

A new Approach to Perform Fuzzy Queries over NoSQL Database

Zahra Ahmadi

Msc in Computer Science;
Department of Mathematics and Computer Science;
Allameh Tabataba'i University Email: z_ahmadi@atu.ac.ir

Farzam Matinfar*

PhD in Computer Engineering; Assistant Professor;
Department of Mathematics and Computer Science;
Allameh Tabataba'i University Email: f.matinfar@atu.ac.ir

Fereshteh Azadi Parand

PhD in Computer Engineering; Assistant Professor;
Department of Mathematics and Computer Science;
Allameh Tabataba'i University Email: parand@atu.ac.ir

Iranian Journal of
Information
Processing and
Management

Received: 26, Feb. 2019 Accepted: 24, Nov. 2019

Abstract: Today, due to massive amount of information, NoSQL databases are used to compute complex data. These databases are used to store semi-structured and unstructured data for big data management. In this paper, fuzzy queries are executed by users on the information stored in the Neo4j graph database, which show results in defuzzy manner. To evaluate the efficiency of the proposed method, we considered a database of manufacturing company. Price and quality fuzzy variables were considered and used to define and execute the fuzzy queries. The results can verify the performance of the proposed method. Additionally, a counselor (who has information about the data) defines fuzzy terms, membership functions, and fuzzy rules table. The most important process in this method is to find the center of gravity in order to defuzzify the final result. So, an algorithm has been implemented for this purpose by C # programming language. The greatest query time is due to find center of gravity. The evaluation results show that the increase in query time by using the proposed approach in comparison to query time by using Cypher language is acceptable due to the complexity of fuzzy concepts. Therefore, the proposed solution will be suitable for using ambiguous and fuzzy queries in large databases.

Keywords: NoSQL, Neo4j Graph Database, Fuzzy Query, Defuzzification, Center of Gravity

Iranian Research Institute
for Information Science and Technology
(IranDoc)

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 35 | No. 2 | 463-484

Winter 2020



* Corresponding Author

راهکاری جدید برای انجام پرس و جوهای فازی بر روی پایگاه داده‌های گرافی NoSQL

زهره احمدی

دانشجوی کارشناسی ارشد؛ گروه ریاضی و علوم رایانه؛
دانشگاه علامه طباطبائی | z_ahmadi@atu.ac.ir

فرزام متین‌فر

دکتری کامپیوتر؛ استادیار؛ گروه ریاضی و علوم رایانه؛
دانشگاه علامه طباطبائی؛
پدیدآور رابط | f.matinfar@atu.ac.ir

فرشته آزادی پوند

دکتری کامپیوتر؛ استادیار؛ گروه ریاضی و علوم رایانه؛
دانشگاه علامه طباطبائی | parand@atu.ac.ir



دریافت: ۱۳۹۷/۱۲/۰۷ | پذیرش: ۱۳۹۸/۰۹/۰۳ | مقاله برای اصلاح به مدت ۷ روز نزد پدیدآوران بوده است.

چکیده: امروزه به دلیل حجم انبوه اطلاعات از پایگاه داده‌های NoSQL برای محاسبه داده‌های پیچیده استفاده می‌شود. این پایگاه داده‌ها با قابلیت ذخیره داده‌های نیمه‌ساختار یافته و بدون ساختار برای مدیریت داده‌های بزرگ مورد استفاده قرار می‌گیرد. در این مقاله پرس و جوهای فازی توسط کاربران بر روی اطلاعات ذخیره شده در پایگاه داده گراف Neo4j انجام شده است که منجر به نمایش نتایج به صورت غیر فازی می‌شود. برای ارزیابی کارایی روش پیشنهادی با در نظر گرفتن یک پایگاه داده از داده‌های مربوط به یک شرکت قطعه‌سازی، دو متغیر قیمت و کیفیت با ویژگی فازی در نظر گرفته شد و از آن‌ها برای تعریف و اجرای پرسش‌های فازی استفاده شد که نتایج می‌تواند صحت عملکرد روش پیشنهادی را تأیید کند. علاوه بر این، یک مشاور که دارای اطلاعات در مورد داده‌هاست، اصطلاح‌های فازی، توابع عضویت، و جدول قوانین فازی را تعریف می‌کند. یکی از فرایندهای مهم در این روش یافتن مرکز ثقل به منظور غیر فازی‌سازی نتیجه نهایی است که برای این منظور الگوریتمی با زبان برنامه‌نویسی C# پیاده‌سازی شده است. بیشترین زمان پرس و جو مربوط به یافتن مرکز ثقل است و پیچیدگی مسائل فازی سربرار زمانی بیشتری تحمیل نخواهد کرد. نتایج ارزیابی نشان می‌دهد که افزایش زمان پرس و جو با استفاده از رویکرد پیشنهادی مبتنی بر منطق فازی نسبت به زمان پرس و جو با زبان

