Extraction of Effective Textual and Semantic Features in Learning to Rank for Web Document Retrieval

Mohaddeseh Mahjoob
Master of Computer Engineering; Faculty of Engineering; Computer and Information Technology Engineering Department; Ferdowsi University; Mashhad, Iran;
Email: mohaddeseh.mahjoob@yahoo.com

Faezeh Ensan*
Assistant Professor; Computer Engineering; Faculty of Engineering; Computer and Information Technology Engineering Department; Ferdowsi University; Mashhad, Iran;
Email: ensan@um.ac.ir

Sanaz Keshvari
PhD Candidate in Computer Engineering; Faculty of Engineering; Computer and Information Technology Engineering Department; Ferdowsi University; Mashhad, Iran Email: s.keshvari@mail.um.ac.ir

Parastoo Jafarzadeh
PhD Candidate in Computer Engineering; Faculty of Engineering; Computer and Information Technology Engineering Department; Ryerson University; Toronto, Canada;
Email: p.jagarzadehhesar@mail.um.ac.ir

Mohammadamin Keyvanzad
Master of Computer Engineering; Faculty of Computer Science; University of New Brunswick; New Brunswick, Canada;
Email: Keyvanzad.amin@unb.ca

Iranian Journal of Information Processing and Management

Received: 24, Oct. 2019 **Accepted: 07, Feb. 2021**

Abstract: Ranking algorithms, as the core of web search systems, are responsible for finding and ranking the most relevant documents to user information needs from the crawled and indexed corpus. With the ever-increasing amount of available training data, ranking technologies are moving towards using Machine Learning methods, described as Learning to Rank algorithms. The basic Learning to Rank systems mainly have used textual features while ignoring semantic features. With the advent of Semantic Web, there is an emerging interest in developing and using semantic features for Learning to Rank systems. An important

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 36 | No. 4 | pp. 1081-1112
Summer 2021
https://doi.org/10.52547/jipm.36.4.1081

* Corresponding Author

challenge is that there is currently no comprehensive study on the combined usage of textual and semantic features for Learning to Rank systems. In this paper, first, we define and implement four new sets of semantic features based on Knowledge Graph, Entity Repetition, Textual Fields and Vector Representation of Words and Texts. For experimental analysis, we used the MQ-2007 dataset from LETOR 4, which includes a set of textual features. The results of running six standard Learning to Rank Algorithms show that by using semantic features, either in isolation or in combination with textual features, significantly increases the performance. The increase in performance is even more significant when we limit the tests to hard queries. We also implemented an existing Feature Selection algorithm to test whether it can improve the results even further. The results showed improvements for some Learning to Rank algorithms, yet failed to improve on others.

Keywords: Learning to Rank, Semantic Features, Web Document Retrieval, Linked Data, Feature Selection

استخراج ویژگی‌های متنی و معنایی در یادگیری رتبه‌بندی جهت بازیابی اسناد وب

محدثه محجوب
کارشناسی ارشد مهندسی کامپیوتر؛ دانشکدهٔ مهندسی؛ گروه کامپیوتر؛ دانشگاه فردوسی مشهد؛ ایران؛
mohaddeseh.mahjoob@yahoo.com

فائزه انسان
دکتری علوم کامپیوتر؛ استادیار؛ دانشکدهٔ مهندسی؛ دانشگاه فردوسی مشهد؛ مشهد، ایران؛
پدیدآور رابط ensan@um.ac.ir

ساناز کشوری
دانشجوی دکتری مهندسی کامپیوتر؛ دانشکدهٔ مهندسی؛ گروه کامپیوتر؛ دانشگاه فردوسی مشهد؛ مشهد، ایران؛
s.keshvari@mail.um.ac.ir

پرستو جعفرزاده
دانشجوی دکتری مهندسی کامپیوتر؛ گروه مهندسی برق کامپیوتر و بیومدیکال؛ دانشگاه رایرسون، کانادا؛
parastoo.jafarzadeh@ryerson.ca

محمدامین کیوان‌زاد
کارشناسی ارشد مهندسی کامپیوتر؛ دانشگاه نیوبرانزویک، کانادا Keyvanzad.amin@unb.ca

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۳۹۸/۰۸/۰۲ پذیرش: ۱۳۹۹/۱۱/۱۹ مقاله برای اصلاح به مدت ۴ ماه نزد پدیدآوران بوده است.

چکیده: با ظهور وب معنایی، تعریف و استفاده از ویژگی‌های معنایی در الگوریتم‌های یادگیری رتبه‌بندی هم مطرح شده است. یک چالش مهم در این زمینه عدم استفاده از ویژگی‌های جامع و همچنین، عدم ترکیب کامل از ویژگی‌های متنی و معنایی است. در این مقاله، با تعریف ویژگی‌های معنایی جدید در چهار دستهٔ ویژگی‌های مبتنی بر گراف و پایگاه دانش، ویژگی‌های مبتنی بر تکرار موجودیت، ویژگی‌های مبتنی بر فیلدهای متنی، و ویژگی‌های مبتنی بر نمایش برداری کلمات و متون به این چالش پاسخ داده شده است. جهت ارزیابی از مجموعهٔ دادهٔ MQ-2007 متعلق به LETOR4، که حاوی ویژگی‌های متنی آماده است، و شش الگوریتم یادگیری رتبه‌بندی استاندارد استفاده شده است. نتایج تجربی نشان می‌دهد که ویژگی‌های معنایی و نیز ترکیب آن‌ها با ویژگی‌های متنی باعث بهبود ۵۰ درصدی نسبت به استفاده از تنها ویژگی‌های متنی می‌شوند. در انتها، از یک الگوریتم انتخاب ویژگی برای انتخاب بهترین ویژگی‌های معنایی استفاده شده که منجر به بهبود ۷ درصدی نسبت به

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۶ | شمارهٔ ٤ | ص‌ص ۱۰۸۱-۱۱۱۲
تابستان ۱۴۰۰
https://doi.org/10.52547/jipm.36.4.1081

الگوریتم‌های رتبه‌بندی بدون انتخاب ویژگی شده است.

کلیدواژه‌ها: یادگیری رتبه‌بندی، ویژگی‌های معنایی، بازیابی اسناد وب، داده‌های پیوندی

۱. مقدمه

در سال‌های اخیر، با رشد سریع وب جهان‌گستر دست‌یابی به اطلاعات مورد نیاز از طریق جست‌وجو در وب به‌طور معمول، برای عموم کاربران غیرممکن شده است. در نتیجه، بازیابی اطلاعات کارا و مؤثر مهم‌تر از قبل شده و موتورهای جست‌وجو برای بسیاری از افراد به ابزاری اساسی تبدیل شده‌اند. به‌طور معمول، شش مؤلفهٔ اصلی در یک موتور جست‌وجو وجود دارد: خزنده[1]، تجزیه‌کننده[2]، نمایه‌ساز[3]، تحلیل‌گر اتصال[4]، پردازش‌گر پرس‌وجو[5]، و رتبه‌بند[6] (Liu 2011).

رتبه‌بند که عنصر مرکزی موتور جست‌وجوست، مسئول تطابق بین پرس‌وجوهای پردازش‌شده و اسناد نمایه‌شده است. به‌دلیل نقش کلیدی رتبه‌بند در موتورهای جست‌وجو، در سال‌های اخیر توجه ویژه‌ای به تحقیقات و توسعه در زمینهٔ فناوری‌های رتبه‌بندی به عمل آمده است. به این منظور، مدل‌های رتبه‌بندی اکتشافی بسیاری پیشنهاد و در ادبیات بازیابی اطلاعات وارد شده‌اند. برای بهبود نتایج حاصل از فناوری‌های یادگیری رتبه‌بندی می‌توان الگوریتم یادگیری ماشین جدیدی ایجاد کرد و یا ویژگی‌هایی را که به این الگوریتم‌ها اعمال می‌شوند، بهبود بخشید. فناوری‌های یادگیری رتبه‌بندی پایه، به‌طور عام از ویژگی‌های متنی و غیرمعنایی که از پرس‌وجو و اسناد استخراج شده‌اند، استفاده می‌کنند. استفاده از ویژگی‌های متنی به‌تنهایی می‌تواند باعث بروز مشکلاتی شود که برخی از آن‌ها عبارت‌اند از: عدم تطابق کلمات در سند و پرس‌وجو، و استفاده از کلمات یکسان با معانی متفاوت در سند و پرس‌وجو.

ظهور وب معنایی حجم زیادی از اطلاعات ساختاری و قابل فهم برای ماشین در مورد طیف وسیعی از اشیا را در دسترس قرار داده است (Guha, McCool and Miller 2003). وجود این اطلاعات فرصتی مناسب برای تعریف ویژگی‌های معنایی و بهبود فناوری‌های یادگیری رتبه‌بندی فراهم آورده است. اغلب روش‌هایی که از این ویژگی‌ها استفاده

1. crawler
2. parser
3. indexer
4. link analyzer
5. query processor
6. ranker

می‌کنند، جنبه‌های خاصی از معنا را در نظر گرفته و تمام ابعاد آن را پوشش نمی‌دهند. همچنین، اکثر این روش‌ها، ویژگی‌های متنی را یا به‌طور کامل نادیده گرفته و یا به شکلی محدود پوشش می‌دهند. ویژگی‌های معنایی اگرچه به‌دلیل وارد کردن معنا در جست‌وجو اغلب نتایج مطلوبی برمی‌گردانند، اما نادیده‌گرفتن ویژگی‌های متنی و تنها استفاده از ویژگی‌های معنایی، در مواقعی که هدف کاربر جست‌وجوی کلیدواژه‌ای خاص است، نتایج مطلوبی را برنمی‌گرداند. بنابراین، به نظر می‌رسد استفاده از ترکیب ویژگی‌های معنایی و متنی می‌تواند نتایج بهتری را به بار آوَرَد. با توجه به مطالب بیان‌شده، هدف کلی این تحقیق ارائهٔ یک سیستم یادگیری رتبه‌بندی است که از ویژگی‌های معنایی در کنار ویژگی‌های متنی به‌منظور بازیابی سند استفاده نماید.

با وجود اینکه در سال‌های اخیر به استفاده از روش‌های یادگیری رتبه‌بندی، به‌عنوان فناوری مورد استفاده در سیستم‌های رتبه‌بند توجه ویژه‌ای شده است، اما این روش‌ها همچنان با چالش‌هایی روبه‌رو هستند؛ چالش‌هایی از قبیل: عدم تعریف و استفاده از ویژگی‌های معنایی پوشش‌دهندهٔ جنبه‌های مختلف معنا در بازیابی سند، عدم استفاده از ویژگی‌های معنایی و متنی به‌طور همزمان در سیستم‌های یادگیری رتبه‌بندی، و عدم مقایسهٔ کارکرد سیستم‌های یادگیری مختلف با ویژگی‌های معنایی.

در راستای رفع چالش‌های گفته‌شده، ارائهٔ یک سیستم یادگیری رتبه‌بندی که از ویژگی‌های معنایی در کنار ویژگی‌های متنی استفاده نماید، هدف کلی این پژوهش است. در پژوهش جاری، نخست موجودیت‌های موجود در پرس‌وجوها و اسناد مجموعهٔ دادهٔ مورد نظر را با هدف تعریف ویژگی‌های معنایی استخراج می‌کنیم. پس از استخراج موجودیت‌ها، ویژگی‌های معنایی متفاوتی تعریف می‌شوند. برای استخراج این ویژگی‌ها از اطلاعات صفحهٔ «ویکی‌پدیا»ی متناظر با هر موجودیت و همچنین، گراف موجود در «دی‌بی‌پدیا»[1] استفاده می‌کنیم. پس از تعریف و پیاده‌سازی ویژگی‌ها در دسته‌های بیان‌شده، باید بهترین ویژگی‌ها را از بین ویژگی‌های پیاده‌سازی‌شده انتخاب کنیم. برای انتخاب ویژگی‌ها، از روش‌های مختلف انتخاب ویژگی[2] استفاده کرده و بهترین ویژگی‌های معنایی انتخاب می‌شوند. پس از این مرحله، ویژگی‌های معنایی منتخب را با ویژگی‌های متنی ترکیب کرده و نتایج را بررسی می‌کنیم. در پایان شش الگوریتم یادگیری رتبه‌بندی

1. DBpedia
2. feature selection

«مـارت»[1]، «رنک‌بوسـت»[2]، «آدارنـک»[3]، «کوردینـت اسـنت»[4]، «لامبادامـارت»[5]، و «رنـدوم فارسـت»[6] را انتخـاب کـرده و آن‌هـا را بـا اسـتفاده از ویژگی‌هـای تعریف‌شـده ارزیابـی می‌کنیـم.

