

Expertise Retrieval: Foundations, Methods and Models

Hashem Atapour

PhD in Information and Knowledge Science; University of Tehran;
Corresponding Author hashematapour@ut.ac.ir

Fatima Fahimnia

PhD in Information and Knowledge Science; Associate
Professor; University of Tehran fahimnia@ut.ac.ir

Nader Naghshineh

PhD in Information and Knowledge Science; Assistant
Professor; University of Tehran nnaghsh@ut.ac.ir

Received: 21, May 2015

Accepted: 25, Aug. 2015

Iranian Journal of
**Information
Processing and
Management**

Abstract: Expertise Retrieval (ER) is defined as finding experts in different subject areas as well as identifying people expertise area(s). For almost two decades, ER within knowledge-based environments has attracted attention of large and medium-sized organizations. It has given rise to a strong community of researchers, especially in Computer and Information Science. It seems, however, that ER has remained neglected among Iranian researchers. So, addressing the foundations, methods and models of ER, current paper intends to take steps in making scientific community and organization managers familiar with main concepts of this area.

This paper is an opinion paper based on library method.

Supporting knowledge management is an important application of ER systems. In these systems, by avoiding the challenging characteristics of expertise concept, researchers simplify the problem of ER and limit its scope to textual documents accumulation and discovering of documents-persons associations. Several models have been presented for discovering of above-mentioned associations, among which Balog's models had important role in promotion of ER field. Document-centric model offered by Balog and his colleagues is one of the most successful models that is the basis for subsequent models.

This paper addresses the issue of Expertise Retrieval and its foundations, methods and models. So, it could be of interest to information professionals, information retrieval area researchers, large and medium-sized organizations' managers, and other knowledge-based environments.

Iranian Research Institute
for Science and Technology

ISSN 2251-8223

eISSN 2251-8231

Indexed in SCOPUS, ISC, & LISTA

Vol. 31 | No. 3 | pp: 651-673

Spring 2016

Keywords: Expertise Retrieval, Information Retrieval, Expert Finding, Expert profiling, Knowledge Management, Models

بازیابی تخصص:

مروری بر مبانی، روش‌ها و مدل‌ها

هاشم عطاپور

دکتری علم اطلاعات و دانش‌شناسی؛
دانشگاه تهران؛

پدیداآور رابط hashematapour@ut.ac.ir

فاطمه فهیم‌نیا

دکتری علم اطلاعات و دانش‌شناسی؛
دانشیار؛ دانشگاه تهران fahimnia@ut.ac.ir

نادر نقشبند

دکتری علم اطلاعات و دانش‌شناسی؛
استادیار؛ دانشگاه تهران nnaghsh@ut.ac.ir



دریافت: ۱۳۹۴/۰۲/۳۱ | پذیرش: ۱۳۹۴/۰۶/۰۳ | مقاله برای اصلاح به مدت ۳ روز نزد پدیداآوران بوده است.

چکیده: بازیابی تخصص به امر یافتن افراد خبره در حوزه‌های موضوعی گوناگون و یا شناسایی حوزه(ها)ی تخصصی افراد می‌پردازد. این حوزه از دهه ۱۹۹۰ میلادی مورد توجه سازمان‌های بزرگ و متوسط و محیط‌های دانش‌مدار قرار گرفته، به طوری که یک اجتماع پژوهشی قوی از پژوهشگران مختلف، به خصوص پژوهشگران علوم رایانه و اطلاعات، حول آن شکل گرفته است. اما به نظر می‌رسد این حوزه در ایران چندان مورد توجه واقع نشده است. از این رو، مقاله حاضر قصد دارد با پرداختن به مبانی، روش‌ها و مدل‌های بازیابی تخصص، در جهت آشنایی بیشتر جامعه علمی و مدیران سازمان‌ها با مفاهیم این حوزه گام بردارد.

مقاله حاضر یک مقاله نظری است که با استفاده از روش کتابخانه‌ای انجام شده است.

در کنار سایر کاربردها، یکی از کاربردهای مهم نظام‌های بازیابی تخصص، پشتیبانی از مدیریت دانش در سازمان‌هاست. در این نظام‌ها، پژوهشگران با دوری جستن از جنبه‌های چالشی مفهوم تخصص، مسئله بازیابی تخصص را به انباشت اسناد متنی و کشف روابط افراد با آن اسناد تقلیل می‌دهند. مدل‌های متعددی برای کشف روابط بین اسناد و افراد ارائه شده‌اند که مدل‌های ارائه شده توسط «بالوگ و همکارانش» سهم به‌سزایی در پیشبرد حوزه بازیابی تخصص داشته‌اند. مدل سند-محور آن‌ها یکی از موفق‌ترین مدل‌هاست که مبانی بسیاری از مدل‌های بعدی قرار گرفته است.

این مقاله برای نخستین بار در ایران به طور نسبتاً جامع به مبحث بازیابی تخصص، مدل‌ها و مفاهیم بنیانی آن پرداخته است. بنابراین، می‌تواند مورد توجه

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در SCOPUS، ISI، LISTA و JIPM
jipm.irandoc.ac.ir
دوره ۳۱ | شماره ۳ | صص ۶۵۱-۶۷۳
بهار ۱۳۹۵

متخصصان علم اطلاعات و پژوهشگران حوزه بازیابی اطلاعات، مدیران سازمان‌های بزرگ و متوسط، و سایر محیط‌های دانش‌مدار قرار گیرد.

کلیدواژه‌ها: بازیابی تخصص، بازیابی اطلاعات، تشکیل پروفایل خبرگان، خبره‌یابی، مدیریت دانش، مدل‌ها

۱. مقدمه

در سال ۱۹۷۸، دولت ایالت متحده پیشنهادی را برای کنترل کاغذ در ادارات دولتی مطرح کرد و این پیشنهاد برای کاهش کاغذبازی به صورت قانون در آمد (حسن زاده ۱۳۸۴). این رویداد، به عنوان نقطه شروع توجه جدی به مدیریت اطلاعات یا همان ذخیره و بازیابی اطلاعات در سازمان‌ها و ادارات به حساب می‌آید. لیکن، آنچه که اکنون در حوزه مدیریت اطلاعات در جریان است، متفاوت از آن چیزی است که در اواخر دهه ۱۹۸۰ مطرح بوده است. امروزه، سرمایه‌گذاری و مدیریت بهینه در حوزه مدیریت اطلاعات، نه به علت صرفه‌جویی در مصرف کاغذ یا لزوم خرید تجهیزات اداری برای ذخیره و بازیابی اطلاعات درون‌سازمانی، بلکه به خاطر ارزش رقابتی و تعیین‌کننده اطلاعات در پیشبرد اهداف سازمانی مطرح است (همان). پژوهش «فلدمن و شرمین» بر اهمیت دسترسی به اطلاعات در سازمان تأکید می‌کند. طبق یافته‌های پژوهش مذکور، ۷۶ درصد از مدیران سازمان‌ها اطلاعات را بسیار حیاتی قلمداد کرده‌اند. همچنین، ۶۰ درصد از مدیران سازمان‌ها احساس می‌کنند که محدودیت‌های زمانی و فقدان آگاهی در مورد چگونگی یافتن اطلاعات، از دستیابی کارکنان به اطلاعات مورد نیازشان جلوگیری می‌کند. آن‌ها بیان می‌کنند که عدم دستیابی به اطلاعات مرتبط، به اتخاذ تصمیمات ضعیف، کاهش بهره‌وری، کاهش فروش و انجام کارهای تکراری منجر می‌شود. پژوهشگران مذکور پس از انجام سه مطالعه موردی، به تخمین هزینه‌هایی می‌پردازند که سازمان به واسطه عدم یافتن اطلاعات متحمل می‌شود: یک سازمان با ۱۰۰۰ کارمند، به واسطه جستجو در میان اطلاعات نامرتب، عدم موفقیت در یافتن اطلاعات موجود، یا تولید مجدد اطلاعات یافت‌نشده، در حدود ۲/۵ الی ۳/۵ میلیون دلار به هدر می‌دهد. بنابراین، به کارگیری نظام اطلاعاتی در سازمان‌های متوسط و بزرگ از اهمیت بالایی برخوردار است (Feldman & Sherman 2003).

با استقرار و توسعه حوزه مدیریت دانش در اوایل دهه ۱۹۹۰ میلادی، کانون توجه بر توسعه نظام‌های اطلاعاتی معطوف گردید که از جستجوی تخصص پشتیبانی کنند؛ یعنی نظام‌هایی که افراد صاحب دانش در حوزه موضوعی خاصی را شناسایی و ارائه کنند یا حوزه‌های تخصصی هر فرد را سیاهه نمایند. همچنین، امروزه سازمان‌ها با مسائل چندبُعدی مواجه می‌شوند که پرداختن به