نشریه علمی | رتبه بین‌المللی

پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۲۲۳۱-۲۲۵۱

شاپا (الکترونیکی) ۸۲۳۱-۸۲۵۱

نمایه در SCOPUS و ISI، LISTA و

jipm.irandoc.ac.ir

دوره ۳۵ | شماره ۲ | صص ۴۶۳-۴۸۴

زمستان ۱۳۹۸



Cypher، با در نظر گرفتن پیچیدگی بیشتر مفاهیم فازی قابل قبول است. در نتیجه، راهکار پیشنهادی گزینه‌ای مناسب برای بهره‌گیری پرس‌وجوهای مبهم درون پایگاه داده‌های بزرگ خواهد بود.

کلیدواژه‌ها: پایگاه داده گراف Neo4j، NoSQL، پرس‌وجوهای فازی، غیرفازی‌سازی، روش مرکز ثقل

۱. مقدمه

پیشرفت فناوری، گسترش اینترنت و استفاده از خدمات الکترونیکی منجر به تولید حجم انبوهی از داده‌ها در زمینه‌های گوناگون مانند پزشکی، علمی، تجارت، نجوم و غیره شده است (Laney 2001). روش‌های رابطه‌ای به‌طوری مؤثر توانایی ما را در تجزیه و تحلیل داده‌های عظیم در مورد حجم، سرعت و یا نمایش داده‌های بزرگ محدود می‌کند (Manyika et al. 2011). هنگامی که به عمل پیوند^۱ پایگاه داده‌های بزرگ نیاز داشته باشیم، معماری پایگاه داده‌های رابطه‌ای^۲ کارآمد نیستند. از این رو، مدل‌های جایگزین جدید بسیاری برای ذخیره و مدیریت این داده‌های بزرگ معرفی شده‌اند (Vicknair 2010). پایگاه داده‌های NoSQL برای ذخیره انواع متفاوتی از داده‌های مختلف معرفی شده‌اند که در آن‌ها هیچ محدودیتی در ساختار و نرمال‌سازی وجود ندارد. علاوه بر این، تفاوت عمده دیگر آن‌ها با پایگاه داده‌های رابطه‌ای، فقدان SQL به‌عنوان یک زبان پرس‌وجوست. چهار نوع ذخیره‌سازی برای داده‌های بزرگ در پایگاه داده NoSQL وجود دارد: مبتنی بر ستون^۳، مبتنی بر سند^۴، کلید-مقدار^۵ و مبتنی بر گراف^۶ (Fowler 2015).

زمانی که در پایگاه داده‌ها پرس‌وجو انجام می‌شود، کاربران ممکن است تقاضای پرسش‌های فازی و مبهم داشته باشند (Castellort and Laurent 2014). از حوزه‌های کاربرد مباحث فازی می‌توان به گزارش‌گیری با استفاده از پرس‌وجوهای فازی مطرح‌شده و بیان نتایج به‌صورت عبارات فازی برای سطوح مدیریتی اشاره کرد که منجر به درک بهتر مفاهیم و بهبود کیفیت ارتباطات می‌شود. علاوه بر این، نرم‌افزار اجرایی^۷ و سیستم‌های پیشنهاددهنده^۸ می‌توانند از مفاهیم فازی و پرسش‌ها استفاده کنند. این موضوع در مورد