در پژوهش‌هـای پیشـین، مقایسـهٔ کاملـی از اعمـال ترکیبـی از ویژگی‌هـای معنایـی و متنـی بـر الگوریتم‌هـای مختلـف یادگیـری رتبه‌بنـدی صـورت نگرفتـه اسـت. مشـخص نیسـت کدام‌یـک از الگوریتم‌هـای پایـهٔ یادگیـری رتبه‌بنـدی بـرای یادگیـری ویژگی‌هـای معنایـی بهتـر هسـتند؟ و همچنیـن، کـدام الگوریتـم رتبه‌بنـدی بـه ترکیـب ویژگی‌هـای معنایـی و متنـی پاسـخ بهتـری می‌دهـد؟ انجـام چنیـن مقایسـه‌ای در پژوهـش جـاری، اولیـن پژوهـش در ایـن زمینـه خواهـد بـود.

در بخـش دوم، پیشـینهٔ پژوهـش آمـده و سـپس، روش پژوهـش در قسـمت سـوم مطـرح شـده اسـت. ارزیابـی روش پیشـنهادی، معیارهـا و نتایـج در قسـمت چهـارم بیـان می‌شـود. در قسـمت پنجـم، بـه نتیجه‌گیـری روش پیشـنهادی پرداختـه و پیشـنهادهایی بـرای کارهـای آتـی ارائـه می‌شـود.

۲. پیشینهٔ پژوهش

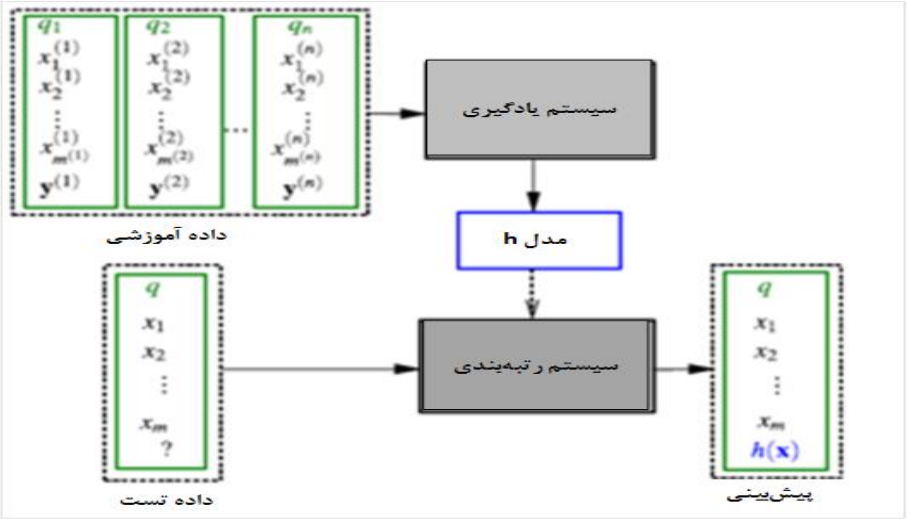
وظایـف بسـیاری در بازیابـی اطلاعـات و پـردازش زبـان طبیعـی[7] وجـود دارد کـه رتبه‌بنـدی مسـئلهٔ اساسـی در آن‌هاسـت (Li 2011). به‌دلیـل نقـش کلیـدی رتبه‌بنـد، توجـه ویـژه‌ای بـه تحقیقات و توسـعهٔ فناوری‌هـای رتبه‌بنـدی شـده اسـت. اخیـراً، داده‌هـای آموزشـی کـه به‌طـور بالقـوه در دسـترس هسـتند، اسـتفاده از فناوری‌هـای یادگیـری ماشـین در سـاخت مدل‌هـای رتبه‌بنـدی کارا را ممکـن سـاخته اسـت (Liu 2011). روش‌هایـی کـه از فناوری‌هـای یادگیـری ماشـین بـرای حـل مسـئلهٔ رتبه‌بنـدی اسـتفاده می‌کننـد، به‌عنـوان روش‌هـای «یادگیـری رتبه‌بنـدی» شـناخته می‌شـوند (Phophalia 2011). در ادامـه، ابتـدا مؤلفه‌هـای اساسـی در مسـائل یادگیـری ماشـین معرفـی می‌شـوند. سـپس، چارچـوب کلـی بـرای یادگیـری رتبه‌بنـدی ارائـه شـده و روش‌هـای موجـود در یادگیـری رتبه‌بنـدی بـر اسـاس مؤلفه‌هـای یادگیـری ماشـین دسـته‌بندی می‌شـوند.

جریـان عمومـی یادگیـری رتبه‌بنـدی در شـکل ۱، آمـده اسـت. بـا توجـه بـه ایـن شـکل، درمی‌یابیـم از آن‌جـا کـه یادگیـری رتبه‌بنـدی، نوعـی از یادگیـری بـا ناظـر[8] اسـت، مجموعـهٔ آموزشـی مـورد نیـاز اسـت. ایجـاد مجموعـهٔ آموزشـی بسـیار شـبیه بـه ایجـاد مجموعـهٔ تسـت

1. MART
2. RankBoost
3. AdaRank
4. Coordinate Ascent
5. LambdaMART
6. Random Forests
7. natural language processing (NLP)
8. supervised learning

برای ارزیابی است. برای مثال، یک مجموعهٔ آموزشی معمولی از سه جزء تشکیل شده است که عبارت‌اند از: n پرس‌وجوی آموزشی $q_i\ (i=1,2,...,n)$، اسناد مرتبط با این پرس‌وجوها که به‌وسیلهٔ بردار ویژگی $x^{(i)}=\{x_j^{(i)}\}_{j=1}^{m^{(i)}}$ نشان داده می‌شوند که در آن $m^{(i)}$ تعداد اسناد مرتبط با پرس‌وجوی q_i است و هر سند امتیاز متناظر را در بردار y دارد.

سپس، الگوریتم یادگیری خاصی برای یادگیری مدل رتبه‌بندی استفاده می‌شود؛ به‌طوری که خروجی مدل رتبه‌بندی می‌تواند برچسب حقیقی در مجموعهٔ آموزشی را پیش‌بینی کند. در مرحلهٔ تست، هنگامی‌که پرس‌وجوی جدیدی وارد می‌شود، مدل یادگرفته‌شده برای مرتب نمودن اسناد و برگرداندن لیست مرتب‌شدهٔ متناظر با پرس‌وجوی کاربر به او استفاده می‌شود (Liu 2009).



شکل ۱. چارچوب یادگیری رتبه‌بندی (Liu 2009)

الگوریتم‌های یادگیری رتبه‌بندی بسیاری را می‌توان در چارچوب نشان داده‌شده در شکل استفاده کرد. به‌منظور درک بهتر، این الگوریتم‌ها بر اساس چهار جزء یادگیری ماشین به سه دسته تقسیم می‌شوند: روش نقطه‌ای[1]، روش جفتی[2] و روش لیستی[3] (Liu 2011).

روش نقطه‌ای: درجهٔ ارتباط برای هر سند در این روش با صفر به معنای نامرتبط، و یک به معنای مرتبط بیان می‌شود. در روش نقطه‌ای تابع زیان، درست بودن پیش‌بینی برای هر

1. pointwise 2. pairwise 3. listwise

سند واحد را بررسی می‌کند (Liu 2011; Guha, McCool and Miller 2003; Li 2011). الگوریتم «رندوم فارست» از الگوریتم‌های این زمینه است (Breiman 2001).

روش جفتی: ترتیب نسبی بین دو سند را بیان می‌کند. در این روش، رتبه‌بندی به‌طور معمول، به مسئلهٔ طبقه‌بندی روی زوج اسناد کاهش می‌یابد تا تعیین شود که کدام سند در جفت سند داده‌شده، اولویت دارد. الگوریتم‌های «رنک‌بوست» (Freund et al. 2003) و «لامبادامارت» (Wu et al. 2010) از جمله الگوریتم‌های مشهور در این دسته هستند.

«مارت» روشی است که خروجی آن ترکیب خطی وزن‌دار از مجموعه درخت‌های رگرسیون و ساخت هر درخت رگرسیون با هدف کمینه‌کردن تابع خطا در امتداد کاهش گرادیان[1] صورت می‌گیرد. «رنک‌بوست» روشی است که مبتنی بر روش تقویت عمل می‌کند و در هر مرحله سعی می‌کند رتبه‌بند ضعیفی بسازد که کمترین خطا را از نظر تشخیص نادرست ترتیب بین جفت اسناد داشته باشد. سپس، در هر مرحله، وزن اسناد را به‌صورت دو-به-دو تغییر می‌دهد. در نهایت، ساخت مدل نهایی با ترکیب این رتبه‌بندها صورت می‌گیرد. «لامبادامارت» در راستای الگوریتم «لامبادارنک» ارائه شد که از شبکهٔ عصبی برای ساخت مدل رتبه‌بند خود استفاده می‌کند. گرادیان تابع خطا به‌عنوان تابع «لامبادا» تعریف می‌شود. این روش به‌جای ساخت مدل رتبه‌بند از طریق شبکهٔ عصبی، از روش درختان تقویت بهره می‌برد.

روش لیستی: فضای ورودی روش لیستی از مجموعه اسناد متناظر با پرس‌وجوی q، برای مثال $x = \{x_j\}_{j=1}^{m}$ تشکیل شده است. همچنین، فضای خروجی این روش شامل لیست رتبه‌بندی‌شدهٔ اسناد است. مجموعه‌ای از اسناد به‌عنوان ورودی دریافت شده و جایگشت آن‌ها توسط این روش پیش‌بینی می‌شود. الگوریتم «آدارنک» (Xu, Li 2007) نمونه‌ای از الگوریتم‌های یادگیری رتبه‌بندی لیستی است. در «آدارنک» یادگیرهای ضعیف بر روی داده‌های آموزشی که توزیع آن‌ها بر اساس یادگیر قبلی تغییر یافته، صورت می‌گیرد و ساخت یادگیر ضعیف در هر مرحله با انتخاب یک ویژگی با هدف کمینه‌کردن تابع خطا صورت می‌گیرد. الگوریتم ارائه‌شده توسط (Ai et al. 2019) الگوریتمی است که روش لیستی را با ترکیب چندین روش جفتی ارائه می‌دهد.

در ادامه، ویژگی‌های استخراج‌شده در سیستم‌های یادگیری رتبه‌بندی مورد ارزیابی

1. gradient

قرار می‌گیرند. الگوریتم‌های یادگیری رتبه‌بندی پایه به‌طور معمول، از ویژگی‌های متنی مستقل و یا مرتبط با پرس‌وجو، به‌منظور رتبه‌بندی اسناد متناظر با پرس‌وجو استفاده می‌کنند. برای مثال، در الگوریتم «ادارنک» (Xu, Li 2007)، برخی از ویژگی‌های استفاده‌شده ویژگی‌های مبتنی بر تکرار مانند TF-IDF و ویژگی‌های مبتنی بر تحلیل لینک مانند PageRank هستند.