آن‌ها مستلزم تفکیک موضوع به اجزاء کوچک‌تر، شناسایی و ترکیب گروهی از خبرگان با تخصص‌های مختلف است؛ امری که از آن به‌عنوان بازیابی تخصص چندرشته‌ای یاد می‌شود (Haruechaiyasak & Kongthon 2008). این عوامل منجر به ایجاد نظام‌هایی شد که به آن‌ها نظام‌های بازیابی تخصص^۱ گفته می‌شود. نظام‌های اولیه^۱ بازیابی تخصص معمولاً برای ورود داده‌های تخصص به خود خبرگان متکی بودند، به‌طوری که خبرگان داده‌های مربوط به مهارت و دانش خود را بر طبق کلیدواژه‌های از پیش تعریف‌شده وارد نظام می‌کردند. این امر، زمان‌بر و مستلزم کار زیادی بود، پروفایل‌های ایجادشده خیلی زود تازگی خود را از دست داده و نیازمند روزآمدسازی مستمری بودند. همچنین، این امکان وجود داشت که کارکنان با آگاهی از استفاده احتمالی کارفرمایان از داده‌های واردشده، در توصیف تخصص خود دچار افراط (به‌دلیل ترس از دست‌دادن شغل) یا تفریط (به‌منظور جلوگیری از افزایش مسئولیت‌ها و وظایف) شوند (MacDonald 2009). در نتیجه، حرکت به سمت فناوری‌های هوشمندی که فرایند ایجاد و روزآمدسازی پروفایل‌های خبرگان را به‌صورت خودکار در بیاورند، آغاز شد. در این میان، پژوهشگران بازیابی اطلاعات با ارائه مدل‌های گوناگون در توسعه ادبیات پژوهشی نظام‌های بازیابی تخصص و همچنین، توسعه محصولات کاربردی این حوزه نقش بسزایی ایفاء نمودند. ارائه مدل‌های گوناگون با هدف بهبودبخشیدن به عملکرد این نظام‌ها صورت می‌گیرد: نظام بازیابی تخصص هر اندازه بتواند در پاسخ به پرس و جوها، خبرگان مربوط بیشتری را بازیابی نموده و خبرگان مربوط را در رتبه‌های بهتری از خبرگان نامربوط قرار دهد، عملکرد مطلوب‌تری خواهد داشت. همچنین، توانمندی نظام بازیابی تخصص در ارائه یک سیاهه دقیق از حوزه‌های تخصصی هر خبره، دیگر معیار عملکرد مطلوب این نظام‌ها محسوب می‌شود. تلاش در جهت تعدیل مدل‌های موجود یا ارائه مدل‌های جدید برای رسیدن به مطلوب‌ترین عملکرد، این حوزه پژوهشی را به حوزه‌ای فعال و پویا تبدیل نموده و توجه بسیاری از پژوهشگران را به خود جلب کرده است. در سویی دیگر، مدل‌های توسعه‌یافته برای بازیابی تخصص در سازمان‌ها، محدود به سازمان‌ها نمانده و در محیط‌های دیگر نیز مورد استفاده قرار گرفته‌اند: یافتن اساتید راهنما و مشاور در محیط‌های دانشگاهی (Alarfaj & et al. 2012)، یافتن نویسندگان خبره در حوزه‌های موضوعی خاص (Bogers, kox, & Bosch 2008; Efron 2009)، یافتن وبلاگ‌نویس‌های خبره در حوزه‌های موضوعی خاص (Macdonald & Ounis 2008; Balog, de Rijke, and Weerkamp 2008)، یافتن داوران خبره برای سپردن داوری مقالات رسیده به کنفرانس‌ها و نشریات علمی (Mimno &

1. expertise retrieval systems

McCallum 2007) نمونه‌ای از به‌کارگیری مدل‌های بازیابی تخصص در محیط‌های دیگر می‌باشد. با وجود اینکه نظام‌های بازیابی تخصص از دهه ۱۹۹۰ میلادی در خارج از ایران به‌طور جدی مورد توجه پژوهشگران قرار گرفته و یک اجتماع پژوهشی قوی، که اعضای آن عمدتاً از متخصصان علوم رایانه و اطلاعات هستند، در این حوزه شکل گرفته است، به‌جرات می‌توان گفت که در کشور ما حوزه مذکور چه از جانب متخصصان علوم رایانه و چه از جانب متخصصان علم اطلاعات و دانش‌شناسی مورد غفلت قرار گرفته است؛ به‌طوری که بجز پژوهش «پرنور و رضایی» (زودآیند) کار پژوهشی در این حوزه صورت نگرفته است. از این‌رو، در این مقاله ضمن تبیین مفاهیم اساسی نظام‌های بازیابی تخصص، به معرفی مدل‌های مشهور بازیابی تخصص پرداخته می‌شود تا زمینه‌ساز آشنایی بیشتر پژوهشگران ایرانی با ادبیات حوزه مذکور و آغاز پژوهش‌های جدی در آن باشد. همچنین، مقاله حاضر به آشناسدن سازمان‌ها و محیط‌های دانش‌مدار کشورمان با نظام‌های بازیابی تخصص کمک نموده و جایگاه برجسته این گونه نظام‌ها را در مدیریت اطلاعات سازمانی تبیین می‌کند. بقیه محتوای مقاله حاضر به این شرح سازماندهی می‌شود: در بخش دوم به انگیزه‌های جستجوی خبرگان اشاره می‌شود؛ در بخش سوم به چالش مفهوم تخصص و چگونگی مواجهه با آن در بازیابی تخصص پرداخته می‌شود؛ در بخش چهارم دو وظیفه مهم نظام‌های بازیابی تخصص شامل خبره‌یابی^۱ و تشکیل پروفایل خبرگان^۲ تشریح می‌شوند؛ در بخش پنجم مدل‌های مشهور بازیابی تخصص ارائه می‌گردد؛ در بخش ششم مجموعه آزمون‌های خبره‌یابی^۳ معرفی می‌شوند؛ و در بخش هفتم به چگونگی انجام قضاوت ربط بر روی خبرگان بازیابی شده پرداخته می‌شود. این مقاله با نتیجه‌گیری در بخش هشتم به پایان می‌رسد.

۲. تخصص: مفهومی چالش‌انگیز

تخصص، مفهومی چالش‌انگیز است، چرا که مفهومی رسمی و جاافتاده نبوده و هر کس تعریف متفاوتی از آن دارد. مقدار دانشی که یک فرد باید داشته باشد تا به‌عنوان یک خبره شناخته شود، مشخص نیست. گفته می‌شود که افراد به‌طور میانگین حداقل به ده سال تجربه نیاز دارند تا در یک حوزه معین به‌عنوان خبره شناخته شوند (Serdyukov & Heimstra 2008). همچنین، از تخصص به‌عنوان دانش ضمنی یاد می‌شود، دانشی که در ذهن افراد نهفته بوده و دسترسی به آن دشوار است (Balog & et al. 2012). این چالش‌ها، بازیابی تخصص را با مشکل مواجه می‌کند. لیکن پژوهشگران نظام‌های بازیابی تخصص خود را درگیر این چالش‌ها نکرده و از ملاحظات

1. expert finding

2. expert profiling

3. expert finding test collections

شناختی و اجتماعی اجتناب می‌کنند. آن‌ها برای یافتن خبرگان یا شناسایی حوزه‌های تخصصی خبرگان به‌طور ساده بر اسناد و شواهد متنی، که از آن‌ها به‌عنوان شواهد تخصص یاد می‌شود، اتکاء نموده و از ارتباطات موجود بین افراد و اسناد برای تصمیم‌گیری در مورد خبرگی فرد یا شناسایی حوزه(های) تخصصی فرد استفاده می‌کنند. بدین ترتیب، هرچقدر فرد با اسناد یک حوزه موضوعی ارتباطات زیاد و محکمی داشته باشد، بر روی او به‌عنوان خبره آن موضوع بیشتر تکیه می‌کنند.

هرچند پژوهشگران با تمرکز بر ارتباطات بین افراد و اسناد برای استخراج و رتبه‌بندی خودکار خبرگان یا شناسایی حوزه‌های تخصصی آن‌ها، خود را از قید و بند چالش‌های شناختی و اجتماعی خبره‌یابی رها می‌کنند، لیکن خود این فرایند با چالش‌هایی روبه‌روست. بنابراین، هنگام استفاده از این روش برای خبره‌یابی یا شناسایی حوزه‌های تخصصی خبرگان باید به چالش‌های مذکور واقف بود و در جهت کنترل آن‌ها تدابیر لازم را اتخاذ نمود. نمونه‌ای از این چالش‌ها عبارت‌اند از:

- ◇ خبرگان کاندیدا معمولاً به‌صورت آیتم‌های قابل بازیابی بازنمون نمی‌شوند: خبرگان به‌طور غیرمستقیم از طریق ارتباط اسناد با آن‌ها شناسایی می‌شوند. کشف این ارتباطات که شامل نویسندگی سند یا مورد استناد قرار گرفتن در سند است، از بازیابی اسناد بازنمایی شده دشوارتر می‌باشد.
- ◇ اسامی خبرگان اغلب ابهاماتی دارد: ذکر اسامی افراد ممکن است ناقص بوده و نامی واحد به چند فرد تعلق داشته باشد.
- ◇ شواهد تخصص از منابع غیرهمگون می‌آید که همه آن‌ها به یک اندازه مهم نیستند: وزن یک ایمیل کوتاه با یک استاندارد فنی متفاوت است.
- ◇ تعیین قوت ارتباط بین یک خبره کاندیدا و شاهد متنی تخصص وی یک تصمیم پیچیده است (همان).

۳. انگیزه‌های جستجوی خبرگان

«یمام-سید و کوسا» پس از انجام مصاحبه‌هایی با پژوهشگران یک مؤسسه پژوهشی، دو انگیزه اصلی جستجوی خبرگان را چنین بیان می‌کنند: (۱) فرد خبره به‌عنوان منبع اطلاعات، (۲) فرد خبره به‌عنوان کسی که قادر است نقش سازمانی یا اجتماعی را انجام دهد (Yimam-Seid & Kobsa 2003).

کاربران ممکن است یک خبره را به‌عنوان منبع اطلاعاتی مکمل یا جایگزین سایر منابع مورد

جستجو قرار دهند. نمونه‌ای از سناریوهایی که در آن، کاربران ممکن است یک فرد خبره را به‌عنوان منبع اطلاعاتی مورد جستجو قرار دهند، به‌شرح زیر شناسایی شده‌اند:

الف) دسترسی به اطلاعات غیرمستند. همه اطلاعاتی که در سازمان هست، به‌طور کامل مستند نمی‌شود. بسیاری از اطلاعات مهم از طریق گذراندن دوره کارورزی، تجربه‌اندوزی و مکالمات غیررسمی کسب می‌شود. در بسیاری از موقعیت‌های پیش‌آمده جدید، اطلاعات مستندشده به‌ندرت کمک‌کننده است. همچنین، در برخی مواقع، اطلاعات مورد نیاز بنا به دلایل مختلف اقتصادی، اجتماعی و سیاسی و ... در دسترس عموم قرار نمی‌گیرد.

ب) تصریح نیاز اطلاعاتی. اطلاعاتی که کاربران نیاز دارند، اغلب صریح و مشخص نیست. بنابراین، لازم است که آن‌ها با یک فرد خبره برای مشخص کردن دقیق نیاز اطلاعاتی خود مشورت کنند.

پ) اتکاء به تخصص دیگران. کاربران اغلب می‌خواهند تلاش و وقت کمتری را برای یافتن اطلاعات مورد نیاز صرف کنند. این امر موجب می‌شود که آن‌ها از خبرگان برای انتخاب اطلاعات مفید از بین حجم عظیم اطلاعات استفاده کنند.