1. join

2. relational database

3. column-based

4. document-based

5. key-value

6. graph-based

7. administrative software application

8. recommender systems

پایگاه داده‌های رابطه‌ای و NoSQL مطالعه شده است. با این حال، هیچ چارچوبی به منظور ارائه نتایج به صورت غیرفازی در پایگاه داده‌های گرافی مطرح نشده است. یکی از رویکردهای توسعه یافته برای پذیرش پرسش‌های فازی به عنوان ورودی در پژوهش «کاستلتورت و مارتین» پیشنهاد شده است. آن‌ها زبان فازی Cypher را با گسترش زبان Cypher برای تعریف واژگان فازی بر روی پایگاه داده‌های گراف NoSQL معرفی کردند (Castelltort and Martin 2018). با استفاده از این زبان، کاربران می‌توانند اصطلاح‌های فازی خود را در یک فایل تعریف کرده و سپس، آن‌ها را به عنوان ورودی‌های یک تابع از پیش تعریف شده استفاده کنند. با این همه، کاربران ترجیح می‌دهند نتایج دقیق و شفاف را به عنوان پاسخ نهایی به دست آورند تا تصمیمات خود را به صورت مؤثر انجام دهند. از این رو، روش‌های غیرفازی‌سازی برای تبدیل نتایج فازی به داده‌های دقیق نیاز است و ما این هدف را برای رویکرد پیشنهادی در نظر می‌گیریم. در روش پیشنهادی در این مقاله، جواب فازی خروجی با استفاده از تکنیک‌های غیرفازی‌سازی شناخته شده به یک جواب دقیق تبدیل شده و به کاربر ارائه می‌گردد.

ادامه مقاله به این صورت ساختار بندی شده است: در بخش دوم، تعاریف و مفاهیم اولیه مرتبط با مسئله و همچنین کارهای مرتبط انجام شده در این حوزه را ارائه می‌شود. بخش سوم، پیرامون روش پیشنهادی در این مقاله صحبت می‌کند. بخش چهارم، به تجزیه و تحلیل یافته‌ها اختصاص یافته و در انتها، نتیجه گیری و کارهای آتی در بخش پنجم تشریح شده است.

۲. پیشینه پژوهش

در ابتدای این بخش به معرفی مختصر مفاهیم استفاده شده در این مقاله پرداخته شده و سپس، کارهای مرتبط با این پژوهش در پایگاه داده‌های رابطه‌ای و گرافی بررسی شده است.

۲-۱. مفاهیم اولیه و ویژگی‌ها

در این بخش به بیان مفاهیم و تعاریفی که در ساختار این مقاله مطرح شده، پرداخته می‌شود. همان‌طور که در مقدمه گفته شد، پایگاه داده‌های NoSQL اغلب برای ذخیره

داده‌های بزرگ^۱ بوده و بر اساس نوع ذخیره‌سازی و ارتباط داده‌ها به چهار دسته تقسیم می‌شود: کلید-مقدار^۲، مبتنی بر ستون، مبتنی بر سند، و مبتنی بر گراف. یکی از کاربردی‌ترین دسته‌ها، پایگاه داده گراف است که در زیر به اختصار معرفی می‌شود.

۱-۱-۲. پایگاه داده گراف

پایگاه داده گراف، پایگاه داده‌ای است که از ساختار گراف با گره‌ها، یال‌ها و خواص برای نمایش و ذخیره داده‌ها استفاده می‌کند. اغلب پایگاه داده‌های گرافی اطلاعاتشان را به صورت کلید-مقدار یا سندمحور ذخیره می‌کنند؛ کاربرد این دسته از پایگاه داده‌ها برای زمانی است که ارتباطاتی معین و مشخص میان رکوردها -مانند شبکه‌های اجتماعی- وجود دارد. روابط در پایگاه داده گرافی اجازه می‌دهد که سلسله‌مراتب‌های پیچیده به سرعت طی شود و در نتیجه، سرعت دسترسی و بهیگی سیستم افزایش یابد (Angles and Gutierrez 2008). چند نمونه از پایگاه داده‌های ذخیره‌سازی این دسته Neo4J, Sparksee, InfoGrid, AllegroGraph هستند.