در پژوهشی که Bendersky, Croft and Diao (2011) انجام داده‌اند، ویژگی‌هایی مانند نسبت کلمات ایست[1] به کلمات غیر ایست، تعداد کلمات در فیلد عنوان، عمق مسیر URL و نسبت جداول در متن تعریف شده‌اند. این ویژگی‌ها که مبتنی بر کیفیت محتوای سند هستند، به‌منظور رتبه‌بندی اسناد وب پیشنهاد شده‌اند. در مقالهٔ Macdonald, Santos and Ounis (2013) نیز ویژگی‌های متنی در دسته‌های مختلفی تعریف شده است. پژوهش Macdonald, Dinçer and Ounis (2015) از ویژگی‌های تعریف‌شده در دو مجموعه دادهٔ LETOR (Qin et al. 2010) و MSLR[2] استفاده کرده و آن‌ها را با ویژگی‌های تعریف‌شده در مقالهٔ Macdonald, Santos and Ounis (2013) ترکیب نموده است. استفاده از ویژگی معنایی برای درک بهتر از پرس‌وجوی کاربر صورت می‌گیرد. در روشی دیگر برای دریافت بهتر مفهوم پرس‌وجوی کاربر از ویژگی‌های متنی مبتنی بر رویداد[3] استفاده می‌شود (Zhoa et al. 2020). در تحقیقاتی که مبتنی بر رویداد هستند، پارامتر زمان پرس‌وجو بسیار پُراهمیت است و در تحقیق Maio et al. (2019) از شبکه‌های عصبی عمیق برای یادگیری این رویدادها استفاده شده است. در بیشتر الگوریتم‌های اخیر از شبکه‌های عصبی عمیق استفاده می‌شود (Ai et al. 2018; Pasumarthi et al. 2019).

دستهٔ دیگر از کارهای مرتبط، سیستم‌های یادگیری هستند که از ویژگی‌های رتبه‌بندی مبتنی بر گراف دانش و معنایی برای رتبه‌بندی استفاده می‌کنند. هدف پژوهش‌های انجام‌شده توسط Dali et al. (2012), Semeraro (2016), Schuhmacher, Dietz and Ponzetto (2015), Chen, Xiong & Callan (2016) استخراج ویژگی‌های معنایی با هدف رتبه‌بندی موجودیت[4] است. جست‌وجوی معنایی[5]، کاربردی از وب معنایی برای جست‌وجوست. جست‌وجوی معنایی در تلاش است که نتایج حاصل از جست‌وجوی سنتی را با استفاده

1. stop words
2. Http://research.microsoft.com/~MSLR
3. event-oriented text retrieval
4. entity
5. semantic search

از داده‌های وب معنایی بهبود بخشد. استفاده از جست‌وجوی معنایی در کنار یادگیری رتبه‌بندی می‌تواند باعث بهبود نتایج جست‌وجو گردیده و چالش‌های جست‌وجوی سنتی را برطرف کند. پژوهش حاضر، رتبه‌بندی اسناد با استفاده از ویژگی‌های معنایی خواهد بود. یکی دیگر از سیستم‌های یادگیری رتبه‌بندی در پژوهش Han et al. (2013) با هدف بازیابی و رتبه‌بندی توییت معرفی شده است.

از دیگر کاربردهای سیستم‌های یادگیری رتبه‌بندی، استفاده از آن‌ها در بازیابی خبره[1] است. بازیابی خبره، فرایند شناسایی خبره برای یک موضوع خاص است. در پژوهش Zheng (2013) از اطلاعات پروفایل خبره و همچنین اسناد مرتبط با آن استفاده شده است. در این روش، با کمک مدل‌های زبانی، احتمال اینکه یک کاندید، در زمینهٔ مورد نظر خبره باشد، محاسبه گردیده و به‌عنوان ویژگی در نظر گرفته می‌شود.

در پژوهش Dali et al. (2012) با هدف رتبه‌بندی موجودیت‌های RDF[2]، ویژگی‌های مبتنی بر فرکانس[3] و ویژگی‌های مبتنی بر مرکزیت[4] استخراج گردیده و در سیستم یادگیری رتبه‌بندی استفاده شده است. ویژگی‌های مبتنی بر فرکانس از شمارش الگوهای مختلف در گراف RDF یا شمارش تعداد رخداد در نتایج جست‌وجوی وب یا پایگاه داده‌های n-gram به‌دست می‌آید، در حالی که ویژگی‌های مبتنی بر مرکزیت با اعمال الگوریتم‌های تئوری گراف، مانند PageRank یا HITS[5] بر گراف RDF حاصل می‌شوند.

برای استخراج ویژگی‌های معنایی به‌طور معمول از اطلاعات موجود در پایگاه‌های دانش مختلف استفاده می‌شود. برای نمونه، در پژوهش Semeraro (2016) از اطلاعات پایگاه دانش «ویکی‌پدیا» و در پژوهش Xiong & Callan (2015) از اطلاعات پایگاه دانش «فری‌بیس»[6] برای استخراج ویژگی‌های معنایی از جمله شباهت متنی، همپوشانی هستی‌شناسی[7]، تعداد تکرار شیء، میزان شباهت شیء با سایر اشیا و اتصالات گراف استفاده شده است.

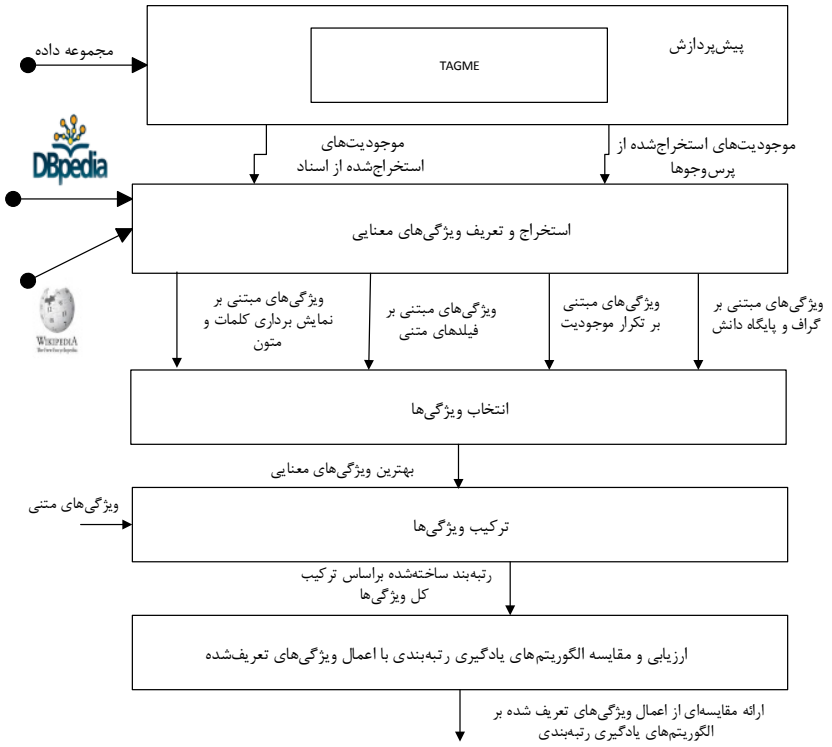
در پژوهش Ensan et al. (2017) به‌منظور بازیابی سند، از ویژگی‌های مبتنی بر تعبیهٔ کلمه[8] و تعبیهٔ سند[9] استفاده شده است Soldaini & Goharian (2017). به بازیابی اسناد در دامنهٔ پزشکی پرداخته و افزون بر تحلیل معنایی نهان، از تکنیک تعبیهٔ کلمه نیز استفاده کرده و ویژگی‌هایی مبتنی بر آن را تعریف نموده است. روش‌های ارائه‌شده اگرچه هدفی یکسان

1. expert finding
2. resource description framework
3. frequency
4. centrality
5. hyperlink-induced topic search (HITS)
6. Freebase
7. ontology overlap
8. word embedding
9. document embedding

با هدف این مقاله را دنبال می‌کنند، اما روش‌های پیشنهادی ویژگی‌های متفاوتی را که ابعاد مختلف موجودیت را در نظر بگیرند، ارائه نکرده‌اند. برای نمونه، در مقالهٔ Ensan et al. (2017) تنها ویژگی‌های مبتنی بر نمایش بُرداری کلمات و موجودیت‌ها تعریف شده است. افزون بر آن، بر خلاف روش پیشنهادی در این مقاله که محدود به دامنهٔ خاصی نیست، برخی از روش‌ها مانند پژوهش Soldaini & Goharian (2017) در دامنهٔ خاصی هستند. در این تحقیقات اگرچه رتبه‌بندی اسناد انجام شده، ولی ابعاد مختلف ویژگی‌ها بررسی نشده است. ما در این مقاله گزارشی دقیق از مقایسهٔ روش‌های مختلف و تأثیر ویژگی‌های مختلف بر پرس‌وجوها خواهیم داشت.

۳. روش پیشنهادی

در شکل ۲، چارچوب کلی سیستم ارائه‌شده در این مقاله نمایش داده شده است. این سیستم ابتدا پیش‌پردازشی روی مجموعهٔ داده، که در بخش ۳-۱ توضیح داده شده، انجام داده و موجودیت‌های موجود در پرس‌وجوها و اسناد مجموعه داده را استخراج می‌نماید. در مرحلهٔ بعد، ویژگی‌های معنایی بین این موجودیت‌ها تعریف و پیاده‌سازی می‌شود. در مرحلهٔ ۳، بهترین ویژگی‌های معنایی انتخاب شده و در فاز بعدی به‌کار می‌رود. در مرحلهٔ ترکیب، تمام ویژگی‌های معنایی و همچنین، بهترین ویژگی‌های معنایی با ویژگی‌های متنی ترکیب شده و سرانجام، نتیجهٔ اعمال ویژگی‌های تعریف‌شده بر روی الگوریتم‌های یادگیری رتبه‌بندی مختلف مورد بررسی قرار می‌گیرد.



شکل ۲. چارچوب سیستم پیشنهادی

در نخستین مرحله از راهکار پیشنهادی، نخست، مجموعهٔ دادهٔ مورد نظر را که مناسب هدف پژوهش بوده و در بخش‌های بعدی معرفی می‌گردد، جمع‌آوری نموده و سپس، موجودیت‌های موجود در پرس‌وجوها و اسناد را با هدف تعریف ویژگی‌های معنایی استخراج کرده‌ایم.

پس از تعریف ویژگی‌های معنایی متفاوت، باید در این مرحله بهترین ویژگی‌های معنایی تعریف‌شده را انتخاب نماییم. در این مقاله نخست، از هر ویژگی به‌تنهایی برای ساخت مدل استفاده شده و سپس، از معیار NDCG@10 که در بخش بعد توضیح داده می‌شود، استفاده نموده و امتیاز هر ویژگی را محاسبه کرده‌ایم. برای اندازه‌گیری شباهت بین ویژگی‌ها نیز از معیار «اسپیرمن» که در پژوهش Geng et al. (2007) معرفی شده است. سپس، امتیاز هر ویژگی را به‌عنوان امتیاز اهمیت آن ویژگی در نظر می‌گیریم. شباهت بین ویژگی‌ها نیز به‌عنوان امتیاز شباهت در نظر گرفته می‌شود. در نهایت، از الگوریتم

حریصانهٔ[1] ارائه‌شده در پژوهش Geng et al. (2007) استفاده می‌شود تا بهترین زیرمجموعه از ویژگی‌های معنایی برای مجموعهٔ دادهٔ مخصوص انتخاب شود.

پس از انتخاب بهترین ویژگی‌های معنایی در مرحلهٔ انتخاب ویژگی، ویژگی‌های منتخب را با ویژگی‌های متنی مجموعهٔ دادهٔ مورد نظر ترکیب کرده و نتیجه را ارزیابی نموده‌ایم. افزون بر ترکیب بیان‌شده، تمام ویژگی‌های معنایی استخراج‌شده را نیز با ویژگی‌های متنی ترکیب کرده و نتیجهٔ ترکیب حاصل را ارزیابی کرده‌ایم. در جدول (الف) مندرج در پیوست، مقایسه‌ای از کارهای گذشته ارائه می‌شود.

