقل از پرس و جوی کاربر بودند و رتبه اسناد تنها با استفاده از کلمات موجود در آنها و ماتریس همبستگی کلمات مشخص می شد. بنابراین، در این مرحله هدف این است که کاستی مدل فضای برداری با استفاده از نتایج دو مرحله پیش بهبود داده شود و کاستی ناشی از عدم وابستگی به پرس و جو با استفاده از مدل فضای برداری رفع گردد. پس از اجرای مدل فضای برداری، هر سند شامل رتبه ای خواهد شد که میزان مشابهت آن را به پرس و جو نشان می دهد. در این مرحله، دو رتبه پیچ رنک و فضای برداری هر سند طبق رابطه (۴) با هم ترکیب خطی شده و رتبه نهایی هر سند نتیجه می شود.

$$TotalRank(d_i) = VSRank(d_i) \times 0.5 + PageRank(d_i) \times 0.5 \quad (4)$$

مجموعه رزومه های گردآوری شده برای شبیه سازی، شامل ۳۵۳ رزومه علمی در ۱۵ رشته تحصیلی و از رزومه های اساتید دانشگاه های کشورهای انگلیس (دانشگاه منچستر، دانشگاه کمبریج، دانشگاه لستر<sup>۱</sup>، کانادا (دانشگاه یورک)، استرالیا (دانشگاه بالارات)، ایران (دانشگاه بوعلی سینا، دانشگاه صنعتی شریف، دانشگاه قم، دانشگاه تهران، دانشگاه صنعتی همدان) بوده است. تعداد کل کلمات استفاده شده (به غیر از کلمات زاید) ۶۲۳۸ کلمه بوده و زبان نگارش رزومه ها همگی انگلیسی بوده است. پس از ایجاد ماتریس همبستگی کلمات، همبستگی بین هر دو کلمه موجود در مجموعه اسناد، مشخص شده و کلمات با میزان همبستگی بالا به عنوان کلمات همبسته قوی مشخص شده اند. سپس، با تعیین لینک ارتباطات بین رزومه ها، گراف ارتباطات معنایی رزومه ها تشکیل شده است.

1. University of Leicester

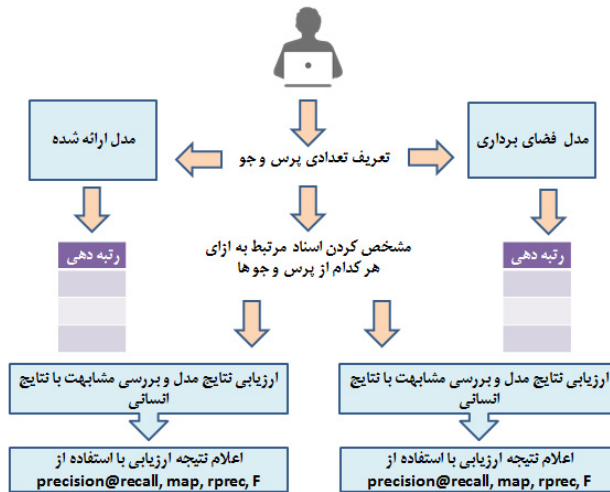
این گراف شامل ۳۵۳ نود و ۷۵۴۹ لینک می‌باشد که توسط الگوریتم پیچ‌رنک و با مقدار همگرایی ۰/۰۰۱ و با احتمال شروع ۰/۸۵ مورد تحلیل قرار گرفته و رتبه اسناد مشخص شده است. بعد از اجرای الگوریتم پیچ‌رنک بر روی گراف مشابهت اسناد، مدل بازیابی فضای برداری روی داده‌ها اجرا شده و رتبه‌بندی حاصل از نتایج آن با نتایج حاصل از الگوریتم پیچ‌رنک ترکیب خطی گردیده و نتایج نهایی مشخص شده‌اند.