ت) نیاز به تفسیر. در بسیاری از مواقع، کاربران به خود اطلاعات علاقه‌ای ندارند، بلکه به کاربردها یا تفاسیر آن علاقه‌مندند. همچنین، در برخی مواقع آن‌ها قادر به درک اطلاعات بازیابی شده نیستند. در چنین شرایطی آن‌ها به خبرگان برای تفسیر یا فهم اطلاعات متوسل می‌شوند.

ث) نیاز به اجتماعی کردن. کاربران ممکن است به‌جای تعامل با اسناد و کامپیوترها، نیاز اطلاعاتی خود را با انسان‌ها مطرح کنند.

دسته دوم عوامل انگیزاننده جستجوی خبرگان، نیاز به خبرگانی است که قادر به انجام وظیفه یا نقشی خاص در سازمان هستند. در این موارد، به کسانی نیاز هست که نوع معینی از تخصص لازم برای ایفای نقش در شرایط ویژه را دارا هستند. این حالت زمانی روی می‌دهد که وارد کردن فرد خبره در فعالیتی معین یا همکاری پیوسته بین او و جستجوگر مد نظر باشد. نمونه‌ای از موارد عبارت‌اند از:

الف) جستجوی یک مشاور، کارمند یا پیمانکار

ب) جستجوی یک همکار، عضو تیم، عضو کمیته، یا داور مقالات نشریه و کنفرانس

پ) جستجوی یک گوینده، سخنران، پژوهشگر، و مصاحبه‌شونده خبره برای رسانه.

جستجوی خبرگان به‌عنوان منبع اطلاعاتی را «نیاز اطلاعاتی»، و جستجوی خبرگان با هدف

سپردن نقش یا وظیفه‌ای به آن‌ها را «نیاز به تخصص» قلمداد کرده‌اند (Yimam-Seid & Kobsa, 2003).

۴. خبره‌یابی و تشکیل پروفایل خبرگان

یک نظام موفق بازیابی تخصص به سازمان کمک می‌کند تا دو وظیفه مهم را به جا آورد: (۱) خبره‌یابی، (۲) تشکیل پروفایل خبرگان. در ادامه به تشریح این دو عمل و فرمالیزه کردن آن‌ها پرداخته می‌شود.

خبره‌یابی: خبره‌یابی یعنی شناسایی و ارائه سیاهه‌ای رتبه‌بندی شده از افرادی که در موضوعی معین صاحب دانش هستند (در موضوع x چه کسی خبره است؟). به بیان ریاضی، امر خبره‌یابی به دنبال تعیین $P(e|q)$ یعنی احتمال حضور کاندیدای e به شرط حضور پرس وجوی q می‌باشد. هر اندازه میزان احتمال مذکور برای کاندیدایی بیشتر باشد، وی خواهد توانست در رتبه‌بندی ارائه شده جایگاه خوبی را به خود اختصاص داده و محتمل‌ترین خبره برای آن موضوع به حساب آید. چالشی که وجود دارد، نحوه تعیین این احتمال به شکلی صحیح است. از آنجا که پرس وجو معمولاً حاوی تعداد کمی اصطلاح برای توصیف خبرگی است، می‌توان از طریق قضیه «بیز»^۱ تخمین بسیار صحیحی را به دست آورد:

$$P(e|q) = \frac{P(q|e) \cdot P(e)}{P(q)}$$

که در آن $P(q)$ احتمال پرس وجوی q ، و $P(e)$ احتمال کاندیدای e است. از آنجا که $P(q)$ برای هر پرس وجویی ثابت است، می‌توان آن را نادیده گرفت. بدین ترتیب، احتمال خبرگی کاندیدای e در رابطه با q ، با $P(q|e) \cdot P(e)$ معادل است. $P(e)$ نشانگر دانش پیشین درباره خبرگی کاندیدای e است که معمولاً واحد در نظر گرفته می‌شود. در نتیجه، $P(e|q)$ با $P(q|e)$ معادل می‌شود (Balog, Azzopardi & de Rijke 2006; Balog 2008) که تخمین آن در ادامه و در بخش مدل‌ها ارائه می‌شود.

تشکیل پروفایل خبرگان: شناسایی و ارائه سیاهه موضوعاتی می‌باشد که یک فرد در ارتباط با آن‌ها صاحب دانش است (فرد γ در مورد چه موضوعاتی می‌داند؟). در واقع، همان پروفایل موضوعی کاندیداست. هدف از تشکیل پروفایل خبرگان، کاستن از تلاش‌های انسانی در امر نگهداری و به‌روزرسانی پروفایل‌های موضوعی از طریق تمرکز بر روش‌های خودکار است.

1. Bayes theorem

پروفایل موضوعی یک کاندیدا به صورت برداری در نظر گرفته می‌شود که هر مؤلفه بردار یعنی $S(e, k_i)$ با تخصص کاندیدای e در موضوع k_i مطابقت دارد. k_i نشانگر یک حوزه دانشی یا مهارتی است که سازمان به استفاده از آن برای تعریف پروفایل موضوعی کاندیدا علاقه‌مند است. n برابر با تعداد موضوعاتی است که محاسبه تخصص کاندیدای e در هر کدام از آن‌ها مد نظر بوده و تعداد آن از پیش تعیین می‌شود.

$$profile(e) = \langle S(e, k_1), S(e, k_2), \dots, S(e, k_n) \rangle$$

در تشکیل پروفایل خبرگان، احتمال حضور حوزه دانشی k_i در پروفایل موضوعی یک کاندیدا مد نظر است. بنابراین، $S(e, k_i)$ به صورت $P(k_i|e)$ تعریف می‌شود که معادل به دست آوردن $P(q|e)$ است، جایی که موضوع k_i به صورت پرس و جوی q بازنمون می‌شود (Balog & de Rijke 2008; Balog 2007). بنابراین، هم خبره‌یابی و هم تشکیل پروفایل خبرگان بر روی تخمین صحیح $P(q|e)$ را تأکید می‌کنند. از نظر ریاضی، تنها فرق خبره‌یابی و تشکیل پروفایل خبرگان، دانش پیشین در باره خبرگی هر فرد ($P(e)$) است که به طور طبیعی می‌توان آن را در امر خبره‌یابی وارد نمود. اما در تشکیل پروفایل خبرگان با توجه به ثابت بودن هر کاندیدا، وارد کردن آن معنادار نیست (Balog 2008).

۵. مدل‌های بازیابی تخصص

همان‌طور که مشاهده شد، نظام بازیابی تخصص، چه در امر خبره‌یابی و چه در تشکیل پروفایل خبرگان، بر شناسایی ارتباط بین پرس و جو (یعنی موضوع تخصص) و کاندیدا (یعنی فرد خبره) تأکید دارد. هرچقدر ارتباط مابین پرس و جو و کاندیدا قوی‌تر باشد، احتمال خبره‌بودن کاندیدا در حوزه موضوعی پرس و جو بیشتر است. برای به دست آوردن ارتباط بین پرس و جو و کاندیدا مدل‌های گوناگونی توسعه یافته‌اند. در مقاله حاضر به همه این مدل‌ها پرداخته نمی‌شود^۱، بلکه مدل‌هایی مورد توجه قرار می‌گیرند که یا نقشی بنیادی داشته و پایه و اساس سایر مدل‌ها قرار گرفته‌اند، و یا به توسعه مدل‌های پایه کمک قابل توجهی نموده‌اند. بر این اساس، ابتدا مدل سند مجازی (Craswell & et al. 2001) معرفی می‌شود که یکی از اولین تلاش‌های فرمالیزه کردن بازیابی تخصص به شمار رفته و در مدل‌های ارائه شده توسط (Balog et al. 2006) تأثیر گذاشته

۱. علاقه‌مندان برای آشنایی کلی با سایر مدل‌های بازیابی تخصص می‌توانند به اثر «بالوگ و دیگران» (۲۰۱۲) مراجعه نمایند که اطلاعات کتابشناختی آن در بخش منابع همین مقاله آمده است.

است. سپس، به مدل‌های کاندیدا-محور^۱ و سند-محور^۲ که توسط (همان) ارائه شده‌اند، پرداخته می‌شود که می‌توان گفت مبنای تقریباً تمامی مدل‌های بعدی قرار گرفته است. در نهایت، مدل رأی‌دهی (MacDonald 2009) معرفی می‌شود که بر مبنای مدل سند-محور (Balog et al. 2006) استوار است و آن را به‌طور قابل توجهی توسعه داده است.

مدل سند مجازی کراسول و همکاران: در این مدل، اصطلاحات تمام اسناد موجود در پروفایل کاندیدا، در یک سند مجازی (θ_e) جمع می‌شود و سپس، با استفاده از مدل‌های سنتی بازیابی اطلاعات، امتیاز آن در ارتباط با پرس و جوی Q مشخص می‌گردد. امتیاز کاندیدای e در ارتباط با پرس و جوی Q برابر با امتیاز سند مجازی وی در ارتباط با پرس و جوی Q می‌باشد:

$$\text{Score}(e, Q) = \text{score}(\theta_e, Q)$$

به عبارتی دیگر، رتبه کاندیدای e در ارتباط با پرس و جوی Q همان رتبه سند θ_e در ارتباط با پرس و جوی Q است (Craswell & et al. 2003).

مدل‌های ارائه‌شده توسط بالوگ و همکاران: در این مدل‌ها برای محاسبه $P(e|q)$ یعنی احتمال حضور کاندیدا به شرط حضور پرس و جو، با استفاده از قضیه «بیز» و فرض واحد بودن $P(e)$ و $P(q)$ ، محاسبه عکس $P(e|q)$ یعنی $P(q|e)$ در دستور کار قرار می‌گیرد. بنابراین، مدل‌های توسعه یافته، هم برای امر خبره‌یابی و هم برای تشکیل پروفایل خبرگان قابل استفاده هستند. هدف، پی‌بردن به این امر است که کاندیدای e تا چه اندازه می‌تواند پرس و جوی q را تولید کند. «بالوگ، آزوپاردی و دو رایک» به‌منظور تعیین $P(q|e)$ از رویکرد مدل زبانی احتمالاتی که یکی از مدل‌های بازیابی اطلاعات است، به دو روش متفاوت استفاده می‌کند.^۳ در مدل اول که به مدل کاندیدا-محور معروف است، برای هر کاندیدا یک بازنمون متنی از اسناد مرتبط با وی درست می‌شود و سپس از طریق این بازنمون، میزان احتمال پیش‌گفته مشخص می‌شود. در مدل دوم که به مدل سند-محور معروف است، ابتدا اسنادی که به بهترین وجه موضوع پرس و جو را توصیف می‌کنند بازیابی شده، سپس افرادی که با اسناد بازیابی شده مرتبط هستند به‌عنوان خبره در موضوع پرس و جو مورد توجه قرار می‌گیرند (Balog, Azzopardi, & de Rijke 2006).