۳-۱. مجموعهٔ داده‌ها

در این مقاله از مجموعهٔ دادهٔ LETOR 4.0 استفاده شده است. نسخهٔ چهار مجموعهٔ دادهٔ LETOR در ماه جولای ۲۰۰۹، منتشر شده است. مجموعهٔ اسناد استفاده‌شده در این مجموعهٔ داده، مجموعهٔ صفحات وب Gov2 است که شامل ۲۵ میلیون صفحه است. همچنین، دو مجموعه پرس‌وجوی MQ2007 و MQ2008 در این مجموعهٔ داده استفاده شده است. این دو مجموعهٔ به‌ترتیب، شامل حدود ۱۷۰۰ و ۸۰۰ پرس‌وجو هستند. مجموعهٔ دادهٔ گفته شده حاوی ۴۶ ویژگی متنی است که برخی از آن‌ها عبارت‌اند از: ویژگی‌هایی مانند تعداد تکرار کلمات (TF)[2]، معکوس تکرار کلمات در اسناد (IDF)[3]، و TF*IDF که این معیارها برای هر سند محاسبه شده است (Qin, Liu 2013).

۳-۲. استخراج ویژگی

به‌منظور پیاده‌سازی این روش، نخست، مجموعهٔ دادهٔ معرفی‌شده در قسمت قبل را جمع‌آوری نموده و سپس، موجودیت‌های موجود در پرس‌وجوها و اسناد را با هدف تعریف ویژگی‌های معنایی استخراج کرده‌ایم. برای انجام این کار از ابزار «تگ‌می»[4] استفاده شده است که موجودیت‌های موجود در متن را استخراج کرده و شناسهٔ صفحهٔ «ویکی‌پدیا»ی متناظر با آن موجودیت را تولید می‌کند. «تگ‌می» ابزاری برای حاشیه‌نویسی است و نحوهٔ کار آن به این شکل است که متن ورودی را گرفته و حاشیه‌هایی مناسب و مرتبط با موضوع متن را به ما برمی‌گرداند. به‌عبارت دیگر، «تگ‌می» یک سیستم حاشیه‌نویس

1. greedy algorithm
2. term frequency (TF)
3. inverse document frequency (IDF)
4. TagMe

خودکار و ابزار قدرتمندی برای تشخیص زیررشته‌های معنادار در یک متن غیرساخت‌یافته بوده که حاشیهٔ موجود در متن ورودی را به صفحهٔ مربوط در «ویکی‌پدیا» لینک می‌دهد (Ferragina, Scaiella 2010).

در این مرحله، پس از استخراج موجودیت‌ها، ویژگی‌های معنایی مختلفی را تعریف کرده‌ایم. برای استخراج این ویژگی‌ها از اطلاعات پیشینهٔ تحقیق و صفحات «ویکی‌پدیا»ی متناظر با هر موجودیت و همچنین، گراف موجود در «دی‌بی‌پدیا»[1] استفاده شده است. ویژگی‌های تعریف‌شده، بر اساس مکانیزم استخراج آن‌ها و همچنین، مرتبط یا مستقل بودن از پرس‌وجو دسته‌بندی شده‌اند. این دسته‌بندی‌ها و برخی از ویژگی‌های هر دسته در ادامه آمده است.

- ◇ ویژگی‌های مبتنی بر گراف و پایگاه دانش: این‌ها ویژگی‌هایی هستند که از گراف موجودیت‌ها و روابط بین آن‌ها استخراج می‌شوند. از جملهٔ این ویژگی‌ها می‌توان به تشابه دسته‌بندی، تشابه صفحاتی که موجودیت‌های سند و پرس‌وجو به آن‌ها لینک داده‌اند، تعداد لینک‌های مستقیم و متمایز موجودیت‌های پرس‌وجو به موجودیت‌های سند، تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند نقش فاعل[2] را دارد، و تعداد انواع سندهای خروجی از نود متناظر با موجودیت سند اشاره کرد.
- ◇ ویژگی‌های مبتنی بر تکرار موجودیت: این ویژگی‌ها عبارت‌اند از: تعداد رخداد موجودیت‌های پرس‌وجو در سند، تعداد موجودیت‌های موجود در سند، نرمال‌شدهٔ تعداد رخداد موجودیت‌های پرس‌وجو در سند، رابطهٔ کل اسناد مجموعه و تعداد اسنادی که موجودیت مورد نظر ما در آن‌ها ظاهر شده است، امتیاز BM25 رخداد موجودیت‌های پرس‌وجو در موجودیت‌های سند را بیان می‌کند. در «دی‌بی‌پدیا» خصیصه‌ای وجود دارد که موجودیت‌های مرتبط با موجودیت مورد نظر را معرفی می‌کند. در این دسته از ویژگی‌ها، برای موجودیت‌های سند و پرس‌وجو، شناسهٔ موجودیت‌های مرتبط با آن‌ها را استخراج کرده و به مجموعهٔ موجودیت‌های سند و پرس‌وجو اضافه نموده‌ایم. سپس، سایر ویژگی‌های مبتنی بر تکرار موجودیت که در این بخش تعریف شده، بر روی مجموعهٔ جدید محاسبه شده است.

1. DBpedia
2. subject

- ویژگی‌های مبتنی بر فیلدهای متنی: ویژگی‌هایی که از فیلدهای متنی موجودیت‌ها استخراج می‌شوند عبارت‌اند از: امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند، امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند، و امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند.
- ویژگی‌های مبتنی بر نمایش برداری کلمات و متون: استفاده از روش‌های تعبیهٔ کلمه و تحلیل معنایی صریح و همچنین، ساخت بردارهای متناظر با عنوان موجودیت‌های سند و پرس‌وجو. ویژگی‌های موجود در این دسته شامل شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تعبیهٔ کلمه، شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تحلیل معنایی صریح هستند. لیست کامل ویژگی‌های تعریف‌شده در این مقاله، در جدول (ب) پیوست آمده است. در مرحلهٔ انتخاب ویژگی نیز روش ارائه‌شده در پژوهش Ferragina, Scaiella (2010)، پیاده‌سازی شده است.

پس از انتخاب بهترین ویژگی‌های معنایی در مرحلهٔ انتخاب ویژگی، ویژگی‌های منتخب را با ویژگی‌های متنی مجموعهٔ دادهٔ مورد نظر ترکیب کرده و نتیجه را ارزیابی نموده‌ایم. افزون بر ترکیب بیان‌شده، تمام ویژگی‌های معنایی استخراج‌شده را نیز با ویژگی‌های متنی ترکیب کرده و نتیجهٔ ترکیب حاصل را ارزیابی کرده‌ایم.

۳-۳. ارزیابی و مقایسهٔ سیستم‌های یادگیری رتبه‌بندی

در این مرحله، شش الگوریتم یادگیری رتبه‌بندی مشهور که در بخش ۲، توضیح داده شد، انتخاب شده است. از بین الگوریتم‌های انتخاب‌شده الگوریتم «رندوم‌فارست» الگوریتم نقطه‌ای است، در حالی که سه الگوریتم انتخابی «مارت»، «لامبدامارت» و «رنک‌بوست» الگوریتم‌های جفتی هستند. علاوه بر چهار الگوریتم بیان‌شده، دو الگوریتم دیگر از دستهٔ الگوریتم‌های لیستی انتخاب شده‌اند که عبارت‌اند از: «آدارنک» و «کوردینت اسنت» (Metzler & Croft 2007). سپس، این الگوریتم‌ها را با استفاده از ویژگی‌های تعریف‌شده ارزیابی نموده و بررسی می‌کنیم که کدام روش یادگیری رتبه‌بندی برای ویژگی‌های معنایی و ترکیب ویژگی‌های معنایی و متنی مناسب‌تر است.

٤. ارزیابی سیستم

تمام پیاده‌سازی‌های سیستم پیشنهادی با زبان برنامه‌نویسی «جاوا» و «پایتون» و با استفاده از کتابخانهٔ استاندارد «رنکلیب»[1] انجام شده است. در این مقاله از سه معیار ارزیابی MAP[2]، NDCG@10[3] و NDCG@20 به‌منظور ارزیابی سیستم پیشنهادی استفاده شده است. اهمیت آماری[4] کارایی روش‌هایی که از ویژگی‌های معنایی و همچنین، ترکیب ویژگی‌های معنایی و متنی استفاده می‌کنند، نسبت به روش‌های پایه با استفاده از معیار p-value بیان شده است. مقادیر کمتر از ۰٫۰۵ نشان‌دهندهٔ آن است که کارایی بیان‌شده دارای اهمیت آماری است.

بررسی کارایی روش‌هایی که از ویژگی‌های معنایی و همچنین، ترکیب آن‌ها با ویژگی‌های متنی استفاده می‌کنند، در مقایسه با روش‌های پایه در جدول ۱، آمده است. در جدول ۱، اهمیت آماری با استفاده از † و درصد اختلاف نسبی با $\Delta\%$ نشان داده شده است. همچنین، در این جدول عبارت «متنی»[5] نشان‌دهندهٔ روش پایه است و متنی و متنی + معنایی به‌ترتیب، بیانگر حالت‌هایی هستند که فقط از ویژگی‌های معنایی و ترکیب ویژگی‌های متنی و معنایی استفاده شده است.

همان‌گونه که از جدول ۱، مشخص است، استفاده از فقط ویژگی‌های معنایی در رتبه‌بند «رندوم فارست» که یک الگوریتم نقطه‌ای است، برای تمام معیارهای ارزیابی دارای بهبود است، اما برای سایر رتبه‌بندها استفاده از فقط ویژگی‌های معنایی، به‌طور معمول، باعث کاهش کارایی شده است، به‌طوری که، تنها رتبه‌بند «مارت» برای معیارهای NDCG@20 و MAP و رتبه‌بند «کوردینت اسنت» برای معیار MAP دارای بهبود هستند.

اگرچه استفاده از ویژگی‌های معنایی به‌تنهایی و به‌طور معمول، باعث کاهش کارایی می‌شود، اما استفاده از ترکیب ویژگی‌های معنایی و متنی با یکدیگر در مقایسه با حالت پایه، در تمام رتبه‌بندها و برای معیارهای ارزیابی MAP و NDCG@10 دارای بهبود بوده و این بهبود دارای اهمیت آماری است. این ترکیب در «آدارنک»، باعث کاهش NDCG@20 شده است، اما در سایر رتبه‌بندها باعث بهبود NDCG@20 شده و این بهبود دارای اهمیت آماری است.