#### ۴. یافته‌های پژوهش

در سیستم‌های بازیابی اطلاعات، زمانی که مدل جدیدی ارائه می‌گردد، برای ارزیابی میزان دقت مدل ارائه‌شده، آن را با یک مدل مرجع مقایسه کرده و با کمک ارزیابی انسانی می‌سنجند. در این ارزیابی‌ها از معیارهای ارزیابی دقت در بازخورد، F Measure، E Measure، MAP، RPREC و سایر معیارهای ترکیبی دقت و بازخورد استفاده می‌شود. در مقاله حاضر نیز از این تکنیک استفاده شده است. جهت ارزیابی مدل ترکیبی ارائه‌شده و مدل فضای برداری (مدل مرجع)، همان‌طور که در شکل ۴ مشاهده می‌شود، ابتدا پرسش‌هایی توسط عوامل انسانی (اساتید و دانشجویان کارشناسی ارشد تیم متن‌کاوی گروه فناوری اطلاعات) طراحی شده و رزومه‌های مرتبط با پرسش‌ها نیز توسط این عوامل مشخص گردیده‌اند. مثلاً برای پرسش «کامپیوتر، داده‌کاوی و بازیابی اطلاعات» عوامل انسانی تمام رزومه‌ها را مطالعه کرده و از بین آنها رزومه افرادی را که در این زمینه‌ها تخصص دارند، مشخص و رتبه‌دهی کرده‌اند. این رتبه‌دهی دقیق‌ترین نوع رتبه‌دهی است، زیرا انسان بسیار دقیق‌تر از ماشین توانایی تشخیص رزومه‌های مرتبط با پرسش را دارد. سپس، بدون توجه به نتایج انسانی، مدل مرجع (مدل فضای برداری) و مدل ارائه‌شده (که ترکیبی از مدل فضای برداری، الگوریتم پیچ‌رنک و ماتریس همبستگی کلمات است) هر کدام به صورت مجزا نتایج خود را برای پرسش ارائه می‌کنند. در این مرحله برای پرسش مطرح شده سه رتبه‌دهی مشخص شده است: رتبه‌دهی انسانی، مدل مرجع و مدل ارائه‌شده. اکنون باید مشخص گردد که آیا رتبه‌دهی مدل مرجع به رتبه‌دهی انسانی نزدیک‌تر است یا رتبه‌دهی مدل ارائه‌شده؟ بنابراین، نتایج هر کدام از مدل مرجع و مدل ارائه‌شده به صورت مجزا با نتایج انسانی مقایسه می‌شوند. در این مقایسه معیارهای ارزیابی P@R

1. precision at recall

دقت در بازخورد)، MAP، RPREC و F Measure (میانگین هارمونیک) اعدادی را به این دو مدل نسبت می‌دهند که میزان موفقیت هر یک را در میزان مشابه بودن به نتایج انسانی مشخص می‌کند. در نهایت، هر کدام از مقادیر  $P@R$ ، MAP، RPREC و F برای مدل مرجع و مدل پیشنهادی با هم مقایسه می‌شوند و هر کدام که مقدار بالاتری داشت، نتایج دقیق‌تری دارد.

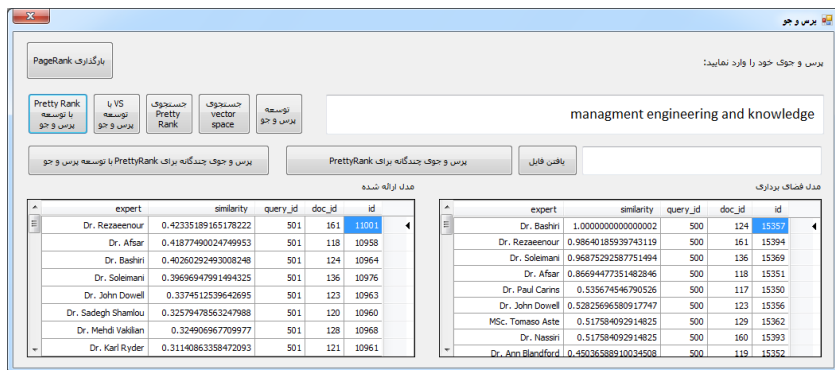


شکل ۴. نحوه ارزیابی سیستم خبره یاب

در سیستم‌های بازیابی، مقدار دقت را برای مقادیر استاندارد بازخوردهای استاندارد (۵ درصد، ۱۰ درصد، ۲۰ درصد، ۳۰ درصد و ۴۰ درصد بازیابی) می‌سنجند. معیارهای دیگر بازیابی مثل میانگین هارمونیک ترکیبی از دو مقدار دقت و بازخورد هستند. برای آزمایش نتایج، برنامه کاربردی پرتی‌رنک<sup>۱</sup> شامل: نمایه‌ساز<sup>۲</sup>، ریشه‌یاب<sup>۳</sup>، جستجوگر<sup>۴</sup> و ارزیاب<sup>۵</sup> توسعه داده شده است. این برنامه، داده‌ها (رزومه‌ها) را به صورت