۲-۱-۲. پایگاه داده Neo4j

پایگاه داده گراف Neo4j یکی از انواع پایگاه داده‌های گراف است که برای ذخیره‌سازی، مدیریت و گزارش‌گیری‌های بسیار پیچیده و کاربردی استفاده می‌شود. از این پایگاه داده به منظور پیشنهادها لحظه‌ای^۳ (همانند سرویس‌های پیشنهاددهی اینستاگرام و فیسبوک و ...)، تشخیص تقلب در کلاه‌برداری‌ها، تشخیص هویت و بررسی دسترسی‌ها، شبکه‌ها و عملیات‌های فناوری اطلاعات استفاده می‌شود (Smets 1997; Neo4j Inc.'s website). سیستم Neo4J زبان رسمی Cypher را ارائه می‌دهد که به کاربر اجازه می‌دهد با استفاده از یک شیوه بسیار بصری، گراف را فراخوانی کند.

۳-۱-۲. مفاهیم منطق فازی

تئوری فازی چارچوبی با توانایی مدل کردن واقعیت است. متغیرهای زبانی در سیستم‌های فازی برای توصیف متغیرهای موجود در سیستم مثل دما، فشار، دوز دارو و غیره به کار می‌روند (Zimmermann 1985). یکی از پارامترهای بسیار مهم در سیستم‌های

1. big data

2. Key-value

3. instannt suggestions

فازی، توابع عضویت^۱ هستند که متغیرها را مشخص می‌کنند و می‌توانند به صورت گسسته و یا پیوسته تعریف شوند. روش‌های مختلفی برای تعیین توابع عضویت وجود دارند که همگی آن‌ها مبتنی بر ذهن و تجربه افراد است و در هر حوزه‌ای معمولاً توسط متخصصان آن حوزه تعیین می‌شوند. یک سیستم فازی حداقل دارای سه مرحله زیر است (Kruse, Gebhardt and Klowon 1994):

◇ فازی‌سازی: وظیفه این مرحله تبدیل یک عدد غیرفازی به یک مقدار فازی است. به این فرایند فازی‌سازی می‌گویند.

◇ تعریف پایگاه قواعد و موتور استنتاج: قواعد سیستم فازی به نحوی ایجاد می‌گردند که کلیه ترکیبات ممکن از حالت‌های مختلف ورودی را شامل شده و مشخص کنند که در هر حالت، خروجی چه باید باشد. قواعد فازی در موتور استنتاج فازی مورد استفاده قرار می‌گیرند تا پاسخ مناسب ایجاد گردد. این موتور میزان تطابق ورودی سیستم را با هر قاعده معین می‌کند. پس از استنتاج در صورتی که چندین عمل فازی‌سازی انجام شده باشد، برای به دست آوردن نتیجه نهایی نیاز به ترکیب نتایج حاصل از قواعد داریم که به این فرایند، ترکیب می‌گویند. برای ترکیب از دو روش Max-Min و Max-Product استفاده می‌شود.

◇ غیرفازی‌سازی: در غیرفازی‌سازی، مجموعه فازی به دست آمده توسط موتور استنتاج به یک مقدار عددی تبدیل می‌شود (Saade and Diab 2004). روش‌های زیر شناخته شده‌ترین روش‌ها برای غیرفازی‌سازی هستند: روش مرکز جمع‌ها (COS)^۲، روش مرکز ثقل (COG)^۳، روش مرکز مساحت (COA)^۴، روش مساحت بیسکتور (BOA)^۵، روش میانگین وزنی (WA)^۶، روش ماکزیمم (اولین (FOM)^۷، آخرین (LOM)^۸، میانی (MOM)^۹).