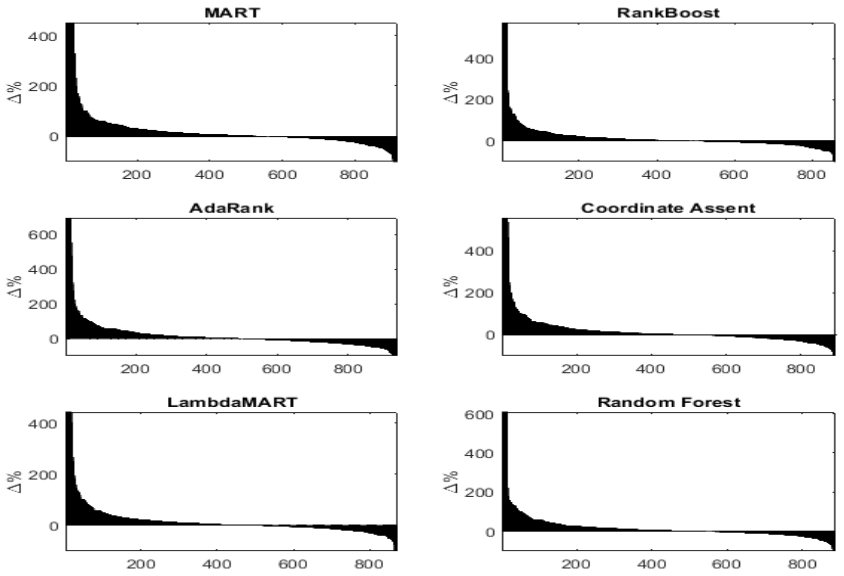
---

1. ranklib
2. mean average precision
3. normalized discounted cumulative gain
4. statistical significant
5. textual

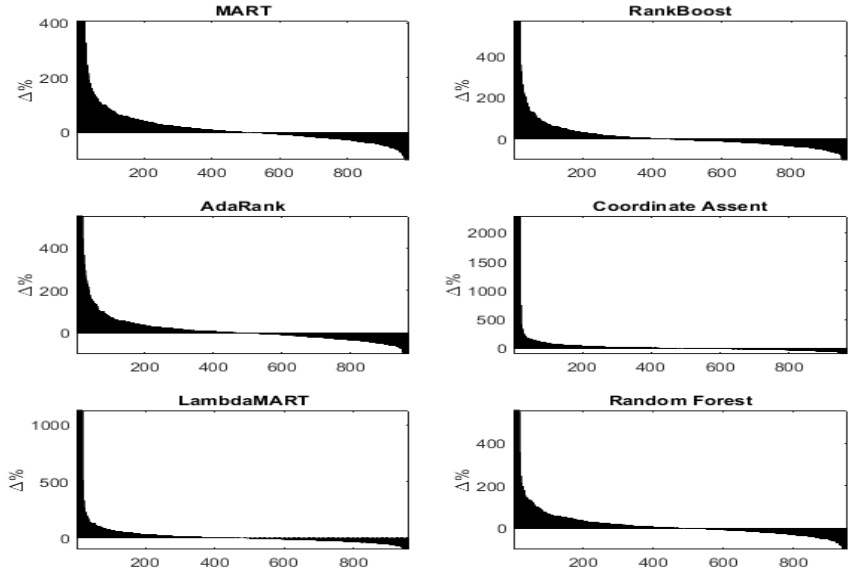
جدول ۱. ارزیابی نتایج برای تعامل ویژگی‌های معنایی و ترکیب ویژگی‌های متنی و معنایی با حالت پایه

رتبه‌بند	ویژگی‌ها	MAP	Δ%	p-value	NDCG @10	Δ%	p-value	NDCG @20	Δ%	p-value
مارت	متنی	۰/۵۰۳			۰/۵۴۲			۰/۴۶۹		
	معنایی	۰/۵۰۷	۰/۸	۰/۶۵۶۱	۰/۵۴۰	-۰/۳	۰/۵۱۵۴	۰/۴۷۲	۰/۸	۰/۷۵۵۰
	متنی+ معنایی	†۰/۵۲۴	۴/۲	۰/۰۰۰۰	†۰/۵۵۶	۲/۷۰	۰/۰۰۰۱	†۰/۴۹۰	۴/۶	۰/۰۰۰
رنک‌بوست	متنی	۰/۴۶۹			۰/۵۰۴			۰/۵۴۳		
	معنایی	۰/۴۶۳	۱/۴	۰/۱۵۲۴	۰/۴۹۹	-۱/۴	۰/۱۰۷۵	†۰/۵۳۳	-۱/۷	۰/۰۲۲۶
	متنی+ معنایی	†۰/۴۶۸	۳/۶	۰/۰۰۰۱	†۰/۵۱۹	۳/۲	۰/۰۰۰۱	†۰/۵۵۳	۰/۲	۰/۰۰۰۱
آدارنک	متنی	۰/۴۶۸			۰/۵۰۲			۰/۵۴۲		
	معنایی	۰/۴۶۷	-۰/۱	۰/۵۰۴۰	۰/۴۹۸	-۰/۷	۰/۳۹۱۲	۰/۵۳۵	-۱/۳	۰/۱۰۰۷
	متنی+ معنایی	†۰/۴۷۴	۱/۳	۰/۰۰۰۰	†۰/۵۰۵	۰/۶	۰/۰۰۰۰	†۰/۵۴۱	۰	۰/۰۰۰۱
کوردسنت‌اسنت	متنی	۰/۴۷۲			۰/۵۰۴			۰/۵۴۴		
	معنایی	۰/۴۷۲	۰/۱	۰/۵۶۱۶	۰/۵۰۳	-۰/۱	۰/۵۳۲۳	۰/۵۳۹	-۰/۹	۷۸۰۸/۰
	متنی+ معنایی	†۰/۴۸۵	۲/۹	۰/۰۰۰۱	†۰/۵۱۸	۲/۸۰	۰/۰۰۰۱	†۰/۵۵۲	۱/۵	۰/۰۰۰۸
لامبادامارت	متنی	۰/۴۷۸			۰/۵۱۰			۰/۵۴۷		
	معنایی	۰/۴۷۱	-۱/۴	۰/۲۳۳۸	۰/۵۰۳	-۱/۲	۰/۲۷۵۱	۰/۵۳۹	-۱/۵	۰/۰۸۰۷
	متنی+ معنایی	†۰/۴۹۳	۳/۱	۰/۰۰۰۲	†۰/۵۲۷	۳/۳	۰	†۰/۵۵۹	۲/۲	۰/۰۰۰۳
رندوم فارست	متنی	۰/۴۷۲			۰/۵۰۷			۰/۵۴۵		
	معنایی	۰/۴۸۲	۲/۲	۰/۶۵۲۶	۰/۵۱۶	۱/۸	۰/۶۶۲۷	۰/۵۴۸	۰/۶	۰/۶۳۶۹
	متنی+ معنایی	†۰/۴۹۶	۵/۲	۰/۰۰۰۱	†۰/۵۳۰	۴/۵	۰/۰۰۱	†۰/۵۶۱	۱/۳	۰

افزون بر مقایسهٔ میانگین معیارهای ارزیابی به‌ازای کلیهٔ پرس‌وجوها، هر یک از پرس‌وجوها را نیز به‌تنهایی بررسی کرده‌ایم تا میزان بهبود و یا ضعف آن‌ها را نشان دهیم. شکل ۳، تحلیلی از پرس‌وجوهایی را که کارایی آن‌ها در روش استفاده از ترکیب ویژگی‌های متنی و معنایی در مقایسه با حالت پایه بهبود یا ضعف داشته‌اند، ارائه می‌دهد. شکل ۴، نیز همین تحلیل را بر روی استفاده از ویژگی‌های معنایی در مقایسه با حالت پایه نشان می‌دهد. در دو شکل بیان‌شده، محور x، نشان‌دهندهٔ پرس‌وجوهاست که بر اساس درصد اختلاف نسبی NDCG@10، مرتب شده‌اند. پرس‌وجوهایی که در هر دو روش مورد مقایسه دارای NDCG@10های برابر بوده‌اند، حذف شده‌اند. محور y، درصد اختلاف نسبی روش مورد بررسی را در مقایسه با حالت پایه برای NDCG@10 نشان می‌دهد.



شکل ۳. درصد اختلاف نسبی روش استفاده از ترکیب ویژگی‌های متنی و معنایی در مقایسه با حالت پایه به ازای هر پرس‌وجو و برای NDCG@10



شکل ۴. درصد اختلاف نسبی روش استفاده از ویژگی‌های معنایی در مقایسه با حالت پایه به ازای هر پرس‌وجو و برای NDCG@10

همان‌گونه که در جدول ۲، مشخص است، استفاده از ترکیب ویژگی‌های معنایی و متنی برای تمام رتبه‌بندها منجر به بهبود بیش از ۵۰ درصد پرس‌وجوها شده است. در این میان، بیشترین بهبود مربوط به دو رتبه‌بند «مارت» و «رندوم فارست» است، در حالی که کمترین بهبود برای رتبه‌بند «آدارنک» گزارش شده است. اما با استفاده از فقط ویژگی‌های معنایی، تنها دو الگوریتم «مارت» و «رندوم فارست» دارای بهبود بیش از ۵۰ درصد هستند و سایر رتبه‌بندها تضعیف شده‌اند.

جدول ۲. درصد تعداد پرس‌وجوهایی که در دو روش استفاده از ویژگی‌های معنایی و ترکیب آن‌ها با ویژگی‌های متنی در مقایسه با حالت پایه دارای بهبود یا ضعف هستند

	استفاده از ویژگی‌های معنایی + متنی		استفاده از ویژگی‌های معنایی	
رتبه‌بند	درصد بهبود	درصد ضعف	درصد بهبود	درصد ضعف
مارت	۴/۶۱	۶/۳۸	۳/۵۱	۷/۴۸
رنک‌بوست	۷/۵۷	۳/۴۲	۶/۴۷	۴/۵۲
آدارنک	۴/۵۲	۶/۴۷	۶/۴۸	۴/۵۱
کوردینت‌اسنت	۵/۵۶	۵/۴۳	۵/۴۸	۵/۵۱
لامباداامارت	۶/۵۸	۴/۴۱	۷/۴۸	۳/۱۵
رندوم فارست	۴/۶۱	۶/۳۸	۲/۵۲	۸/۴۷

پس از بررسی کل ویژگی‌های معنایی و ترکیب آن‌ها با ویژگی‌های متنی و با استفاده از روش انتخاب ویژگی بیان‌شده در بخش قبل، بهترین ویژگی‌های معنایی را انتخاب کرده و بررسی‌های قبلی را روی این ویژگی‌ها اعمال می‌کنیم. نتایج بررسی کارایی روش‌هایی که از ویژگی‌های معنایی منتخب و همچنین ترکیب آن‌ها با ویژگی‌های متنی استفاده می‌کنند، در مقایسه با روش‌های پایه در جدول ۳، آمده است. در این جدول ویژگی متنی نشان‌دهندهٔ روش پایه است و بهترین ویژگی‌های معنایی بیانگر حالت‌هایی هستند که فقط از ویژگی‌های معنایی منتخب استفاده شده است. همچنین، عبارت کلی بیانگر حالت‌هایی است که از ترکیب ویژگی‌های معنایی منتخب و ویژگی‌های متنی استفاده شده است.

استفاده از فقط ویژگی‌های معنایی منتخب، در دو رتبه‌بند «مارت» و «رندوم فارست» در معیارهای MAP و NDCG@10 باعث بهبود کارایی، ولی برای همهٔ حالت‌های دیگر باعث کاهش کارایی شده است. در مقابل، استفاده از ترکیب ویژگی‌های معنایی منتخب

و ویژگی‌های متنی در مقایسه با حالت پایه، در تمام رتبه‌بندها و برای تمام معیارهای ارزیابی دارای بهبود است و این بهبود دارای اهمیت آماری است. حتی رتبه‌بند «آدارنک» که در حالت قبل از انتخاب ویژگی باعث کاهش NDCG@20 شده بود، در اینجا دارای بهبود است.