1. PrettyRank
2. Indexer
3. Stemmer
4. Searcher
5. Evaluator

فایل‌های متنی گرفته و مراحل نمایه‌سازی، ریشه‌یابی و ذخیره‌سازی نمایه‌ها را در بانک‌های اطلاعاتی به صورت آفلاین انجام داده است. برنامه پرتی‌رنک شامل بخش‌هایی برای دریافت پرس‌وجو و گزارش نتایج جستجو می‌باشد. این برنامه، پرسش کاربر را دریافت نموده و افراد خیره در زمینه این پرسش را بر اساس اطلاعات رزومه آنها (که توسط مدل ارائه‌شده پالایش شده‌اند) مشخص می‌کند. شکل ۵، برنامه پرتی‌رنک را نمایش می‌دهد. این برنامه از نظر میزان دقت بازیابی اسناد با نرم‌افزار عمومی Lemur مورد مطابقت قرار گرفته است.



شکل ۵. نرم‌افزار پرتی‌رنک

برای ارزیابی آزمایش، تعداد ۹۵ جمله پرسش توسط ۲ نفر از اساتید گروه و تعدادی از دانشجویان کارشناسی ارشد مرتبط با حوزه متن‌کاوی مشخص شده و رزومه‌های مرتبط با هر کدام از پرسش‌ها توسط عوامل انسانی ذکر شده مشخص شده‌اند. در انتخاب رزومه‌های مرتبط، علاوه بر کلمات مشترک استفاده شده در متن پرس‌وجو و رزومه‌ها، رزومه‌هایی که توسط عوامل انسانی به عنوان متخصصان مرتبط با پرس‌وجوها مشخص شده‌اند، شناسایی گردیده‌اند. پس از اجرای نرم‌افزار با استفاده از معیارهای ذکر شده، میزان شباهت رتبه‌دهی مدل ارائه‌شده و مدل مرجع با رتبه‌دهی انسانی مورد سنجش قرار گرفته است که نتایج آن در جداول (۲) و (۳) ارائه شده است.

آنچه که در حوزه سیستم‌های بازیابی بسیار مهم است، این است که اسناد مرتبط در بازیابی‌های اولیه حضور داشته باشند. همان‌طور که در نمودار (۶) مشاهده می‌شود، مقدار

دقت در ۵ درصد اول بازیابی برای مدل ارائه شده برابر ۹۱/۲ درصد می باشد که این مقدار برای مدل فضای برداری برابر ۸۳/۳ درصد است. با توجه به نتایج به دست آمده، مشاهده می شود که نتایج مدل ارائه شده به ازای مقادیر استاندارد بازیابی (۵ درصد، ۱۰ درصد، ۲۰ درصد، ۳۰ درصد و ۴۰ درصد بازیابی) از نتایج مدل فضای برداری بهتر می باشد. این نتایج نشان دهنده این موضوع می باشد که استخراج روابط معنایی بین اسناد رزومه می تواند در شناسایی موضوع یک سند مفید واقع شود. بنابراین، در شناسایی مشابهت یک سند با یک پرس و جو، علاوه بر سنجش میزان مشابهت بر اساس کلمات مشترک استفاده شده در آنها، می توان از کلمات مترادف (همبسته قوی) آنها و ارتباطات معنایی بین اسناد استفاده کرد. در این پژوهش، با استخراج این ارتباطات معنایی و تشکیل گراف مشابهت اسناد و تحلیل آن، میزان دقت پیشنهاد فرد خبره افزایش داده شده است. به عبارت دیگر، سیستم ارائه شده در این مقاله پیشنهادهای دقیق تری به کاربران دارد و در ازای نیاز دانشی کاربران، افراد خبره مرتبط تری را پیشنهاد می کند.

## ۵. نتیجه گیری

با توجه به افزایش میزان حجم داده های بدون ساختار داخل سازمان ها و همچنین روی وب، یافتن دانش مورد نیاز از میان حجم انبوه داده ها امری بسیار مشکل می باشد. سیستم های پیشنهادگر دانش، فروم ها و سیستم های پاسخ به پرسش جهت آسان کردن راه دسترسی به دانش مورد نیاز و پاسخ گویی به نیاز اطلاعاتی کاربران به وجود آمده اند. این سیستم ها در مقایسه با موتورهای جستجو به جای پیشنهاد صفحات وب رتبه بندی شده به کاربران سعی دارند جواب پرسش را بدهند. کاربران با ثبت سؤالات خود در فروم ها، منتظر کاربران دیگر می مانند تا متخصصان دارای دانش، آنها را راهنمایی کنند. سیستم های پاسخ به سؤال با ایده پاسخ دهی کوتاه و مفید با استفاده از مخازن دانشی ثبت شده، دسترسی به دانش مورد نیاز را بهبود داده اند. ادبیات پیشین سیستم های مدیریت دانشی بر این باورند که پیشنهاد متخصص، به عنوان منبع به وجود آورنده دانش، بسیار مفیدتر از پیشنهاد اطلاعات است. متخصصان با انتقال دانش و تجربه خود باعث انتقال مفاهیم شده و این، به مراتب بسیار بهتر از انتقال اطلاعات خواهد بود. سیستم های پیشنهادگر خبره، متخصصان به وجود آورنده دانش را به کاربر جستجو کننده معرفی می کنند تا ارتباط بین