مدل کاندیدا-محور: در این مدل، که به آن روش پروفایل-محور (Fang & Zhai 2007) یا رویکرد مستقل از پرس و جو (Petkova & Croft 2008) نیز گفته می‌شود، برای تخمین $P(q|e)$ از

1. candidate-centric model

2. document-centric model

۳. البته در معادلات ارائه‌شده برای مدل‌های اول و دوم، هر جا که بازیابی اسناد مد نظر باشد، می‌توان از سایر مدل‌های بازیابی که قابلیت وزن‌دهی به نتایج و رتبه‌بندی آن‌ها را داشته باشند، استفاده کرد.

فون مدل‌سازی زبانی استفاده می‌شود (Balog, Azzopardi, & de Rijke 2006). بدین ترتیب که ابتدا با استفاده از اسناد مرتبط با هر کاندیدا، یک بازنمون متنی (θ_e) از دانش وی ایجاد می‌شود که بدان مدل زبانی کاندیدا نیز گفته می‌شود. سپس، به این امر پرداخته می‌شود که بازنمون متنی هر کاندیدا، تا چه حد می‌تواند پرس و جوی q را تولید کند. به عبارتی دیگر، هدف، محاسبه $P(q|\theta_e)$ است؛ امری که مشخص شدن آن، میزان احتمال ربط پرس و جوی به کاندیدا ($P(q|e)$) را مشخص خواهد کرد. در مدل زبانی استاندارد، $P(q|\theta_e)$ برابر است با ضرب احتمال حضور تک‌تک اصطلاحات پرس و جوی در بازنمون متنی:

$$P(q|\theta_e) = \prod_{t \in Q} P(t|\theta_e)^{n(t,q)}$$

که در آن $n(t,q)$ برابر با تعداد دفعات حضور اصطلاح t در پرس و جوی Q ، و $P(t|\theta_e)$ احتمال حضور اصطلاح t (از اجزای پرس و جوی Q) در بازنمون متنی کاندیدا است. در این مورد، اگر تنها یکی از اصطلاحات پرس و جوی در بازنمون متنی کاندیدا حضور نداشته باشد، حاصل معادله فوق برابر با صفر خواهد شد. این مسئله به «تنگی داده‌ها»^۱ مشهور است. به منظور جلوگیری از بروز احتمال صفر لازم است به جای $P(t|\theta_e)$ مقدار هموارشده آن جای گیرد. با استفاده از روش هموارسازی «جلینیک-مرسر»^۲، مقدار هموارشده آن با $(1-\lambda)P(t|\theta_e) + \lambda.P(t)$ معادل خواهد بود که در آن λ پارامتر هموارسازی بوده و باید مقدار آن بین صفر و یک طوری تنظیم شود که مدل بهترین عملکرد را به جا بگذارد. $P(t)$ برابر با احتمال حضور اصطلاح t در کل مجموعه اسناد است. در نتیجه، معادله ۱ به شکل زیر تغییر خواهد کرد:

$$P(q|\theta_e) = \prod_{t \in Q} ((1-\lambda)P(t|\theta_e) + \lambda.P(t))^{n(t,q)}$$

تنها مسئله باقیمانده در معادله فوق، نحوه محاسبه $P(t|\theta_e)$ می‌باشد. برای محاسبه این مؤلفه، هر یک از اسناد موجود در بازنمون متنی کاندیدا در ایجاد ارتباط بین کاندیدا و اصطلاح نقش ایفا می‌کنند:

$$P(t|\theta_e) = \sum_d P(t|d)P(d|e)$$

که در آن $P(t|d)$ احتمال حضور اصطلاح t در سند d است و در ساده‌ترین حالت با استفاده از تخمین احتمال بیشینه (فراوانی نسبی اصطلاح t در سند d) به دست می‌آید. $P(d|e)$ احتمال ارتباط بین سند d و کاندیدی e است. در صورت عدم حضور نام کاندیدا در سند، مقدار آن صفر، در غیر این صورت مقدار آن غیر صفر بوده و بسته به راهبردهای اتخاذشده می‌تواند مقادیر

مختلفی را به خود اختصاص دهد (همان). در برخی آثار نظیر (MacDonald & Ounis 2006) قوت ارتباط اسناد با کاندیداها کمی سازی نمی‌شود. در نتیجه، در صورت حضور نام کاندیدا در سند، $P(d|e)$ برابر با یک، و در غیر این صورت برابر با صفر در نظر گرفته می‌شود. اما در برخی آثار دیگر مانند (Fang & Zhai 2007; Petkova & Croft 2008)، قوت ارتباط اسناد با کاندیداها متناسب با فراوانی وقوع نام کاندیدا در اسناد است. رویکرد اول را مجموعه-محور^۱ و رویکرد دوم را فراوانی-محور^۲ می‌گویند (Balog & et al. 2012). معادله نهایی مدل کاندیدا-محور (Balog, Azzopardi, & de Rijke 2006) به صورت زیر است:

$$P(q|\theta_e) = \prod_{t \in Q} ((1 - \lambda) \sum_d P(t|d)P(d|e) + \lambda P(t))^{n(t,q)}$$

مدل سند-محور: در این مدل که بدان مدل وابسته به پرس وجو (Petkova & Croft 2008) نیز می‌گویند، به جای ایجاد بازنمون متنی هر کاندیدا، فرایند یافتن خبرگان اندکی متفاوت مورد توجه قرار می‌گیرد، به طوری که در آن کاندیدا به طور مستقیم مدل سازی نمی‌شود. در عوض، اسناد، مدل سازی شده و مورد پرس وجو قرار می‌گیرند. سپس، کاندیداهای مرتبط با اسناد بازیابی شده به عنوان خبرگان احتمالی حوزه موضوعی پرس وجو در نظر گرفته می‌شوند. امتیاز هر کاندیدا در ارتباط با پرس وجو، از طریق جمع امتیاز اسناد مرتبط با وی یا جمع تعداد اسناد مرتبط با وی یا جمع معکوس رتبه اسناد مرتبط با وی به دست می‌آید^۳. نمایش ریاضی مدل سند-محور به صورت ذیل است:

$$P(q|e) = \sum_{d \in D} P(q|d) \cdot P(d|e)$$

که در آن $P(q|d)$ احتمال ارتباط بین سند d با پرس وجوی q ، و $P(d|e)$ احتمال ارتباط بین سند d و کاندیدای e است که راجع به آن در مدل کاندیدا-محور توضیح داده شد. برای محاسبه $P(q|d)$ می‌توان از مدل‌های مختلف بازیابی که قابلیت وزن دهی به نتایج و رتبه بندی آن‌ها را داشته باشند، استفاده کرد. از آنجا که ابداع کنندگان این مدل‌ها در (Balog, Azzopardi, & de Rijke 2006) روش مدل سازی زبانی احتمالاتی را برای بازیابی اسناد مبنا قرار داده‌اند، در اینجا نیز روش مذکور برای بسط رابطه^۲ به کار گرفته می‌شود. در مدل زبانی استاندارد، $P(q|d)$ برابر است با ضرب احتمال حضور تک تک اصطلاحات پرس وجو در سند d :

$$P(q|d) = \prod_{t \in Q} P(t|d)^{n(t,q)}$$

1. set-based

2. frequency-based

۳. در مدل رأی دهی به چگونگی تجمع امتیاز اسناد مرتبط با هر کاندیدا بیشتر پرداخته می‌شود.

که در آن $n(t, q)$ برابر با تعداد دفعات حضور اصطلاح t در پرس و جوی Q ، و $P(t|d)$ احتمال حضور اصطلاح t (از اجزای پرس و جوی Q) در سند d است. با توجه به آنچه پیش تر راجع به تُنکی داده‌ها گفته شد و به منظور جلوگیری از بروز احتمالات صفر، لازم است مقدار هموارشده $P(t|d)$ در معادله فوق قرار گیرد. با قراردادن مقدار هموارشده $P(t|d)$ ، رابطه فوق به شکل زیر در خواهد آمد:

$$P(q|d) = \prod_{t \in Q} ((1 - \lambda)P(t|d) + \lambda \cdot P(t))^{n(t,q)}$$

بدین ترتیب، معادله نهایی مدل سند-محور به صورت زیر خواهد بود:

$$\sum_{d \in D} \prod_{t \in Q} ((1 - \lambda)P(t|d) + \lambda \cdot P(t))^{n(t,q)} \cdot P(d|e)$$

تقریباً همه نظام‌هایی که در ویرایش‌های ۲۰۰۵ تا ۲۰۰۷ «ترک»^۱ شرکت کرده‌اند، یک یا هر دوی این رویکردها را پیاده‌سازی کرده‌اند. نتایج ارزیابی‌ها نشان داده است که مدل سند-محور در مقایسه با مدل کاندیدا-محور، چه در امر خبره‌یابی و چه در تشکیل پروفایل خبرگان، عملکرد بهتری داشته است (همان).