۲-۲. کارهای مرتبط

رویکردهای بسیاری به منظور انجام پرس‌وجوهای فازی با توجه به داده‌های نامعلوم و مبهم مطرح شده است. این مقاله با هدف پیاده‌سازی و اجرای پرسش‌های فازی به روی داده‌های معلوم و دقیق بر مبنای منطق فازی ارائه شده است (Smets 1997). برخی

1. membership functions
4. center of area
7. first of maxima method

2. center of sum
5. bisector of area
8. last of maxima method

3. center of gravity
6. weighted average
9. mean of maxima method

پژوهش‌ها در زمینه توسعه زبان SQL و ایجاد دو زبان پرس‌وجوی فازی FSQL/ SQLf و FQL در پایگاه داده‌های رابطه‌ای است (Bosc and Pivert 1995 و Takahashi 1991). این زبان‌ها، پرسش‌های فازی را بر روی برچسب‌های فازی، فاکتورهای مقایسه و جمع‌بندی انجام می‌دهند. علاوه بر این، چندین موتور پایگاه داده فازی با ویژگی‌های فازی ارائه شده‌اند که در آن دامنه‌های فازی توسط کاربران تعریف و بعد از آن پرس‌وجوها انجام می‌شوند (Galindo et al. 1998; Zadrożny and Kacprzyk 1998). همچنین، f-SPARQL یک برنامه افزودنی قابل انعطاف برای SPARQL را فراهم می‌کند که به کاربران اجازه می‌دهد فیلترهای فازی و اپراتورهای فازی را اعمال کنند (Cheng, Ma & Yan 2010).

اولین کار برای تعریف پرسش‌های فازی در پایگاه داده‌های گراف NoSQL در سال ۲۰۱۳ توسط «آرنود کاستلتورت» ارائه شد (Castelltort and Laurent 2014). «کاستلتورت» چندین رویکرد برای پیاده‌سازی و اعمال پرس‌وجوهای فازی در پایگاه داده گرافی Neo4j توسط زبان پرس‌وجوی Cypher را به صورت زیر ارائه داد (Castelltort and Martin 2018):

- ◇ ایجاد یک زبان همپوشانی به روی زبان Cypher؛
- ◇ توسعه پرس‌وجوهای Cypher و استفاده از API^۱ سطح پایین موجود؛
- ◇ توسعه API سطح پایین با توابع بهینه فقط در دسترس کاربران پیشرفته با مهارت‌های توسعه؛
- ◇ استفاده از یک زبان پرس‌وجوی توسعه Cypher در یک API سطح پایین پیشرفته.

استفاده از یک زبان پرس‌وجوی توسعه Cypher در یک API سطح پایین پیشرفته بهترین رویکرد است، اما نیاز به هزینه‌های سنگین توسعه و نگهداری دارد. «کاستلتورت» برای این راه حل دو پیاده‌سازی ارائه داد: پیاده‌سازی اول، پیاده‌سازی ساده و بی‌تکلف است که مجموعه‌ای از توابع فازی را برای کاربران تعریف می‌کند و پیاده‌سازی دیگر، زبان پرس‌وجوی Cypher را معرفی می‌کند. برای ارائه یک پیاده‌سازی کلی لازم بود که یک زبان خاص دامنه اضافه شود تا کاربران بتوانند در نهایت، عبارات فازی را با DSL فازی بیان کنند (همان).

علاوه بر این، عبارات کمی فازی نیز می‌توانند در پرسش‌های فازی مورد استفاده قرار گیرند. استفاده از نمایش داده‌های کمی فازی در پایگاه داده‌های شبکه اجتماعی بدون