پرسش‌کننده و پاسخ‌دهنده را برقرار سازند و به این وسیله سرعت و دقت دسترسی به دانش را فراهم آورند. سیستم‌های پیشنهادگر خبره از تحلیل محتوای پروفایل متنی و سوابق متخصصان، تحلیل ارتباطات آنها و ترکیب این دو، برای یافتن متخصصان در یک حوزه استفاده می‌کنند. این سیستم‌ها در فضای وب و اینترنت، همچنین درون سازمان‌ها جهت اشتراک دانش مورد استفاده قرار گرفته‌اند.

ما در این مقاله با استفاده از ماتریس همبستگی کلمات و ترکیب آن با مدل فضای برداری و با استفاده از الگوریتم پیچ‌رنک مدلی ترکیبی ارائه کرده‌ایم که در مقایسه با روش فضای برداری (مدل مرجع) نتایج دقیق‌تری در بازیابی داشته است. ما با استفاده از ماتریس همبستگی کلمات ارتباطات معنایی بین اسناد را استخراج کرده و گراف ارتباطات اسناد را تشکیل داده و با استفاده از الگوریتم پیچ‌رنک، اسناد را رتبه‌بندی می‌کنیم. با ترکیب خطی این نتایج با نتایج حاصل از مدل فضای برداری، رتبه‌بندی نهایی حاصل می‌شود. این نتایج نشان می‌دهند که مدل ارائه‌شده دقت بیشتری نسبت به مدل فضای برداری دارد و افراد مرتبط‌تری را برای پرسش کاربران معرفی می‌کند. این نکته بیانگر این موضوع است که با استفاده از استخراج روابط معنایی اسناد می‌توان میزان دقت را در استخراج اسناد مرتبط به پرس و جوی کاربر افزایش داد. در این مقاله، نرم‌افزاری کاربردی توسعه و ارائه شده است که از مدل ارائه‌شده استفاده کرده و به محض دریافت پرسش کاربران، افراد متخصص دارای دانش را (که قبلاً اطلاعات رزومه آنها در سیستم ذخیره و نمایه‌سازی شده) به آنها پیشنهاد می‌کند. برای کارهای آینده در این زمینه می‌توان اقدامات زیر را انجام داد:

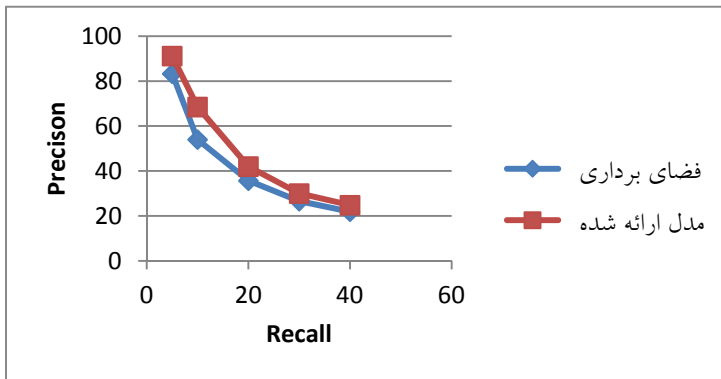
- ◇ استفاده از مدل‌های بازیابی دیگر مثل مدل‌های شبکه عصبی، مدل‌های احتمالی و ...
- ◇ استفاده از مدل‌های تحلیل شبکه دیگر همانند الگوریتم هیتس
- ◇ تغییر مدل برای پشتیبانی از زبان‌های دیگر مثل فارسی
- ◇ استفاده از ضرایب وزنی متفاوت در ترکیب خطی
- ◇ استفاده از روش‌های ترکیب دیگر مثل: فیلترینگ