مدل رأی‌دهی: همان‌طور که پیش تر نیز بحث شد، در مدل سند-محور خبره‌یابی، نظام ابتدا با به کارگیری یک مدل بازیابی، اسنادی را که گمان می‌کند با پرس و جو مرتبط هستند، بازیابی می‌کند ($P(q|d)$). اگر سند d که در میان اسناد بازیابی شده قرار دارد، با کاندیدای e در ارتباط باشد، $P(d|e)$ غیر صفر، و اگر در ارتباط نباشد، عدد صفر به آن تعلق می‌یابد. در مدل رأی‌دهی نیز، هر یک از اسناد بازیابی شده که نام خبرگان کاندیدا در آن ذکر شده باشد، به‌عنوان رأی برای آن کاندیدا محسوب می‌شود. به عبارتی دیگر، در صورت وقوع نام کاندیدای e در سند d ، $P(d|e)$ برابر با یک در نظر گرفته می‌شود، در غیر این صورت برابر با صفر شده و آن سند رأی برای کاندیدا محسوب نمی‌شود. بنابراین، مدل رأی‌دهی کاملاً بر مبنای مدل سند-محور و منطبق با آن است. نوآوری مدل رأی‌دهی، در ارائه شیوه‌های گوناگون تلفیق آرائی است که اسناد به کاندیداها اعطا کرده‌اند: ممکن است امتیاز سند بر طبق مدل بازیابی به کاندیدا داده شود، یا رتبه سند به کاندیدا داده شود، یا به‌طور ساده هر سند یک امتیاز محسوب شود. (MacDonald & Ounis (2006) و MacDonald (2009) با استفاده از فنون هم‌جوشی داده‌ها^۲ روش‌های گوناگونی را برای تلفیق آراء ارائه کرده‌اند که در ادامه به آن‌ها پرداخته می‌شود. اما قبل از پرداختن به آن‌ها، توضیح مختصری

در مورد فنون هم‌جوشی ارائه می‌شود.

فنون هم‌جوشی داده‌ها به‌عنوان ابزاری برای ادغام رتبه‌بندی‌های صورت گرفته توسط نظام‌های بازیابی اطلاعات مختلف به کار می‌رود (Fox & et al. 1993). برای مثال، ابرموتورهای جستجو برای رتبه‌بندی نتایج از فنون هم‌جوشی داده‌ها بهره می‌برند. در معماری ابرموتور جستجو چندین موتور جستجو قرار دارند که هر کدام از آن‌ها ممکن است در پاسخ به پرس‌وجوی عرضه شده، رتبه‌بندی متفاوتی از صفحات بازیابی شده ارائه دهد. ابرموتور جستجو برای تلفیق رتبه‌بندی صورت گرفته توسط موتورهای جستجو و ارائه یک رتبه‌بندی نهایی از صفحات بازیابی شده، فنون هم‌جوشی داده‌ها را به کار می‌گیرد. فنون هم‌جوشی داده‌ها بسیار گسترده‌اند: به‌عنوان مثال، برای به‌دست آوردن رتبه نهایی هر صفحه، می‌توان امتیازی را که توسط نظام‌های مختلف به آن صفحه تخصیص داده شده، با هم جمع کرده و از نمره نهایی حاصل شده برای رتبه‌بندی استفاده کرد، یا می‌توان رتبه‌ای را که هر صفحه در نظام‌های مختلف به‌دست آورده، معکوس نموده و سپس با هم جمع زده و از نمره نهایی حاصل شده برای رتبه‌بندی استفاده کرد و نظایر آن. در نظام‌های بازیابی اطلاعات فنون هم‌جوشی داده‌ها، رتبه‌بندی‌های ایجاد شده توسط نظام‌های مختلف بازیابی اطلاعات را در یک رتبه‌بندی واحد ادغام می‌نماید، اما در نظام خبره‌یابی، فنون هم‌جوشی داده‌ها با تلفیق آرائی که کاندیدا از اسناد به‌دست آورده، رتبه واحدی برای وی تولید می‌کند. دوازده فن هم‌جوشی که MacDonald & Ounis (2006) و MacDonald (2009) برای ادغام آراء کاندیداها ارائه می‌دهند، به شرح زیر است:

۱. روش ترکیب آرای تأییدی^۱: در این روش تعداد اسناد بازیابی شده که حاوی نام کاندیدا هستند، برای رتبه‌بندی کاندیداها ملاک قرار می‌گیرد. هرچه قدر فراوانی اسنادی که حاوی نام کاندیدا هستند بیشتر باشد، رتبه وی بهتر خواهد بود.
۲. روش مبتنی بر رتبه معکوس^۲: در این روش، رتبه آن دسته از اسناد بازیابی شده که حاوی نام کاندیدا هستند، معکوس شده و سپس با هم جمع می‌شوند تا رتبه نهایی کاندیدا مشخص شود.
۳. روش ترکیبی بُردا^۳: در این روش، رتبه هر یک از اسناد بازیابی شده که حاوی نام کاندیداست، از تعداد کل اسناد بازیابی شده کم می‌شود. سپس اعداد به‌دست آمده با هم جمع می‌شوند تا امتیاز نهایی کاندیدا مشخص گردد. هرچه امتیاز کاندیدا بالاتر باشد، رتبه وی بهتر خواهد بود. برای مثال، اگر کلاً ۱۰۰ سند بازیابی شود و رتبه یکی از اسنادی که نام کاندیدا در آن واقع شده، ۱۰۰ باشد، نمره‌ای که فرد از آن سند به‌دست می‌آورد ($100 - 100 = 0$) صفر خواهد بود.

1. ApprovalVotes

2. Reciprocal Rank

3. BordaFuse

۴. روش مبتنی بر میانه امتیازات^۱: در این روش، آن دسته از اسناد بازیابی شده که حاوی نام کاندیدا هستند، بر طبق امتیازی که مدل بازیابی اطلاعات به آن‌ها تخصیص داده است، مرتب می‌شوند و امتیاز سند میانی به عنوان امتیاز کاندیدا در نظر گرفته شده و برای رتبه‌بندی به کار می‌رود.

روش مبتنی بر امتیاز حداقل^۲: در این روش، در میان اسناد بازیابی شده حاوی نام کاندیدا، امتیاز سندی که کمترین امتیاز را از مدل بازیابی اطلاعات گرفته، به عنوان امتیاز کاندیدا در نظر گرفته شده و برای رتبه‌بندی به کار می‌رود. هرچقدر این امتیاز بالاتر باشد، رتبه کاندیدا بهتر خواهد بود. روش مبتنی بر امتیاز حداکثر^۳: در این روش، در میان اسناد بازیابی شده حاوی نام کاندیدا، امتیاز سندی که بیشترین امتیاز را از مدل بازیابی اطلاعات گرفته، به عنوان امتیاز کاندیدا در نظر گرفته شده و برای رتبه‌بندی به کار می‌رود. هرچقدر این امتیاز بالاتر باشد، رتبه کاندیدا بهتر خواهد بود. جمع امتیازات^۴: در این روش، امتیاز هر یک از اسناد بازیابی شده که حاوی نام کاندیدا هستند، با هم جمع شده و امتیاز نهایی فرد را تشکیل می‌دهند.

میانگین امتیازات^۵: در این روش، امتیاز هر یک از اسناد بازیابی شده که حاوی نام کاندیدا هستند، با هم جمع شده، و حاصل، بر تعداد اسناد حاوی نام فرد تقسیم می‌شود تا امتیاز نهایی کاندیدا به دست آید.

ضرب مجموع امتیازات در تعداد حامل‌ها^۶: در این روش، امتیاز هر یک از اسناد بازیابی شده که حاوی نام کاندیدا هستند، با هم جمع شده، سپس حاصل در تعداد اسناد حاوی نام فرد ضرب می‌شود تا امتیاز نهایی کاندیدا به دست آید.

جمع مقادیر تابع نمایی امتیازات^۷: در این روش، امتیاز هر یک از اسناد بازیابی شده که حاوی نام کاندیدا هستند، در تابع نمایی (exp) قرار می‌گیرد، سپس اعداد به دست آمده با هم جمع می‌شوند تا امتیاز نهایی فرد را تشکیل دهند.

میانگین مقادیر تابع نمایی امتیازات^۸: در این روش، امتیاز هر یک از اسناد بازیابی شده که حاوی نام کاندیداست، در تابع نمایی (exp) قرار می‌گیرد، سپس اعداد به دست آمده با هم جمع می‌شوند. حاصل، بر تعداد اسناد بازیابی شده که حاوی نام کاندیدا هستند، تقسیم می‌شود تا امتیاز نهایی فرد به دست آید.

ضرب مجموع مقادیر تابع نمایی امتیازات در تعداد حامل‌ها^۹: در این روش، امتیاز هر یک از اسناد

1. CombMED
4. CombSUM
7. expCombSUM

2. CombMIN
5. CombANZ
8. expCombANZ

3. CombMAX
6. CombMNZ
9. expComMNZ

بازیابی شده که حاوی نام کاندیداست، در تابع نمایی (exp) قرار می‌گیرد، سپس اعداد به دست آمده با هم جمع می‌شوند. حاصل، در تعداد اسناد بازیابی شده که حاوی نام کاندیدا هستند، ضرب می‌شود تا امتیاز نهایی فرد به دست آید (MacDonald 2009; MacDonald & Ounis 2006).

به‌طور کلی، تلفیق آراء کاندیداها را می‌توان در سه گروه جای داد: زمانی که تعداد اسناد رأی‌دهنده رتبه کاندیدا را تعیین می‌کند (مثل روش ترکیب آراء تأییدی)؛ زمانی که امتیاز اسناد رأی‌دهنده رتبه کاندیدا را تعیین می‌کنند (مثل روش‌های مبتنی بر میانه، مبتنی بر حداقل امتیاز، مبتنی بر حداکثر امتیاز، جمع امتیازات، میانگین امتیازات، ضرب مجموع امتیازات در تعداد حامل‌ها، جمع مقادیر تابع نمایی امتیازات، میانگین مقادیر تابع نمایی امتیازات، ضرب مجموع مقادیر تابع نمایی امتیازات در تعداد حامل‌ها)؛ و زمانی که رتبه اسناد رأی‌دهنده رتبه کاندیدا را تعیین می‌کند (مثل روش‌های معکوس رتبه و ترکیب بردا).

«مک‌دونالد و اونیس» و «مک‌دونالد» به ارزیابی و مقایسه ۱۲ فن هم‌جوشی پرداختند. نتایج نشان داد که دو روش جمع امتیازات و ضرب مجموع امتیازات در تعداد حامل‌ها، در بین سایر روش‌ها بهترین عملکرد را دارند (MacDonald & Ounis 2006; MacDonald 2009). همچنین، ویرایش‌های آن‌ها در قالب تابع نمایی یعنی روش‌های جمع مقادیر تابع نمایی امتیازات و ضرب مجموع مقادیر تابع نمایی امتیازات در تعداد حامل‌ها عملکرد بازیابی را بیشتر بهبود می‌بخشند.