جدول ۳. ارزیابی نتایج استفاده از ویژگی‌های معنایی منتخب و استفاده از ترکیب آن‌ها با ویژگی‌های متنی در مقایسه با حالت پایه

رتبه بند	ویژگی	MAP	Δ%	p-value	NDCG @10	Δ%	p-value	NDCG @20	Δ%	p-value
مارت	متنی	۰/۴۶۹			۰/۵۰۳			۰/۵۴۲		
	بهترین ویژگی‌های معنایی	۰/۴۷۱	۰/۶	۰/۶۵۶۱	۰/۵۰۷	۰/۷	۰/۵۱۵۴	۰/۵۴۰	-۰/۳	۰/۷۵۵
	کلی	†۰/۴۹۱	۴/۸	۰/۰۰۱	†۰/۵۲۵	۴/۴	۰/۰۰۱	†۰/۵۵۸	۳/۱	۰/۰۰۰۱
رنک‌بوست	متنی	۰/۴۶۹			۰/۵۰۴			۰/۵۴۳		
	بهترین ویژگی‌های معنایی	۰/۴۶۰	-۱/۸	۰/۱۵۲۴	۰/۴۹۴	-۱/۸	۰/۱۰۷۵	۰/۵۳۱	-۲/۱	۰/۰۲۲۶
	کلی	†۰/۴۸۳	۳/۰	۰/۰۰۰۱	†۰/۵۱۷	۲/۶	۰/۰۰۰۱	†۰/۵۵۲	۱/۸	۰/۰۰۱
آدارنک	متنی	۰/۴۶۴			۰/۵۰۲			۰/۵۴۳		
	بهترین ویژگی‌های معنایی	۰/۴۶۴	-۰/۹	۰/۵۰۴	۰/۴۹۷	-۱/۰	۰/۳۹۱۲	۰/۵۳۴	-۱/۷	۰/۱۰۰۷
	کلی	†۰/۴۸۵	۳/۶	۰/۰۰۱	†۰/۵۱۸	۲/۳	۰/۰۰۰۱	†۰/۵۵۴	۲/۲	۰/۰۰۰۱
کوردینت‌اسنت	متنی	۰/۴۷۲			۰/۵۰۴			۰/۵۴۴		
	بهترین ویژگی‌های معنایی	۰/۴۷۵	۰/۸	۰/۵۶۱۶	۰/۵۰۸	۰/۷	۰/۲۳۵۳	۰/۵۴۲	-۰/۳	۰/۷۸۷
	کلی	†۰/۴۸۸	۴/۳	۰/۰۰۱	†۰/۵۲۰	۳/۲	۰/۰۰۰۱	†۰/۵۵۵	۰/۲	۰/۰۰۰۸
لامبادا مارت	متنی	۰/۴۷۸			۰/۵۱۰			۰/۵۴۷		
	بهترین ویژگی‌های معنایی	۰/۴۷۰	-۱/۷	۰/۲۳۳۸	۰/۵۰۴	-۰/۱	۰/۲۷۵۱	۰/۵۳۹	-۱/۵	۰/۰۸۰۷
	کلی	†۰/۴۹۵	۳/۵	۰/۰۰۰۲	†۰/۵۲۶	۳/۳	۰/۰۰۰۱	†۰/۵۶۰	۲/۳	۰/۰۰۳
رندوم فارست	متنی	۰/۴۷۲			۰/۵۰۷			۰/۵۴۵		
	بهترین ویژگی‌های معنایی	۰/۴۷۴	۰/۵	۰/۶۵۲۶	۰/۵۰۹	۰/۵	۰/۶۶۲۷	۰/۵۴۲	-۰/۴	۰/۶۳۶۹
	کلی	†۰/۴۹۵	۴/۹	۰/۰۰۰۱	†۰/۵۳۰	۴/۴	۰/۰۰۱	†۰/۵۶۱	۰/۳	۰/۰۰۰۱

همانند بررسی‌های قبل از انتخاب ویژگی، هر یک از پرس‌وجوها را نیز به تنهایی بررسی کرده‌ایم تا میزان بهبود و یا ضعف آن‌ها را نشان دهیم.

همان‌گونه که در جدول ۴، مشخص است، استفاده از ترکیب ویژگی‌های معنایی منتخب و ویژگی‌های متنی برای تمام رتبه‌بندها منجر به بهبود حداقل ۷/۵ درصد پرس‌وجوها شده است، در حالی که قبل از انجام انتخاب ویژگی رتبه‌بند «آدارنک» دارای بهبود ۵۲٫۸ درصد بوده است. با استفاده از فقط ویژگی‌های معنایی منتخب، افزون بر دو الگوریتم «مارت» و «رندوم فارست»، که قبل از انتخاب ویژگی هم دارای بهبود بیش از ۵۰ درصد بودند، دو رتبه‌بند «لامبادامارت» و «کوردینت اسنت» نیز دارای بهبود بیش از ۵۰ درصد است. الگوریتم «آدارنک» نیز نسبت به حالت قبل دارای بهبود شده است. اما دو روش «رنک‌بوست» و «رندوم فارست» بدتر از روش قبل از انتخاب ویژگی عمل کرده‌اند.

جدول ٤. درصد تعداد پرس‌وجوهایی که در دو روش استفاده از ویژگی‌های معنایی منتخب و ترکیب آن‌ها با ویژگی‌های متنی در مقایسه با حالت پایه دارای بهبود یا ضعف هستند

	استفاده از ویژگی‌های معنایی منتخب + ویژگی‌های متنی		استفاده از ویژگی‌های معنایی منتخب	
رتبه‌بند	درصد بهبود	درصد ضعف	درصد بهبود	درصد ضعف
مارت	۹/۵۸	۱/۴۱	۶/۵۱	۴/۴۸
رنک‌بوست	۰/۵۷	۰/۴۳	۲/۴۶	۸/۵۳
آدارنک	۲/۵۸	۸/۴۱	۵/۴۹	۵/۵۰
کوردینت اسنت	۵/۵۸	۵/۴۱	۶/۵۰	۴/۴۹
لامبادامارت	۶/۵۷	۴/۴۲	۴/۵۰	۶/۴۹
رندوم فارست	۴/۶۰	۶/۳۹	۰/۵۱	۴۹

افزون بر بررسی ویژگی‌ها به‌صورت دسته‌های معنایی، متنی و ترکیب آن‌ها با یکدیگر، ویژگی‌های موجود در هر یک از چهار دستهٔ ویژگی‌های مبتنی بر گراف و پایگاه دانش، ویژگی‌های مبتنی بر فیلدهای متنی، ویژگی‌های مبتنی بر تکرار موجودیت و ویژگی‌های مبتنی بر نمایش برداری کلمات و متون را مورد بررسی قرار داده‌ایم و دو معیار ارزیابی NDCG@10 و NDCG@20 را برای آن‌ها محاسبه کرده‌ایم.این نتایج در جدول ۵، آمده است.

نتایج بررسی‌ها نشان می‌دهد که ویژگی‌های مبتنی بر تکرار موجودیت در تمام موارد بهتر از سایر ویژگی‌ها عمل می‌کنند. به‌جز رتبه‌بند «رنک‌بوست»، در سایر رتبه‌بندها، ویژگی‌های

مبتنی بر گراف و پایگاه دانش، دومین دستهٔ ویژگی مؤثر هستند.

جدول ۵. نتایج ارزیابی ویژگی‌های گروه‌های مختلف

رتبه‌بند	ویژگی‌ها	NDCG@10	NDCG@20
مارت	مبتنی بر تکرار موجودیت	۰/۴۹۴۴	۰/۵۳۲۴
	مبتنی بر فیلدهای متنی	۰/۴۶۲۳	۰/۵۰۶۱
	مبتنی بر گراف و پایگاه دانش	۰/۴۹۰۱	۰/۵۲۷۶
	مبتنی بر نمایش برداری کلمات و متون	۰/۴۶۳۲	۰/۵۰۷۶
رنک‌بوست	مبتنی بر تکرار موجودیت	۰/۴۸۷۷	۰/۵۲۷۲
	مبتنی بر فیلدهای متنی	۰/۴۵۷۰	۰/۵۰۰۹
	مبتنی بر گراف و پایگاه دانش	۰/۴۵۴۱	۰/۵۰۰۱
	مبتنی بر نمایش برداری کلمات و متون	۰/۴۴۶۰	۰/۴۹۱۸
آدارنک	مبتنی بر تکرار موجودیت	۰/۴۸۷۶	۰/۵۲۶۶
	مبتنی بر فیلدهای متنی	۰/۴۶۵۶	۰/۵۰۸۸
	مبتنی بر گراف و پایگاه دانش	۰/۴۷۰۷	۰/۵۱۲۵
	مبتنی بر نمایش برداری کلمات و متون	۰/۴۳۰۸	۰/۴۸۲۹
کوردینت اسنت	مبتنی بر تکرار موجودیت	۰/۴۹۹۱	۰/۵۳۶۳
	مبتنی بر فیلدهای متنی	۰/۴۷۲۶	۰/۵۱۴۸
	مبتنی بر گراف و پایگاه دانش	۰/۴۷۷	۰/۵۱۸۱
	مبتنی بر نمایش برداری کلمات و متون	۰/۴۴۷۳	۰/۴۹۴۰
لامبادامارت	مبتنی بر تکرار موجودیت	۰/۴۸۷۰	۰/۵۲۵۱
	مبتنی بر فیلدهای متنی	۰/۴۶۲۳	۰/۵۰۵۹
	مبتنی بر گراف و پایگاه دانش	۰/۴۸۴۹	۰/۵۲۴۵
	مبتنی بر نمایش برداری کلمات و متون	۰/۴۶۴۷	۰/۵۰۷۵
رندوم فارست	مبتنی بر تکرار موجودیت	۰/۴۹۴۴	۰/۵۳۲۸
	مبتنی بر فیلدهای متنی	۰/۴۶۷۱	۰/۵۱۰۳
	مبتنی بر گراف و پایگاه دانش	۰/۴۹۱۰	۰/۵۲۲۸
	مبتنی بر نمایش برداری کلمات و متون	۰/۴۶۶۷	۰/۵۱۱

۵. بحث و نتایج

در این تحقیق ویژگی‌های معنایی استخراج شده و سیستم بازیابی با اضافه‌کردن این ویژگی‌ها به ویژگی‌های متنی ارائه شده است. این سیستم نسبت به کارهای گذشته بسیار کاراتر بوده و عملکرد بهتری دارد. علت بهبود نتایج این تحقیق در نظر گرفتن هر دو ویژگی‌های معنایی و متنی است؛ چرا که در کارهای گذشته که از ویژگی‌های معنایی استفاده شده، ویژگی‌های متنی به‌کار برده نشده است. این نادیده‌گرفتن ویژگی‌های متنی در مواقعی که جست‌وجوی کاربر تنها بر اساس کلیدواژه است، مطلوب نخواهد بود. همچنین، سیستم‌های یادگیری رتبه‌بندی پیشین که از ویژگی‌های معنایی استفاده کرده‌اند، هر یک ویژگی‌های معنایی محدودی را بیان نموده‌اند، به‌طوری که تنها جنبهٔ محدودی از معنا پوشش داده شده و سایر جنبه‌های آن در نظر گرفته نشده است.

وجود ابهام در پرس‌وجو موجب بازیابی‌های ضعیف می‌شود. برای برطرف‌کردن این مشکل، موجودیت‌ها را توسط ابزار «تگ‌می» به‌دست آورده‌ایم. از آن جا که «تگ‌می» دارای جزئی برای انجام ابهام‌زدایی است، فرض می‌کنیم که موجودیت‌های برگردانده‌شده، موجودیت‌هایی هستند که ابهام‌زدایی شده‌اند. اما به منظور برطرف‌کردن خطاهای احتمالی، برخی از موجودیت‌های استخراج‌شده از پرس‌وجو را به‌صورت دستی بررسی کرده‌ایم تا از صحت موجودیت‌های برگردانده‌شده اطمینان حاصل کنیم.

با داشتن ویژگی‌های معنایی مبتنی بر گراف و پایگاه دانش به اطلاعاتی از صفحات «ویکی‌پدیا»ی متناظر با هر موجودیت دست یافته‌ایم. اطلاعاتی مانند تشابه صفحاتی که موجودیت‌های سند و پرس‌وجو به آن‌ها لینک داده‌اند و یا تعداد لینک‌های مستقیم و متمایز موجودیت‌های پرس‌وجو به موجودیت‌های سند و همچنین، ویژگی‌های موجود در دستهٔ ویژگی‌های مبتنی بر تکرار موجودیت مانند تعداد رخداد موجودیت‌های پرس‌وجو در سند هر یک از این ویژگی‌ها در بهبود عملکرد سیستم کارا هستند و در هیچ‌یک از ویژگی‌های متنی این اطلاعات وجود ندارد. ویژگی مؤثر دیگر در بهبود عملکرد سیستم ارائه‌شده استخراج ویژگی از فیلدهای متنی موجودیت است. به‌عنوان مثال، در «دی‌بی‌پدیا» خصیصه‌ای به‌نام چکیده[1] داریم که چکیدهٔ طولانی متناظر با موجودیت مورد نظر را ارائه می‌دهد. در این ویژگی برای کلیهٔ موجودیت‌های سند، چکیدهٔ طولانی آن‌ها را استخراج

1. abstract

کرده و امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند را محاسبه کرده‌ایم. در نظر گرفتن این ویژگی‌های معنایی و به‌کار بردن ویژگی متنی در کنار این ویژگی‌ها نتایج قابل قبولی به‌دست داده‌اند.