جدول ۲. مقایسه دقت در بازخورد برای دو مدل فضای برداری و مدل ارائه شده

مدل	p@5	p@10	p@20	p@30	p@40
فضای برداری	۸۳/۳	۵۴	۳۵/۷	۲۶/۶	۲۱/۹
مدل ارائه شده	۹۱/۲	۶۸/۵	۴۲	۳۰	۲۴/۸

جدول ۳. مقایسه مدل فضای برداری با مدل ارائه شده با استفاده از معیارهای دیگر

مدل	F	MAP	R-PREC
فضای برداری	۲۱/۷	۴۴/۳	۵۲/۶
مدل ارائه شده	۲۲/۱	۵۱/۳	۶۳/۱



شکل ۶. دقت در بازخورد برای مدل فضای برداری و مدل ارائه شده

### فهرست منابع

- Ackerman, M. S., and D. McDonald. 1996. *Answer Garden 2: merging organizational memory with collaborative help*. Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, Boston, MA. (pp.: 97-105)
- Ackerman, M. S., and T. W. Malone. 1990. *Answer Garden: a tool for growing organizational memory*. Proceedings of the ACM SIGOIS and IEEE CS TCOA Conference on office Information Systems, Cambridge, MA. (pp.: 31-39)
- Baeza-Yates R., Ribeiro-Neto B., *Modern Information Retrieval: The Concepts and Technology behind Search*, 2<sup>nd</sup> Edition, Boston, USA: ACM Press, 2011.

- Balog, K., L. Azzopardi, and M. de Rijke. 2006. *Formal models for expert finding in enterprise corpora*. Proceedings of the 29<sup>th</sup> annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY. (p.: 50)
- Balog, K., T. Bogers, L. Azzopardi, M. de Rijke and A. van den Bosch. 2007. *Broad expertise retrieval in sparse data environments*. Proceedings of the 30<sup>th</sup> Annual International ACM SIGIR Conference, Amsterdam, The Netherlands. (pp.: 551-558)
- Campbell, C. S., P. P. Maglio, A. Cozzi, and B. Dom. 2003. *Expertise identification using email communications*. Proceedings of the 12th International Conference on Information and Knowledge Management, New Orleans, LA.
- Chen, W., Q. Zing, and W. Liu. 2006. *A user reputation model for a user-interactive question answering system*. Proceedings of the 2<sup>nd</sup> international conference on semantics, knowledge, and grid., (p.: 40)
- Cooke, R. M., S. ElSaadany, and X. Huang. 2008. On the performance of social network and likelihood-based expert weighting schemes. *Reliability Engineering and System Safety* 93 (5): (pp.: 745–746).
- D'Amore, R. 2004. *Expertise community detection*. Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK. (pp.: 498–499)
- Filemon, A., and Uriarte Jr. 2008. *Introduction to Knowledge Management*, Jakarta, Indonesia: ASEAN Foundation .
- Fu, Y., R. Xiang, Y. Liu, M. Zhang, and S. Ma. 2007. *Finding experts using social network analysis*. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Fremont, CA. (pp.: 77–80)
- Hu, M., and A. H. Elamy. 2007. *Speeding up QA: An index structure for question queries*. Canadian conference on electrical and computer engineering. Columbia, Canada. (pp.: 1215–1218)
- Jurczyk, P., and E. Agichtein. 2007. *Discovering authorities in question answer communities by using link analysis*. CIKM'07: Proceedings of the 16th international conference on information and knowledge management. Lisboa, Portugal: ACM Press.
- Kautz, H., B. Selman, and M. Shah. 1997. Referral web: combining social networks and collaborative filtering. *Communications of the ACM* 40 (3): 63–65.
- Krulwich, B., and C. Burkey. 1996. *Contactfinder agent: answering bulletin board questions with referrals*, The 13<sup>th</sup> National Conference on Artificial Intelligence, Portland, Or.
- Krulwich, B., C. Burkey. 1996. *Contactfinder agent: answering bulletin board questions with referrals*, The 13<sup>th</sup> National Conference on Artificial Intelligence, AAAI 96, IAAI 96, Portland, Oregon, Volume 1.
- Lawrence, Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford: InfoLab.
- Lederer, A. L., D. J. Maupin, M. P. Sena, and Y. Zhuang. 2000. The technology acceptance model and the world wide web. *Decision Support Systems* 29 (3): 269-282.