۶. مجموعه آزمون‌های بازیابی تخصص

وجود مجموعه آزمون‌ها برای ارزیابی عملکرد الگوریتم‌ها و مدل‌های توسعه یافته ضروری است. مجموعه آزمون که در ارتباط با بازیابی اطلاعات معنا پیدا می‌کند، مشتمل بر مجموعه عظیمی از اسناد است و دارای سه مؤلفه اساسی است: مجموعه‌ای از مدارک، مجموعه‌ای از عناصر موضوعی (پرس و جواها)، مجموعه‌ای از قضاوت‌های ربط (Sanderson 2010, 250). این مجموعه در واقع در حکم یک آزمایشگاه است که تجهیزات و لوازم مورد نیاز برای آزمون و ارزیابی الگوریتم‌های ایجاد شده را داراست. بدین ترتیب، می‌توان الگوریتم‌ها یا مدل‌های مختلف را مورد آزمون قرار داده و مشخص کرد که کدام یک از آن‌ها عملکرد بهتری دارد.

از معروف‌ترین مجموعه‌های آزمون بازیابی اطلاعات، مجموعه‌های تهیه شده توسط «ترک»^۱ است که با هدف ایجاد زمینه و امکانات لازم جهت انجام آزمایش‌هایی برای بهبود نظام‌های بازیابی اطلاعات متنی و ایجاد فضای رقابتی برای بهبود مستمر این نظام‌ها به وجود آمده است. از

1. TREC

دیگر مجموعه‌های معروف جهانی می‌توان به «کلیف»^۱ و «لتور»^۲ اشاره کرد (درودی و همکاران، ۱۳۸۸). در ایران نیز، برای پیشبرد تحقیقات مربوط به بازیابی اطلاعات متنی فارسی و نیز سایر حوزه‌های مرتبط با بازیابی اطلاعات مثل روش‌های نمایه‌سازی، ریشه‌یابی و مباحث زبان‌شناسی، مجموعه‌های آزمون‌ی توسط گروه تحقیقاتی پایگاه داده دانشگاه تهران^۳ ایجاد شده است که از جمله مهم‌ترین آن‌ها می‌توان به مجموعه «بیجن‌خان»^۴، مجموعه روزنامه همشهری و مجموعه دات‌آی‌آر^۵ اشاره نمود (صفحه خانگی گروه تحقیقاتی پایگاه داده دانشگاه تهران). اخیراً در علم اطلاعات و دانش‌شناسی نیز مجموعه آزمون کوچکی از مقالات دو نشریه «علوم و فناوری اطلاعات»، و «نما» توسط «فهمی‌فر و فهم‌نیا» (۱۳۹۱) ایجاد شده است.

در ارتباط با امر بازیابی تخصص نیز چندین مجموعه آزمون ایجاد شده است که در اینجا نیز «ترک» پیشرو است. «ترک» در سال ۲۰۰۵ مجموعه آزمون «دابلوتتری‌سی» (کنسرسیوم جهانی وب (W3C))^۶ را، که حاصل یک خزش در صفحات وب سایت کنسرسیوم جهانی وب در سال ۲۰۰۴ بود، برای خبره‌یابی ارائه کرد. این مجموعه با ۳۳۱۰۳۷ سند، و حجمی بالغ بر ۵/۷ گیگابایت و با سیاهه ۱۰۹۲ نفری از خبرگان کاندیدا، نه تنها صفحات وب، بلکه سیاهه‌های پست الکترونیکی، اسناد فنی، و سایر انواع داده‌ها را که عملیات روزمره کنسرسیوم وب جهانی را منعکس می‌کند، شامل می‌شود. بعد از اتمام خزش، ۱۰۹۲ خبره کاندیدا در گروه‌های مختلف کاری^۷ که موضوعات متنوعی را پوشش می‌دادند، جای گرفتند. پنجاه پرس‌وجو که موضوع آن‌ها، همان موضوع گروه‌های کاری بود، ایجاد شد. در این پیکره، با مطرح کردن پرس‌وجویی که موضوع آن یکی از گروه‌های کاری است، انتظار می‌رود افرادی که به صورت پیش‌فرض در آن گروه موضوعی، خبره قلمداد شده‌اند، بازیابی شوند. در صورت عدم تطابق کامل افراد بازیابی شده با افراد موجود در سیاهه، شاخص عملکرد مدل اجرا شده از ۱۰۰ پایین می‌آید. به عبارتی دیگر، مبنای ارزیابی عملکرد مدل خبره‌یابی، سیاهه افرادی است که در گروه‌های موضوعی به عنوان خبره قلمداد شده‌اند (Craswell, de vries, & Soboroff 2005). اقدام «ترک» در تهیه مجموعه آزمون «دابلوتتری‌سی» منجر به پیشرفت اساسی در حوزه بازیابی تخصص گردید و از نظر مدل‌سازی، الگوریتم‌ها، و روش‌های ارزیابی، پیشرفت‌های سریعی در حوزه بازیابی تخصص حاصل شد.

«ترک» در سال ۲۰۰۷ مجموعه آزمون جدید دیگری را با نام «سی‌اس‌آی‌رو»^۸ ارائه کرد که مشتمل بر ۳۷۰۳۱۵ سند، دارای حجمی برابر با ۴/۲ گیگابایت، و سیاهه‌ای ۳۶۷۸ نفری از خبرگان کاندیدا بود. این پیکره حاصل خزش در وب‌سایت‌های CSIRO.AU* است و در ساخت

1. CLEF: Conference and Labs of the Evaluation Forum

2. LETOR: Learning To Rank for information retrieval

4. Bijankhan

7. working groups

3. Database Research Group

6. W3C: World Wide Web Consortium

5. dotIR

8. CSIRO: Commonwealth Scientific and Industrial Research Organization

پرس‌وجوهای آن از افرادی استفاده شده که از آن‌ها با نام مروّجان علم «سی‌اس‌آی‌رو»^۱ یاد می‌شود. نقش مروّجان علم، ارتقای تصور عمومی از «سی‌اس‌آی‌رو» و بالابردن قابلیت‌های «سی‌اس‌آی‌رو» از طریق مدیریت اطلاعات و تعامل با گروه‌های صنعتی، آژانس‌های دولتی، گروه‌های حرفه‌ای، رسانه و عموم مردم است. پنجاه پرس‌وجوی ساخته‌شده توسط مروّجان علم به دو منظور می‌توانند مورد استفاده قرار گیرند: بازیابی اسناد و خبره‌یابی. در ارتباط با بازیابی اسناد، از مروّجان علم خواسته شده تا منابع و اسناد کلیدی مرتبط با پرس‌وجویی را که مطرح کرده‌اند، سیاهه کنند. این سیاهه در پژوهش‌های بازیابی اسناد به‌عنوان مبنای ارزیابی قرار می‌گیرد. در ارتباط با خبره‌یابی، از مروّجان علم خواسته می‌شود نام افرادی را که در حوزه موضوعی پرس‌وجو به‌عنوان خبره کلیدی به شمار می‌روند، سیاهه کنند. مروّجان علم برای هر موضوع، حداقل ۱ خبره و حداکثر ۱۱ خبره ذکر کرده‌اند. این سیاهه در پژوهش‌های خبره‌یابی به‌عنوان مبنای ارزیابی قرار می‌گیرد (Bailey & et al. 2007).

مجموعه‌های دیگری نیز که بازنمون نوع متفاوتی از سازمان‌ها هستند، ایجاد شده‌اند، مانند مجموعه خبرگان «یووی‌تی»^۲ یا همان مجموعه خبرگان دانشگاه «تیلبورگ» هلند. این مجموعه که از خزش وب‌سایت دانشگاه «تیلبورگ» استخراج شده، به دو زبان هلندی و انگلیسی بوده و تعداد خبرگان کاندیدا در آن ۱۱۶۸ نفر می‌باشد. ایجادکنندگان این مجموعه آزمون معتقدند که مجموعه آن‌ها با پیکره «دبلیوتری‌سی» متفاوت است؛ چراکه این مجموعه تخصص‌های گسترده‌تری را تحت پوشش قرار می‌دهد؛ دربردارنده چندین ساختار مثل ساختار موضوعی و ساختار سازمانی بوده و مشتمل بر اسناد پژوهشی، دوره‌های آموزشی، انتشارات و صفحات خانگی دانشگاهی افراد است؛ شرایطی که از دیدگاه سازندگان این مجموعه آزمون ممکن است بر امر بازیابی تخصص اثر بگذارد. در این مجموعه، خبرگان کاندیدا خودشان بیان می‌کنند که در چه حوزه‌ای صاحب تخصص هستند؛ بدین ترتیب که از لیست حوزه‌های موضوعی خبرگی که شامل ۱۴۹۱ موضوع است، حوزه‌های تخصص خود را می‌توانند انتخاب کنند. اگر حوزه تخصصی آن‌ها در بین ۱۴۹۱ موضوع قرار نداشت، می‌توانند افزوده شدن آن را به لیست تخصص‌ها درخواست کنند. بدین ترتیب، مشخص می‌شود که در دانشگاه، چه افرادی در یک حوزه موضوعی خاص دارای تخصص هستند (Balog & et al. 2007). حال، باید پرس‌وجوهای موضوعی به مجموعه آزمون عرضه شوند و مدل‌های بازیابی تخصص پیاده‌سازی شوند تا مشخص گردد کدام‌یک از مدل‌ها با سیاهه افراد خبره آن حوزه همپوشانی بیشتری دارند. از آنجا که در این مجموعه آزمون سیاهه حوزه‌های تخصصی هر فرد موجود است، عملکرد مدل بازیابی تخصص را علاوه بر

خبره‌یابی، می‌توان برای تشکیل پروفایل خبرگان نیز سنجید. بدین ترتیب که با وارد کردن نام خبرگان، سیاهه حوزه‌های تخصصی آن‌ها را بازیابی کرد. همپوشانی بالای تخصص‌های بازیابی‌شده برای هر کاندیدا با تخصص‌های موجود در پروفایل وی، نشانگر عملکرد مطلوب مدل بازیابی تخصص است.