٦. نتیجه گیری و پیشنهادها

در این مقاله یک سیستم یادگیری رتبه‌بندی ارائه شد که از ویژگی‌های معنایی در کنار ویژگی‌های متنی استفاده می‌کند. ویژگی‌های معنایی در مواقعی کارا هستند که منظور واقعی کاربر در پرس‌وجو مطالبه نشده و دارای ابهام و ایهام بوده است. علت عملکرد بهتر الگوریتم‌های مبتنی بر ویژگی‌های ترکیبی معنایی-متنی نسبت به عملکرد سیستمی که تنها از ویژگی متنی استفاده می‌کند، در نظر گرفتن معنای کلمات موجود در پرس‌وجوی کاربر است. موجودیت‌های استخراج‌شده از مجموعهٔ داده توسط ابزار «تگ‌می» به‌دست آمده‌اند. ویژگی‌های معنایی در چهار دستهٔ ویژگی‌های مبتنی بر تکرار موجودیت، ویژگی‌های مبتنی بر گراف و پایگاه دانش، ویژگی‌های مبتنی بر فیلدهای متنی، و ویژگی‌های مبتنی بر نمایش بردار کلمات و متون قرار می‌گیرند. از بین چهار دسته ویژگی پیاده‌سازی‌شده، دستهٔ ویژگی‌های مبتنی بر تکرار موجودیت بهترین دسته مشاهده شده. برای استخراج این ویژگی‌ها از اطلاعات موجود در گراف «دی‌پی‌پدیا» و صفحات «ویکی‌پدیا» استفاده شده است.

نتایج به‌دست‌آمده در این پژوهش به این صورت است که عملکرد بهترین الگوریتم ویژگی‌های ترکیبی معنایی-متنی ۵۰ درصد نسبت به عملکرد الگوریتم مبتنی بر ویژگی‌های متنی بهبود یافته است. در پیاده‌سازی‌های دیگری که با اعمال الگوریتم‌های انتخاب ویژگی برخی از ویژگی‌های معنایی انتخاب شده‌اند، این الگوریتم بهبود ۷ درصد نسبت به الگوریتم ترکیبی که از تمام ویژگی‌ها استفاده می‌شود، داشته است.

در این پژوهش از ابزار «تگ‌می» برای استخراج موجودیت‌ها استفاده شده است و این ابزار برای بعضی از اسناد و پرس‌وجوها موجودیتی گزارش نکرده است. در کارهای آینده می‌توان در کنار «تگ‌می» از سایر ابزارهای حاشیه‌نویسی متن استفاده کرده و تعداد اسناد و پرس‌وجوهای بیشتری از مجموعهٔ داده را مورد استفاده قرار داد. با توجه به اینکه ویژگی‌های معنایی بر اساس موجودیت‌ها استخراج شده‌اند، هرچه تعداد موجودیت‌های استخراج‌شده از متن و پرس‌وجو بیشتر باشد، امکان استفاده از معنا در بهبود نتایج بیشتر

به‌وجود می‌آید. بنابراین، استفاده از سایر ابزارهای حاشیه‌نویسی متن به‌منظور استخراج موجودیت‌های بیشتر در این زمینه نیز می‌تواند بهبود ایجاد نماید. در این پژوهش تنها از اطلاعات دو پایگاه دانش «ویکی‌پدیا» و «دی‌بی‌پدیا» استفاده شده است. می‌توان علاوه بر این دو پایگاه دانش، از اطلاعات سایر پایگاه‌های دانش به‌منظور تکمیل اطلاعات معنایی مرتبط با هر موجودیت استفاده نمود.

References

Ai, Q. , K. Bi, J. Guo, & W. Bruce Croft.2018 . Learning a deep listwise context model for ranking refinement. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval

Ai, Q., X. Wang, N. Asadi, N. Golbandi, M. Bendersky, & M.-A. Najor. 2019. Learning groupwise multivariate scoring functions using deep neural networks. In *proceedings of the acm sigir international conference on theory of information retrieval*, pp. 85–92. Santa Clara, CA, USA.

Bendersky, M., W. B. Croft, and Y. Diao. 2011. Quality-biased ranking of web documents. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 95-104. Hong Kong, China.

Breiman, L. 2001. Random forests. *Machine learning* 45: 5-32.

Chen, J. , C. Xiong, and J. Callan. 2016. An empirical study of learning to rank for entity search. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 737-740. Pisa, Italy.

Dali, L. , B. Fortuna, T. Duc, and D. Mladenić. 2012. Query-independent learning to rank for rdf entity search. *The semantic web: Research and applications,* pp. 484-498.

Ensan, F. , E. Bagheri, A. Zouaq, and A. Kouznetsov2017 .. An empirical study of embedding features in learning to rank. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 2059-2. New York ,United States.

Ferragina, P. , and U. Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedin of the 19th ACM international conference on Information and knowledge management*, pp. 1625-1628. Toronto, Ontario, Canada.

Freund, Y. , R. Iyer, R. E. Schapire, and Y. Singer. 2003. An efficient boosting algorithm for combining preferences. *Journal of machine learning research* 4: 933-969.

Geng, X., T.-Y. Liu, T. Qin, and H. Li. 2007. Feature selection for ranking. In *Proceedings of the30 th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 407-414. Beijing China.

Guha, R., R. McCool, and E. Miller. 2003. Semantic search," in Proceedings of the 12th international conference on World Wide Web, pp. 700-709. New York, United States.

Han, Z. ,X. Li, M. Yang, H. Qi, and S. Li, 2013. Feature analysis in microblog retrieval based on learning to rank. In *Natural Language Processing and Chinese Computing*, ed: Springer, pp. 410-416.

Hang, H. 2011. A short introduction to learning to rank. *IEICE TRANSACTIONS on Information and Systems* 94: 1854-1862.

Li, H. 2011. Learning to rank for information retrieval and natural language processing. *Synthesis Lectures on Human Language Technologies* 4: 1-113.

Liu, T.-Y. 2009. Learning to rank for information retrieva. *Foundations and Trends in Information Retrieval* 3: 225-331.

_____.2011 . *Learning to rank for information retrieval*. Berlin Heidelberg: Springer Science & Business Media.

Macdonald, C. , B. T. Dinçer, and I. Ounis, 2015. Transferring learning to rank models for web search. In *Proceedings of the 2015 international conference on the theory of information retrieval*, pp. 41-50. New York, United States.

Macdonald, C. , R. L. Santos, and I. Ounis, 2013. The whens and hows of learning to rank for web search. *Information Retrieval* 16: 584-628.

Maio, C.-D. , G. Fenza, M. Gallo, V. Loia, & M. Parente. 2019. Time-aware adaptive tweets ranking through deep learning. *Future Generation Computer Systems* 93: 924-932.

Metzler, D., and W. B. Croft. 2007. Linear feature-based models for information retrieval. *Information Retrieval* 10: 257-274.

Pasumarthi, R.-K. , S. Bruch, X. Wang, C. Li, M. Bendersky, M. Najork, J. Pfeifer, N. Golbandi, R. Anil, & S. Wolf. 2019. Tf-ranking: scalable tensorflow library for learning-to-rank. *In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2970-2978)*.

Phophalia, A. 2011. A survey on learning to rank (letor) approaches in information retrieval ,in Engineering (NUiCONE), 2011 Nirma University International Conference on, pp. 1-6.

Qin, T. , and T.-Y. Liu. 2013. Introducing letor 4.0 datasets. *arXiv preprint arXiv:1306.2597*.

_____, J. Xu, and H. Li. 2010. LETOR: A benchmark collection for research on learning to rank for information retrieval. *Information Retrieval* 13: 346-374.

Schuhmacher, M. , L. Dietz, and S. P Ponzetto. 2015. Ranking entities for web queries through text and knowledg. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1461-1470. New York, United States.

Semeraro, G. 2016. Learning to Rank Entity Relatedness Through Embedding-Based Features. In *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings*, p. 471.

Soldaini, L. , and N. Goharian. 2017. Learning to rank for consumer health search: a semantic approac. In *European Conference on Information Retrieval*, pp. 640-646. Aberdeen, United Kingdom.

Wu, Q., C. J. Burges, K. M. Svore, and J. Gao. 2010. Adapting boosting for information retrieval measures. *Information Retrieval* 13: 254-270. DOI 10.1007/s10791-009-9112-1

Xiong C. , and J. Callan. 2015. Esdrank: Connecting query and documents through external semi-structured data. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 951-960. Melbourne, Australia.

Xu, J., and H. Li, 2007. Adarank: a boosting algorithm for information retrieval. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 391-398. Amsterdam The Netherlands.

Zhao, L. , M. Li, J. Kou, J. Zhang, & Y. Zhang. 2020. A framework for event-oriented text retrieval based on temporal aspects: a recent review. In *Proceedings of the 12th International Conference on Machine Learning and Computing*, pp. 39–46. Shenzhen China.

Zheng, H.-T. , Q. Li, Y. Jiang, S.-T. Xia, and L. Zhang. 2013. Exploiting multiple features for learning to rank in expert finding. In *International Conference on Advanced Data Mining and Applications*, pp. 219-230. Hangzhou, China.

جدول (الف). مقایسهٔ کارهای گذشته

هدف مقاله	الگوریتم یادگیری ماشین استفاده‌شده	نوع الگوریتم استفاده‌شده	ویژگی‌های مورد استفاده					
			آیا از دانش استفاده می‌کند؟	ویژگی مبتنی بر گراف و پایگاه دانش	ویژگی مبتنی بر تکرار موجودیت	ویژگی مبتنی بر فیلدهای متنی	ویژگی مبتنی بر نمایش برداری کلمات و متون	ویژگی مبتنی بر شبکه‌های اجتماعی	استفاده از ویژگی متنی	استفاده از ویژگی معنایی
رتبه‌بندی موجودیت‌های RDF	RankSVM	Pairwise	بله، استفاده از Dbpedia و Yago	*	*					*
رتبه‌بندی موجودیت	Lambda MART	pairwise/ listwise	بله، استفاده از ویکی‌پدیا و Dbpedia	*		*	*		*	*
رتبه‌بندی اسناد	ListMLE و RankSVM	Listwise و pairwise	بله، استفاده از پایگاه دانش Freebase	*	*	*			*	*
استخراج microblog	RankSVM	Pairwise	بله، استفاده از اطلاعات شبکه‌های اجتماعی			*		*	*	ویژگی شبه معنایی
استخراج خبره در زمینه‌ای خاص	RankSVM	Pairwise	خیر			*	*		*	
محاسبه شباهت بین دو موجودیت	–	–	بله، استفاده از پایگاه دانش Dbpedia	*						*

هدف مقاله	الگوریتم یادگیری ماشین استفاده‌شده	نوع الگوریتم استفاده‌شده	ویژگی‌های مورد استفاده					
			آیا از دانش استفاده می‌کند؟	ویژگی مبتنی بر گراف و پایگاه دانش	ویژگی مبتنی بر تکرار موجودیت	ویژگی مبتنی بر فیلدهای متنی	ویژگی مبتنی بر نمایش برداری کلمات و متون	ویژگی مبتنی بر شبکه‌های اجتماعی	استفاده از ویژگی متنی	استفاده از ویژگی معنایی
پیدا کردن موجودیت‌های مرتبط با پرس‌وجو به‌منظور استفاده در گسترش پرس‌وجو	–	–	بله، استفاده از پایگاه‌های دانش ویکی‌پدیا و DBpedia				*		*	*
محاسبه فاصله معنایی بین داده‌های پیوندی با هدف پیشنهاد منبع	–	–	بله، * استفاده از گراف داده‌های پیوندی							*
محاسبه شباهت منابع با هدف استفاده در سیستم‌های پیشنهاددهنده	–	–	بله، * استفاده از DBpedia							*
محاسبه شباهت منابع با هدف استفاده در سیستم‌های پیشنهاددهنده	–	–	بله، * استفاده از DBpedia							*
لینک موجودیت‌ها	–	–	بله، * استفاده از ویکی‌پدیا و وردنت							*
محاسبه‌ی ارتباط معنایی	–	–	بله، * استفاده از ویکی‌پدیا							*