- Lichtner, V., A. P. Kounkou, A. Dotan, J. P. Kooken, and N. A. M. Maiden. 2009. *An online forum as a user diary for remote workplace evaluation of a work integrated learning system*. Proceedings of the 27<sup>th</sup> international conference on extended abstracts on human factors in computing systems, Boston, MA. (pp.: 2599–2970)
- Lin, C., H. Zhou, Z. Huang, and W. Wang. 2008. REC: *A novel model to rank experts in communities*. Proceedings of the 9th international conference on Webpage information management, Zhangjiajie, China. (pp.: 301–308)
- Liu, Duen-Ren, Yu-Hsuan Chin, Wei-Chen Kao, and Hsiu-Wen Wang. 2013. Integrating expert profile, reputation and link analysis for expert finding in question-answering websites, *Information Processing & Management* 49 (1), (pp. : 312-329).
- Liu, Duen-Ren, X., G. A. Wang, A. Johri, M. Ahou, and W. Fan. 2014. *Harnessing global expertise: a comparative study of expertise profiling methods for online communities*, *Information Systems Frontiers*, 16 (4), (pp.: 715-727).
- Manning, C. D., and H. Schutze. 1999. *Foundations of statistical natural language processing*. Cambridge: MIT Press.
- Mattox, D., M. Maybury, and D. Morey. 1999. *Enterprise expert and knowledge discovery*. Proceedings of the 8th International Conference on Human-Computer Interaction, Munich, Germany. (pp.: 23–27)
- Pendergast, M. 2006. An analysis tool for the assessment of student participation and implementation dynamics in online discussion forums. *ACM SIGITE Newsletter* 3 (2): 10–17.
- Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. Culture as consensus: a theory of culture and informant accuracy. *American Anthropologist* 88 (2): 313-338.
- Serdyukov, P., H. Rode, and D. Hiemstra. 2007. University of Twente at the TREC 2007 *Enterprise Track: modeling relevance propagation for the expert search task*. The 16<sup>th</sup> Text Retrieval Conference (TREC 2007) Enterprise Track, Gaithersburg, MD.
- Streeter, L., and K. Lochbaum. 1988. *An expert/expert-locating system based on automatic representation of semantic structure*. Proceedings of the Fourth Conference on Artificial Intelligence Applications, San Diego, CA. (pp.: 380-388)
- Streeter, L., K. E. Lochbaum. 1988. *An expert/expert-locating system based on automatic representation of semantic structure*. Proceedings of the 4<sup>th</sup> Conference on Artificial Intelligence Applications, (pp. :345-350). IEEE
- Tobias, M. P., and I. Finke. 2006. SELaKT-social network analysis as a method for expert localisation and sustainable knowledge transfer. *Journal of Universal Computer Science* 10 (6): 291–701.

- Updegrave, D. A., S. B. Smith, and W. R. Bollentin. 1988. *CCNEWS: An online forum for newsletter editors*. Proceedings of the 16<sup>th</sup> annual ACM SIGUCCS conference on user services. Long Beach, CA, USA. (pp.: 351–358).
- Wang, G. Alan, Jian Jiao, Alan S. Abraham, Weiguo Fan, and Ahongju Ahang. 2013. ExpertRak: A topic-aware expert finding algorithm for online knowledge communities, *Decision Support Systems*, 54 (3), (pp.: 1442-1451).
- Yimam-Seid, D., and A. Kobsa. 2003. *Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach, Sharing Experiences: Beyond Knowledge Management*. Cambridge: MIT Press.
- Yung-Ming, Li, Liao Tzu-fong, and Lai Cheng-Yang. 2012. A social recommender mechanism for improving knowledge sharing in online forums, *Information Processing & Management*, 48 (5), (pp. : 978-994).
- Zhang, J., M. S. Ackerman, and L. Adamic. 2007. *Expertise networks in online communities: structure and algorithms*. Proceedings of the 16<sup>th</sup> International World Wide Web Conference (WWW), Banff, Canada.
- Zhang, J., Tang, J., & Li, J. (2007). Expert finding in a social network. In *Advances in Databases: Concepts, Systems and Applications*, Springer Berlin Heidelberg, (pp.: 1066-1069).
- Zhang, P., Wu, C., Wang, C., & Huang, X. 2006. *Personalized question answering system based on ontology and semantic web*. Proceedings of the 4<sup>th</sup> IEEE international conference on industrial informatics. Singapore. (pp.: 1046–1051).
- Zhao, Z., S. Feng, Y. Liang, and Q. Zeng. 2009. *Mining user's interest from interactive behaviors in QA system*. Proceedings of the 1<sup>st</sup> international workshop on education technology and computer science. Wuhan, Hubei, China. (pp.: 1025–1029).