۷. قضاوت ربط در مورد خبرگان بازیابی‌شده

انجام قضاوت ربط در مورد خبرگان پیشنهادشده توسط نظام جستجوی خبرگان کار آسانی نیست و طی چندین سال برگزاری کنفرانس‌های سالانه «ترک»، روش‌های ارزیابی متفاوتی مورد بررسی قرار گرفته‌اند. قضاوت ربط در مورد خبرگان پیشنهادشده مشکل‌تر از قضاوت ربط در مورد اسناد بازیابی‌شده است، چراکه کاربران با خواندن یا مرور سند بازیابی‌شده به راحتی می‌توانند در مورد ربط آن قضاوت کنند، در حالی که، قضاوت ربط خبرگان به این سادگی نیست. رضایت کاربر در نظام جستجوی خبرگان، احتمالاً به تعامل موفق او با خبرگان پیشنهادشده (مثلاً دریافت یک اندرز مفید) وابسته است. (MacDonal (2009 سه استراتژی مختلف برای قضاوت ربط در مورد خبرگان شناسایی کرده است که ذیلاً به آن‌ها اشاره می‌شود.

واقعیت قطعی‌ی‌ا‌پ‌م‌م‌ج‌م‌و‌د: در این روش، ساخت پرس‌وجوها و قضاوت‌های ربط بر مبنای واقعیت‌های مطلق استوار می‌شوند. بدین معنا که با توجه به دانش قبلی در مورد فرد، تخصص وی در یک یا چند موضوع، مسلم فرض شده و انتظار می‌رود به‌هنگام جستجوی آن موضوع (ها)، مدل بازیابی تخصص بتواند فرد مذکور را بازیابی کند. برای مثال، کنسرسیوم جهانی وب (W3C) شامل گروه‌های کاری بوده و هر کدام اعضایی دارند. در مجموعه آزمون خبره‌یابی «ترک» ۲۰۰۵ اعضای هر گروه کاری، در حیطه موضوعی آن گروه، خبره محسوب شده و اسامی گروه‌های کاری به‌عنوان پرس‌وجو انتخاب گشته است. هدف، این بود که مشخص شود مدل‌های بازیابی تخصص با دریافت نام گروه کاری به‌عنوان پرس‌وجو، تا چه اندازه می‌توانند اعضای آن گروه را پیش‌بینی کنند. اشکال این روش آن است که در مورد پرس‌وجوهای غیرمنطبق با نام گروه‌های کاری دچار مشکل می‌شود. همچنین، ممکن است افراد در موضوعاتی خبره باشند، لیکن عضو گروه کاری مربوطه نباشند.

پرسش از کاندیداهای خبرگی: این کار به دو روش قابل انجام است: در روش اول، از همه خبرگان کاندیدا خواسته می‌شود تا خبره‌بودن یا خبره‌نبودن خود را در حوزه موضوعی تک‌تک پرس‌وجوها تعیین کنند؛ امری که بسیار زمان‌بر بوده و مستلزم پرسش‌های فراوان از کاندیداهاست. در روش دوم، برای کاستن از حجم کاری روش اول، فرد کاندیدا تنها زمانی مورد پرسش قرار

می‌گیرد که توسط مدل بازیابی تخصص برای پرس‌وجو پیشنهاد شده باشد. به هر حال، پرسش از خود کاندیدا در مورد تخصص‌اش برای سازمان‌های بزرگ که صدها یا هزاران نفر کارمند هستند، مناسب نیست. ممکن است همه کاندیداها در دسترس نبوده یا به پرسشنامه‌ها پاسخ ندهند. در چنین مواردی، شاید روش‌شناسی اجازه دهد تا از همکاران فرد در مورد تخصص وی پرسش شود که این امر ممکن است بر روایی نتایج اثرگذار باشد. روشی که در «ترک» ۲۰۰۷ و در یک محیط سازمانی با اندازه متوسط به کار برده شد، مشتق از روش پرسشنامه‌ای است. در این روش، عده‌ای از افراد که دانش خوبی در مورد حوزه‌های تخصص کاندیداها دارند، در مورد ربط خبرگان به پرس‌وجوها قضاوت می‌کنند. مزیت این روش آن است که تعداد کمی از افراد در فرایند قضاوت ربط درگیر می‌شوند و شناسایی کاندیداها هر پرس‌وجو به احتمال قوی قابل اتکاء خواهد بود. اما ممکن است ارزیابان شناخت جامعی از علائق هر کاندیدا نداشته باشند، و بدین ترتیب، برخی از کاندیداها مرتبط با پرس‌وجو شناسایی نشوند؛ امری که بر جامعیت نتایج بازیابی شده تأثیر می‌گذارد.

روش شواهد پشتیبان: در این روش، کاندیداها رتبه‌بالی پیشنهاد شده برای هر پرس‌وجو در یک انبوه قرار می‌گیرند. سپس، برای هر کاندیدی قرار گرفته در انبوه، ده سندی که از خبرگی وی پشتیبانی می‌کنند در انبوه دیگری قرار می‌گیرند. قضاوت ربط شامل دو مرحله است. ارزیابان قبل از انجام قضاوت ربط کاندیدا به پرس‌وجو، اسناد پشتیبان تخصص آن‌ها را خوانده و در مورد ربط اسناد قضاوت ربط انجام می‌دهند. در واقع، ارزیابان با کمک اسناد پشتیبان، در مورد مربوط بودن فرد به پرس‌وجو قضاوت می‌کنند. در ارزیابی‌های نهایی، تنها ارزیابی‌های صورت گرفته در مورد ربط کاندیداها برای سنجش عملکرد نظام بازیابی تخصص مورد استفاده قرار می‌گیرد. این روش قضاوت ربط که در «ترک» ۲۰۰۶ مورد استفاده قرار گرفته، زمانی مناسب است که ارزیابان هیچ‌گونه دانش پیشینی در مورد حوزه‌های تخصص کاندیداها نداشته باشند. محدودیت این روش، وابسته بودن آن به محتوای پیکره اسناد است. ممکن است کاندیدی مربوط به پرس‌وجو، به دلیل نبود اسناد پشتیبان، نامربوط ارزیابی شود. دوسطحی بودن انبوه‌سازی، مشکلاتی را به وجود می‌آورد که در انبوه‌سازی تک‌سطحی اسناد وجود ندارد: ممکن است یک کاندیدی مربوط به دلیل قرار نگرفتن در انبوه کاندیداها اصلاً مورد ارزیابی قرار نگیرد یا کاندیدی مرتبط با پرس‌وجو به دلیل انبوه شدن اسناد پشتیبان نامربوط، نامربوط ارزیابی شود (Macdonald 2009).

هر کدام از فنون قضاوت ربط که در بالا گفته شد، مزایا و معایبی در خصوص راحتی استفاده، قابلیت استفاده مجدد و قابلیت اعتماد دارد. برای اطمینان از عدم اثرگذاری نوع قضاوت ربط انجام شده بر عملکرد مدل بازیابی تخصص، «مک‌دونالد» پیشنهاد می‌کند که مدل بازیابی

تخصص بر روی مجموعه آزمون‌های مختلف که از فنون متفاوت قضاوت ربط بهره برده‌اند، پیاده‌سازی و آزمون شود (MacDonald 2009).

۸. نتیجه‌گیری

پژوهش‌هایی که در دهه ۱۹۶۰ میلادی در علم اطلاعات و دانش‌شناسی در حوزه رفتار اطلاع‌یابی کارکنان دانشی مانند پژوهشگران و مهندسان صورت گرفت، نشان داد که مراجعه به خبرگان انسانی یکی از روش‌های کسب اطلاعات کارکنان دانشی بوده است (Balog et al. 2012). با این وجود، حدود سه دهه طول کشید تا نیاز کاربران به یافتن خبرگان انسانی مورد توجه نظام‌های اطلاعاتی قرار گیرد. در دهه ۱۹۹۰ میلادی توجه سازمان‌ها به مدیریت دانش باعث شد که آن‌ها بخشی را به نظام‌های اطلاعاتی خود اضافه کنند که از بازیابی تخصص پشتیبانی نماید. گفتنی است نظام‌های بازیابی تخصص به‌عنوان جزئی از نظام‌های اطلاعاتی، در نقش ابزار پشتیبان مدیریت دانش عمل می‌کنند و از مسائل مهم مدیریت دانش مانند فرهنگ اشتراک دانش، پنهان کردن دانش، اعتماد و نظایر آن دوری می‌کنند.

با توجه به مشکلات موجود در ایجاد و نگهداری نظام‌های دستی بازیابی تخصص، از همان ابتدا تلاش‌ها بر روی ایجاد نظام‌هایی معطوف گشت که بتوانند خبرگان را به‌طور خودکار پروفایل کرده و سپس در مواقع نیاز بازیابی نمایند. لازمه این امر، عملیاتی نمودن مفهومی چالشی و پیچیده تخصص بود. بدین‌منظور، پژوهشگران با دوری جستن از چالش‌های شناختی مفهوم تخصص بر روی اسناد متنی به‌عنوان شاهد تخصص افراد تمرکز نموده و چنین تعریفی را ارائه کردند: هرچقدر فرد با اسناد یک حوزه موضوعی ارتباط مستحکمی داشته باشد، بیشتر می‌توان بر روی او به‌عنوان خبره آن موضوع تأکید کرد. بدین ترتیب، مسئله بازیابی تخصص به انباشت اسناد متنی و کشف روابط افراد با آن اسناد تقلیل یافت. برای کشف روابط بین اسناد و افراد، مدل‌های متعددی ارائه شده‌اند که مدل‌های ارائه‌شده توسط «بالوگ» سهم بسزایی در پیشبرد حوزه بازیابی تخصص داشته‌اند. مدل سند-محور وی یکی از موفق‌ترین مدل‌هاست که مبنای بسیاری از مدل‌های بعدی قرار گرفته است. از این رو، «بالوگ» و همکارانش به‌عنوان افرادی شناخته می‌شوند که در سازماندهی به مفاهیم آشفته بازیابی تخصص و فرمالیزه کردن آن نقشی اساسی ایفاء کرده‌اند.