1. application program interface

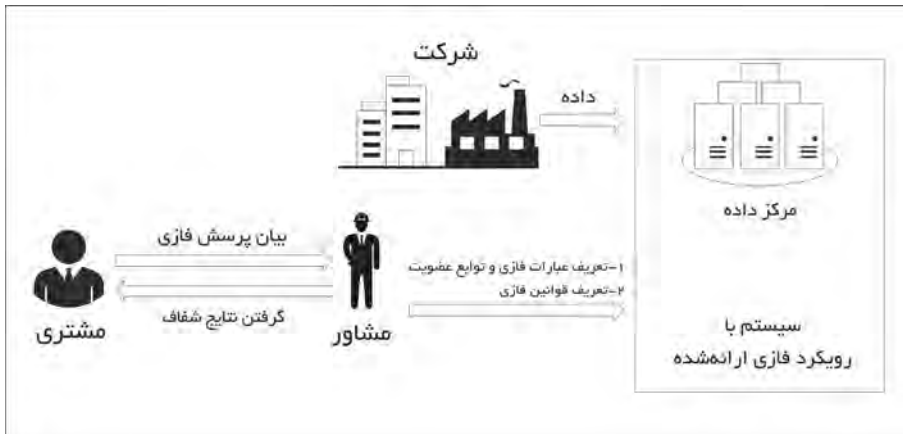
ارائه زبانی برای پرس‌وجو در (2014) Yager بحث شده است. در سال ۲۰۱۹، «پیورت» و همکارانش رویکردی برای بیان نمایش داده‌های کمی با تعریف دستوره‌های زبانی، اضافه بر زبان پرس‌وجوی Cypher در پایگاه داده‌های گرافی فازی ارائه دادند (Pivert, Slama & Thion 2019).

به دلیل این که پیاده‌سازی DSL نیازمند سطح بالای دانش برنامه‌نویسی است و نتایج فازی از پیاده‌سازی دوم به اندازه کافی قابل ملاحظه نیست تا کاربران بتوانند تصمیم بگیرند، ما یک رویکرد جدید برای حل این محدودیت‌ها پیشنهاد می‌کنیم.

۳. روش پیشنهادی

با توجه به نیاز رویکرد جدید در پایگاه داده‌های گراف NoSQL به منظور بهره‌گیری از پرس‌وجوهای فازی، یک رویکرد مبتنی بر فازی ارائه می‌شود که درخواست‌های مبهم را به عنوان ورودی‌ها پذیرفته و پاسخ‌های دقیق را به مشتریان ارائه دهد. رویکرد پیشنهادی از یک چارچوب استفاده می‌کند که شامل سه موجودیت مشتری، مشاور و شرکت است. شکل ۱، روابط بین سه موجودیت را با جزئیات نشان می‌دهد. هر یک از مشتریانی که دارای درخواست‌ها و محدودیت‌های خود هستند، می‌توانند درخواست‌های خود را با استفاده از زبان فازی مطرح کرده و آن‌ها را به مشاور ارسال کنند. مشاور یک فرد متخصص است که درخواست مشتری را بررسی و تحلیل کرده و مشتریان را برای رسیدن به اهداف خود راهنمایی می‌کند. مشاور، اطلاعاتی راجع به ویژگی‌ها و مشخصات داده‌ها دارد که در پایگاه داده گرافی ذخیره شده‌اند. مشاور همانند یک عامل^۱ عمل کرده و مشتری را برای رسیدن به اهداف خود راهنمایی می‌کند. به عنوان مثال، یک مشاور در یک کارخانه تولید محصول، اطلاعاتی در مورد محصولات که در کارخانه تولید می‌شوند، مانند هزینه، کیفیت و غیره دارد. او اطلاعات محصول را به صورت دوره‌ای از طریق شرکت دریافت کرده، داده‌های بازیابی را پردازش می‌کند و سپس، چندین پاسخ دقیق را به مشتریان عرضه می‌کند. در نهایت، مشتریان یکی از گزینه‌های ارائه شده را انتخاب می‌کنند. مشاور مانند یک عامل فازی عمل می‌کند که اطلاعات فازی را دریافت کرده و نتایج را به صورت غیرفازی برمی‌گرداند.