هدف مقاله	الگوریتم یادگیری ماشین استفاده‌شده	نوع الگوریتم استفاده‌شده	ویژگی‌های مورد استفاده						
			آیا از دانش استفاده می‌کند؟	ویژگی مبتنی بر گراف و پایگاه دانش	ویژگی مبتنی بر تکرار موجودیت	ویژگی مبتنی بر فیلدهای متنی	ویژگی مبتنی بر نمایش برداری کلمات و متون	ویژگی مبتنی بر شبکه‌های اجتماعی	استفاده از ویژگی متنی	استفاده از ویژگی معنایی
بهبود نتایج جست‌وجوی پزشکی کاربران	Logistic-regression, random-forests, LambdaMART, AdaRank, ListNET	Pointwise Pointwise Pairwise Listwise listwise	بله، از پایگاه دانش اطلاعات پزشکی و صفحات ویکی‌پدیا	*	*		*		*	*
رتبه‌بندی موجودیت‌ها	RankSVM coordinate ascent	Pairwise listwise	بله، استفاده از ویکی‌پدیا و DBpedia	*	*	*				*
رتبه‌بندی اسناد	AdaRank ListNet RankNet RankBoost	Listwise Listwise Pairwise pairwise	بله، استفاده از ویکی‌پدیا				*		*	*
رتبه‌بندی موجودیت	RankSVM Coordinate-ascent	Pairwise listwise	بله، استفاده از DBpedia	*						*

جدول (ب). ویژگی‌های تعریف‌شده

شماره ستون ویژگی در خروجی	توصیف ویژگی	دسته‌بندی بر اساس مکانیزم استخراج	دسته‌بندی بر اساس وابستگی به پرس‌وجو و سند
۴۷	تعداد رخداد موجودیت‌های پرس‌وجو در سند $\sum_{e_i \in qe \cap de} c(e_i, de)$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۴۸	تعداد رخداد موجودیت‌های پرس‌وجو در سند با استفاده از لگاریتم $\sum_{e_i \in qe \cap de} \log(c(e_i, de)+1)$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۴۹	نرمال‌شدهٔ ویژگی ۴۷ $\sum_{e_i \in qe \cap de} \frac{c(e_i, de)}{\lvert de \rvert}$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۵۰	نرمال‌شدهٔ ویژگی ۴۸ $\sum_{e_i \in qe \cap de} \log(\frac{c(e_i, de)}{\lvert de \rvert}+1)$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۵۱	رابطهٔ کل اسناد مجموعه و تعداد اسنادی که موجودیت مورد نظر ما در آن‌ها ظاهر شده است $\sum_{e_i \in qe \cap de} \log \frac{\lvert C \rvert}{df(e_i)}$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۵۲	حاصل‌ضرب دو ویژگی ۴۷ و ۵۱	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۵۳	تعداد موجودیت‌های سند	مبتنی بر تکرار موجودیت	سند
۵۴	ویژگی ۵۱ برای موجودیت‌های پرس‌وجو $\sum_{e_i \in qe} \log \frac{\lvert C \rvert}{df(e_i)}$	مبتنی بر تکرار موجودیت	پرس‌وجو
۵۵	مشابه ویژگی ۵۴ با استفاده از لگاریتم $\sum_{e_i \in qe} \log(\log \frac{\lvert C \rvert}{df(e_i)})$	مبتنی بر تکرار موجودیت	پرس‌وجو
۵۶	$\sum_{e_i \in qe} \log(\frac{\lvert C \rvert}{c(e_i, C)}+1)$	مبتنی بر تکرار موجودیت	پرس‌وجو
۵۷	$\sum_{e_i \in qe \cap de} \log(\frac{c(e_i, de)}{\lvert de \rvert}) \log(\frac{\lvert C \rvert}{df(e_i)}+1)$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۵۸	$\sum_{e_i \in qe \cap de} \log((\frac{c(e_i, de)}{\lvert de \rvert} \cdot \frac{\lvert C \rvert}{c(e_i, C)})+1)$	مبتنی بر تکرار موجودیت	پرس‌وجو - سند

شماره ستون ویژگی در خروجی	توصیف ویژگی	دسته‌بندی بر اساس مکانیزم استخراج	دسته‌بندی بر اساس وابستگی به پرس‌وجو و سند
۵۹	امتیاز BM25 رخداد موجودیت‌های پرس‌وجو در موجودیت‌های سند	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۶۰	لگاریتم امتیاز BM25 رخداد موجودیت‌های پرس‌وجو در موجودیت‌های سند	مبتنی بر تکرار موجودیت	پرس‌وجو - سند
۶۱ تا ۷۴	ویژگی‌های ۴۷ تا ۶۰ برای ترکیب موجودیت‌های سند و پرس‌وجو با موجودیت‌های مرتبط با آنها	مبتنی بر تکرار موجودیت	مشابه ویژگی‌های ۴۷ تا ۶۰
۷۵	کمینه تشابه Category بین موجودیت‌های سند و پرس‌وجو	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۷۶	بیشینه تشابه Category بین موجودیت‌های سند و پرس‌وجو	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۷۷	میانگین تشابه Category بین موجودیت‌های سند و پرس‌وجو	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۷۸	مینیمم تشابه صفحاتی که موجودیت‌های سند و پرس‌وجو به آن‌ها لینک داده‌اند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۷۹	ماکزیمم تشابه صفحاتی که موجودیت‌های سند و پرس‌وجو به آن‌ها لینک داده‌اند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۸۰	میانگین تشابه صفحاتی که موجودیت‌های سند و پرس‌وجو به آن‌ها لینک داده‌اند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۸۱	مینیمم شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تعبیهٔ کلمه	مبتنی نمایش برداری کلمات و متون	پرس‌وجو-سند
۸۲	ماکزیمم شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تعبیهٔ کلمه	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند
۸۳	میانگین شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تعبیهٔ کلمه	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند
۸۴	مینیمم تعداد لینک‌های مستقیم و متمایز موجودیت‌های پرس‌وجو به موجودیت‌های سند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۸۵	ماکزیمم تعداد لینک‌های مستقیم و متمایز موجودیت‌های پرس‌وجو به موجودیت‌های سند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۸۶	میانگین تعداد لینک‌های مستقیم و متمایز موجودیت‌های پرس‌وجو به موجودیت‌های سند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند

شماره ستون ویژگی در خروجی	توصیف ویژگی	دسته‌بندی بر اساس مکانیزم استخراج	دسته‌بندی بر اساس وابستگی به پرس‌وجو و سند
۸۷	مینیمم تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند، نقش فاعل را دارد	مبتنی بر گراف و پایگاه دانش	سند
۸۸	ماکزیمم تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند، نقش فاعل را دارد	مبتنی بر گراف و پایگاه دانش	سند
۸۹	میانگین تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند، نقش فاعل را دارد	مبتنی بر گراف و پایگاه دانش	سند
۹۰	مینیمم تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند، نقش مفعول را دارد	مبتنی بر گراف و پایگاه دانش	سند
۹۱	ماکزیمم تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند، نقش مفعول را دارد	مبتنی بر گراف و پایگاه دانش	سند
۹۲	میانگین تعداد سه‌گانه‌هایی که در آن‌ها نود متناظر با موجودیت سند، نقش مفعول را دارد	مبتنی بر گراف و پایگاه دانش	سند
۹۳	مینیمم تعداد انواع مسندهای خروجی از نود متناظر با موجودیت سند	مبتنی بر گراف و پایگاه دانش	سند
۹۴	ماکزیمم تعداد انواع مسندهای خروجی از نود متناظر با موجودیت سند	مبتنی بر گراف و پایگاه دانش	سند
۹۵	میانگین تعداد انواع مسندهای خروجی از نود متناظر با موجودیت سند	مبتنی بر گراف و پایگاه دانش	سند
۹۶	مینیمم تعداد انواع مسندهای ورودی به نود متناظر با موجودیت سند	مبتنی بر گراف و پایگاه دانش	سند
۹۷	ماکزیمم تعداد انواع مسندهای ورودی به نود متناظر با موجودیت سند	مبتنی بر گراف و پایگاه دانش	سند
۹۸	میانگین تعداد انواع مسندهای ورودی به نود متناظر با موجودیت سند	مبتنی بر گراف و پایگاه دانش	سند
۹۹	مینیمم شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تحلیل معنایی صریح	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند
۱۰۰	ماکزیمم شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تحلیل معنایی صریح	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند
۱۰۱	میانگین شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تحلیل معنایی صریح	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند
۱۰۲	مینیمم امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند

شماره ستون ویژگی در خروجی	توصیف ویژگی	دسته‌بندی بر اساس مکانیزم استخراج	دسته‌بندی بر اساس وابستگی به پرس‌وجو و سند
۱۰۳	ماکزیمم امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۰۴	میانگین امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۰۵	مینیمم امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۰۶	ماکزیمم امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۰۷	میانگین امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ طولانی موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۰۸	مینیمم امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۰۹	ماکزیمم امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۱۰	میانگین امتیاز BM25 رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۱۱	مینیمم امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۱۲	ماکزیمم امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۱۳	میانگین امتیاز TF-IDF رخداد کلمات پرس‌وجو در چکیدهٔ کوتاه موجودیت‌های سند	مبتنی بر فیلدهای متنی	پرس‌وجو-سند
۱۱۴	امتیاز WMD تشابه Category بین موجودیت‌های سند و پرس‌وجو	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۱۱۵	امتیاز WMD تشابه صفحاتی که موجودیت‌های سند و پرس‌وجو به آها لینک داده‌اند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۱۱۶	امتیاز WMD شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تعبیهٔ کلمه	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند
۱۱۷	امتیاز WMD تعداد لینک‌های مستقیم و متمایز موجودیت‌های پرس‌وجو به موجودیت‌های سند	مبتنی بر گراف و پایگاه دانش	پرس‌وجو-سند
۱۱۸	امتیاز WMD مینیمم شباهت کسینوسی بین زوج بردارهای متناظر با موجودیت‌های سند و پرس‌وجوی استخراج‌شده با استفاده از تحلیل معنایی صریح	مبتنی بر نمایش برداری کلمات و متون	پرس‌وجو-سند

محدثه محجوب

دارای مدرک کارشناسی ارشد مهندسی کامپیوتر از دانشگاه فردوسی مشهد است.

بازیابی اطلاعات از جمله علایق پژوهشی وی است.

فائزه انسان

متولد سال ۱۳۶۰، دارای مدرک تحصیلی دکتری در رشتهٔ علوم کامپیوتر از دانشگاه نیوبرانزویک، کاناداست. ایشان هم‌اکنون استادیار گروه مهندسی کامپیوتر دانشگاه فردوسی مشهد است.

بازیابی اطلاعات منطق و استنتاج از جمله علایق پژوهشی وی است.

ساناز کشوری

متولد ۱۳۷۱، دانشجوی دکتری در رشتهٔ مهندسی نرم‌افزار دانشگاه فردوسی مشهد است.

بازیابی اطلاعات وب و شبکه‌های عصبی از جمله علایق پژوهشی وی است.

پرستو جعفرزاده

متولد سال ۱۳۷۴، دارای مدرک تحصیلی کارشناسی ارشد در رشتهٔ مهندسی کامپیوتر از دانشگاه فردوشی مشهد است. وی هم‌اکنون دانشجوی دکتری در رشتهٔ مهندسی کامپیوتر در دانشگاه رایرسون کاناداست.

بازیابی اطلاعات از جمله علایق پژوهشی وی است.

محمدامین کیوان‌زاد

دارای مدرک کارشناسی ارشد مهندسی کامپیوتر از دانشگاه نیوبرانزویک کاناداست.

بازیابی اطلاعات از جمله علایق پژوهشی وی است.