مجموعه آزمون‌های مختلفی برای خبره‌یابی ایجاد شده‌اند که می‌توان عملکرد مدل‌های گوناگون بازیابی تخصص را بر روی آن‌ها مورد آزمون قرار داد. از جمله این مجموعه‌ها می‌توان به مجموعه‌های «ترک» مابین سال‌های ۲۰۰۵ تا ۲۰۰۸ و مجموعه آزمون دانشگاه «تیلورگ» اشاره کرد. ویژگی مجموعه آزمون‌های مذکور آن است که هر کدام شرایط سازمانی متفاوتی را

نمایندگی می‌کنند، بنابراین، می‌توان عملکرد مدل‌های بازیابی تخصص را در شرایط سازمانی متفاوت مورد آزمون قرار داد. طبعاً مدلی موفق‌تر است که در شرایط سازمانی مختلف عملکرد مطلوبی داشته باشد.

قضاوت در مورد خبرگان بازیابی شده یکی دیگر از حوزه‌هایی است که بسیار چالشی بوده و به نظر می‌رسد تلاش‌ها تا رسیدن به یک روش مورد توافق نسبی ادامه داشته باشد. هیچ‌کدام از روش‌های به‌کاررفته کامل نبوده و مزایا و معایبی دارند. روش جدید پیشنهادی ما استفاده از نظرات یکی از خبرگان بازیابی شده برای قضاوت ربط در مورد سایر خبرگان بازیابی شده است. لیکن نحوه انتخاب خبره قضاوت‌کننده و همچنین جامعیت شناخت وی از سایر خبرگان مسائلی است که باید حل و فصل شوند.

در نهایت، باید گفت که حوزه بازیابی تخصص، حوزه‌ای است نسبتاً نوپا که در خارج از ایران کارهای فراوانی در زمینه توسعه مفاهیم و مدل‌های آن صورت گرفته است و حتی محصولات کاربردی مانند «آمینر» (<https://aminer.org>) توسعه یافته‌اند. اما در ایران به این حوزه چندان پرداخته نشده است. بازیابی تخصص را در کشور از دو بُعد می‌توان مورد توجه قرار داد: بُعد نظری و بُعد عملی. در بُعد نظری می‌توان با هدف مشارکت جهانی در ادبیات حوزه، پژوهش‌هایی را برای توسعه مفاهیم و مدل‌ها تعریف کرد؛ به این معنا که با ایده‌پردازی‌های نو، مدل‌های جدیدی ارائه شده یا با هدف بهبود عملکرد مدل‌های پیشین، در راستای بسط آن‌ها و به‌طور کلی پیشبرد دانش نظری حوزه گام برداشته شود. در بُعد عملی، می‌توان با توجه ضرورت وجود چنین نظام‌هایی برای سازمان‌های بزرگ و سایر محیط‌های دانش‌مدار، طرح‌های پژوهشی عملیاتی برای پیاده‌سازی نظام بازیابی تخصص در سازمان‌ها تعریف نمود. به عبارتی دیگر، نظام بازیابی تخصص برای سازمان طراحی شده و با نظام اطلاعاتی آن ادغام شود.

فهرست منابع

- پرنور، احسان، و جلال رضایی‌نور. زودآیند. مدل جدید پیشنهادگر فرد خبره با استفاده از ماتریس همبستگی کلمات. *پژوهشنامه پردازش و مدیریت اطلاعات*. <http://Jipm.irandoc.ac.ir> (دسترسی در ۱۳۹۳/۱۱/۱۵).
- حسن‌زاده، محمد. ۱۳۸۴. مدیریت اطلاعات و مدیریت دانش (رویکرد مقایسه‌ای). *نشریه اطلاع‌شناسی* ۹ و ۱۰: ۷-۲۱.
- درودی، احسان، هما برادران هاشمی، ابوالفضل آل احمد، امیرحسین حبیبیان، محمدعلی زارع، فرزاد مهدیخانی، آزاده شاکری، و مسعود رهگذر. ۱۳۸۸. *پیکره محک استاندارد برای تحقیقات بازیابی اطلاعات فارسی*. ece.ut.ac.ir/dbgr/webir/files/Papers/WebIR.pdf (دسترسی در ۱۳۹۳/۵/۱۷).
- فهمی‌فر، سپیده، و فاطمه فهمی‌نیا. ۱۳۹۱. طراحی مجموعه‌های آزمون: راهکار ارزیابی کارآمدی نظام‌های بازیابی اطلاعات. *نشریه تحقیقات کتابداری و اطلاع‌رسانی دانشگاهی* ۴۶ (۶۱): ۱۷-۴۴.

گروه تحقیقاتی پایگاه داده دانشگاه تهران. پیکره‌های فارسی <http://ece.ut.ac.ir/dbrg> (دسترسی در ۱۳۹۳/۵/۱۷).

- Alarfaj, F., U. Kruschwitz, D. Hunter, & C. Fox. 2012. *Finding the right supervisor: expert-finding in a university domain*. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: human language technologies: student research workshop (pp. 1-6). Association for Computational Linguistics.
- Bailey, P., N. Craswell, I. Soboroff, & A. P. de Vries. 2007. The CSIRO enterprise search test collection. In *ACM SIGIR Forum* (Vol. 41, No. 2, pp. 42-45). ACM.
- Balog, K. 2008. People search in the enterprise. PhD thesis, University of Amsterdam.
- _____, & M. de Rijke. 2007. *Determining expert profiles (with an application to expert finding)*. In Proceedings of the 20th International Joint Conference on Artificial Intelligence IJCAI 2007 (pp. 2657–2662).
- _____, L. Azzopardi, & M. de Rijke. 2006. *Formal models for expert finding in enterprise corpora*. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 43-50). ACM.
- _____, T. Bogers, L. Azzopardi, M. de Rijke, & A. Van Den Bosch. 2007. *Broad expertise retrieval in sparse data environments*. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 551-558) ACM.
- _____, M. de Rijke, & W. Weerkamp. 2008. *Bloggers as experts*. In *SIGIR '08*. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval.
- _____, Y. Fang, M. de Rijke, P. Serdyukov, & L. Si. 2012. Expertise retrieval. *Foundations and Trends in Information Retrieval* 6 (2–3): 127-256.
- Bogers, T., K. Kox, & A. van den Bosch. 2008. *Using citation analysis for finding experts in workgroups*. In Proc. DIR (pp. 21-28).
- Craswell, N., A. P. de Vries, & I. Soboroff. 2005. Overview of the TREC 2005 Enterprise Track. In *Trec* (5): 199-205.
- _____, D. Hawking, A. M. Vercoestre, & P. Wilkins. 2001. *P@ noptic expert: Searching for experts not just for documents*. In Ausweb Poster Proceedings, Queensland, Australia (Vol. 15, p. 17).
- Efron, M. 2009. *Finding expert authors in institutional repositories*. Proceedings of the American Society for Information Science and Technology 46 (1): 1-15.
- Fang, H., & C. Zhai. 2007. Probabilistic models for expert finding. In *ECIR*, (pp. 418–430).
- Feldman, S. & C. Sherman. 2003. The high cost of not finding information. *technical Report*. 29127 IDC
- Fox, E. A., M. P. Koushik, J. Shaw, R. Modlin, & D. Rao. 1993. *Combining evidence from multiple searches*. In The first text retrieval conference TREC-1 500 (207): 319. US Department of Commerce, National Institute of Standards and Technology.
- Haruechaiyasak, C., & A. Kongthon. 2008. *Multidisciplinary expertise retrieval with an application in r&d management*. In Proceedings of the SIGIR 2008 Workshop on Future Challenges in Expertise Retrieval (fCHER) (pp. 25-28).
- MacDonald, C. 2009. The voting model for people search. Doctoral dissertation, University of Glasgow.
- _____, & I. Unis. 2006. *Voting for candidates: adapting data fusion techniques for an expert search task*. In Proceedings of the 15th ACM international conference on Information and knowledge management (pp. 387-396). ACM.
- _____, & I. Unis. 2008. *Key blog distillation: ranking aggregates*. In Proceedings of the 17th ACM conference on Information and knowledge management (pp. 1043-1052). ACM.

- Mimno, D., & A. McCallum. 2007. *Expertise modeling for matching papers with reviewers*. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 500-509). ACM.
- Petkova, D., & W. B. Croft. 2008. Hierarchical language models for expert finding in enterprise corpora. *International Journal on Artificial Intelligence Tools* 17 (01): 5-18.
- Sanderson M., & J. Zobel. 2005. *Information retrieval system evaluation: effort, sensitivity, and reliability*. In SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (pp.162-169). ACM Press.
- Serdyukov, P. & D. Heimstra. 2008. *Being Omnipresent To Be Almighty: The Importance of the Global Web Evidence for Organizational Expert Finding*. Future Challenges in Expertise Retrieval SIGIR 2008 workshop, July 24, Singapore, pp. 17-24.
- Yimam-Seid, D. & A. Kobsa. 2003. Expert finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce* 13 (1): 1-24.

هاشم عطاپور

متولد سال ۱۳۶۲، دارای مدرک تحصیلی دکتری در رشته علم اطلاعات و دانش‌شناسی از دانشگاه تهران است. ایشان هم‌اکنون بورسیه هیئت علمی دانشگاه تبریز است.

بازیابی اطلاعات، بازیابی تخصص و یافتن خبرگان با اتکاء به مدل‌های بازیابی اطلاعات، سنجش اثرگذاری آکادمیک و اجتماعی علم، و پیاده‌سازی مدیریت دانش از علایق پژوهشی وی است.



فاطمه فهیم‌نیا

متولد سال ۱۳۴۵، دارای مدرک تحصیلی دکتری در رشته کتابداری و اطلاع‌رسانی از دانشگاه تهران است. ایشان هم‌اکنون دانشیار و مدیر گروه علم اطلاعات و دانش‌شناسی دانشگاه تهران است.

حوزه مطالعات اقتصادی و مدیریت اطلاعات، آموزش علم اطلاعات، فراهم‌آوری و مطالعات کودکان از جمله علایق پژوهشی وی است.



نادر نقشینه

متولد سال ۱۳۴۰، دارای مدرک تحصیلی دکتری در رشته کتابداری و اطلاع‌رسانی از دانشگاه تهران است. ایشان هم‌اکنون مشاور فنی کتابخانه مرکزی و مرکز اسناد دانشگاه تهران و نیز رئیس آزمایشگاه پایش دیجیتالی است.

مسائل مربوط به زندگی دیجیتال، Digital Emergence و حفاظت و رخنه دیجیتال از جمله علایق پژوهشی وی است.