1. agent



شکل ۱. روابط بین مشتری، مشاور و شرکت

با توجه به دسترسی مشاور به اطلاعات شرکت، خروجی‌ها در سه مرحله مانند شکل ۲، ایجاد می‌شوند:

مرحله فازی سازی: در این مرحله، اصطلاح‌های فازی^۱، توابع عضویت و مجموعه‌های فازی توسط مشاور تعیین می‌شود.

مرحله تعریف قانون: برای تعیین قوانین فازی^۲ در موتور استنتاج فازی، قوانین اگر- سپس^۳ به‌طور مستقل^۴ و خودسرانه^۵ تعیین می‌شوند.

مرحله غیرفازی سازی: در این مرحله، تعدادی از قوانین بر اساس ترجیحات مشتری انتخاب^۶ می‌شوند. روش «ممدانی»^۷ برای ترکیب قوانین به‌منظور تعیین خروجی هر قانون استفاده می‌شود (Wang 2015). در نهایت، خروجی‌ها ترکیب شده و روش مرکز ثقل (Mogharreban and Dilalla 2006) برای محاسبه نتیجه نهایی استفاده می‌شود.

1. fuzzy statements

2. fuzzy rules

3. if-then

4. independently

5. arbitrarily

۶. در اصطلاح گفته می‌شود که قوانین fire شده‌اند.

7. Mamdani



شکل ۲. مراحل تولید خروجی با استفاده از منطق فازی

روش به دست آوردن مرکز ثقل، پرکاربردترین روش در غیرفازی سازی محسوب می شود. به این منظور، در روش پیشنهادی از الگوریتم ارائه شده در شکل ۳، که از روش محاسبه مرکز ثقل بهره می گیرد و این مقدار را با تقریب مناسبی محاسبه می نماید، استفاده شده است.

شبه کد الگوریتم محاسبه مرکز ثقل

- (۱) بازه‌ی مورد بررسی به تعداد دلخواه تقسیم می شود.
- (۲) مقدار μ به ازای تمامی نقاط میانی بازه‌های تقسیم بندی شده محاسبه می شود $a \leq \mu = \frac{b-a}{2}$ نقطه آغازین بازه و b نقطه پایانی بازه است).
- (۳) به ازای تمامی نقاط میانی، $sum = \sum \mu$
- (۴) $half-sum = sum/2$
- (۵) $Result = 0$
- (۶) از نقطه مقدار بازه دوباره مرحله دو انجام می شود و نتیجه در متغیر $result$ ذخیره شده و با یکدیگر جمع می شوند.
- (۷) این روند تا زمانی که شرط $(Result \leq half-sum)$ درست شود، تکرار می شود.

شکل ۳. شبه کد مربوط به الگوریتم محاسبه مرکز ثقل

نقطه‌ای را که در آن شرط موجود در مرحله هفتم برقرار شده، نقطه مرکز ثقل می نامیم.

فرایند غیرفازی سازی تنها یک بار در طول عمر یک پایگاه داده اجرا می شود. در حقیقت، با توجه به داده‌های ذخیره شده در پایگاه داده تمام حالت‌های ممکن که از طریق ترکیب آن‌ها اتفاق می افتد، تعریف می شوند. پس از آن، مرکز ثقل برای تمام

این حالت‌ها اندازه‌گیری و به‌عنوان متغیر در پایگاه داده ذخیره می‌شود. به این ترتیب، در زمان پردازش درخواست‌های مشتری، این متغیرها مورد استفاده قرار می‌گیرند و نیازی به محاسبه مرکز ثقل برای هر پرس‌وجو وجود ندارد و در نهایت، بهترین گزینه‌ها به کاربران ارائه می‌شود. بنابراین، خروجی‌های مراحل توصیف‌شده پیشنهادهایی است که توسط سیستم فازی تحت نظر مشاور ایجاد می‌شود؛ به‌طوری که هر کدام از این پیشنهادها بر اساس ترجیحات مشتری و دانش مشاور است و در نهایت، این پیشنهادات رتبه‌بندی و به مشتری ارائه می‌شود.

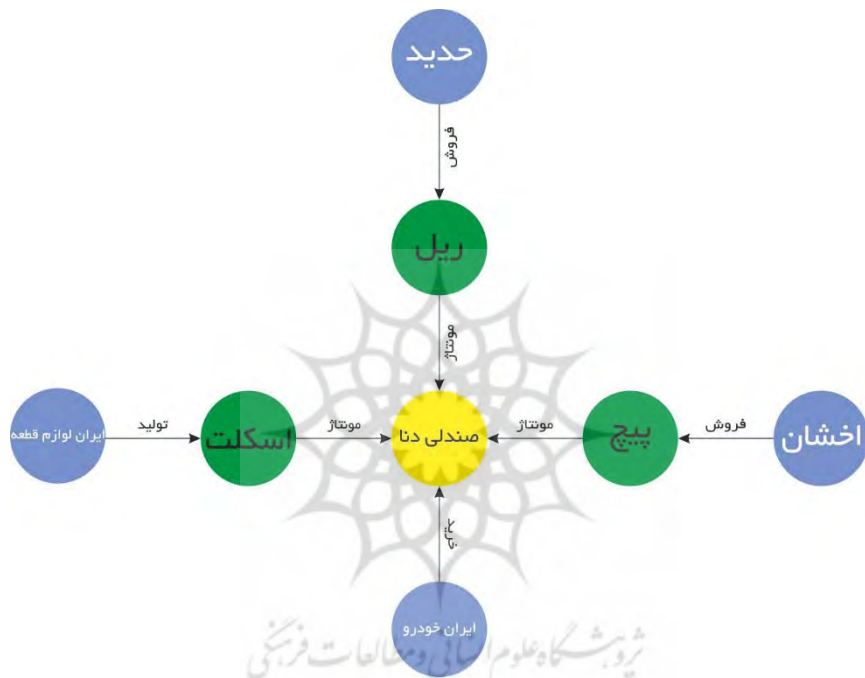
۴. تجزیه و تحلیل یافته‌ها

در این بخش چارچوب پیشنهادی بر روی یک مثال واقعی بررسی و تحلیل شده است. همان‌طور که اشاره شد، ما از پایگاه داده گرافی Neo4j به‌عنوان موتور پایگاه داده استفاده کرده‌ایم. زبان برنامه‌نویسی C# و محیط Visual Studio برای فراخوانی درایورهای Neo4j و تعریف توابع عضویت به‌منظور فازی‌سازی و عملیات غیرفازی‌سازی استفاده شده‌اند.

در این مقاله داده‌های واقعی یک شرکت در حوزه قطعه‌سازی و صندلی خودرو که تولیدکننده قطعات و صندلی خودرو است، بررسی شده است. بعضی از مواد اولیه برای تولید صندلی توسط خود شرکت تولید و سپس استفاده شده و برخی از مواد اولیه نیز توسط تولیدکنندگان قطعات اولیه تأمین می‌شود و در نهایت، صندلی‌های تولیدشده به شرکت‌های دیگر فروخته می‌شود. هر شرکت و یا تأمین‌کننده مواد اولیه به‌صورت یک گره با خاصیت‌های نام، شناسه، شهر، تلفن و نوع تأمین‌کننده یا خریدار بودن آن مشخص می‌شود. قطعات اولیه که تولید و یا از تأمین‌کنندگان خریداری می‌شوند، در گره‌هایی با ویژگی‌های نام، شناسه، بارکد، گروه، کیفیت و قیمت ذخیره می‌شوند. قطعه‌نهایی که همان صندلی‌های تولیدشده هستند نیز توسط گره‌هایی با ویژگی‌های نام، شناسه، بارکد، کیفیت و قیمت مشخص می‌شود. علاوه بر گره‌هایی که برای شرکت‌های تولیدی و محصولات وجود دارد، بین گره‌ها نیز روابطی وجود دارد که ارتباط آن‌ها با یکدیگر و همسایگی آن‌ها را نشان می‌دهد. بین شرکت‌های تولیدکننده و قطعات اولیه رابطه تولید، بین شرکت خریدار قطعات اولیه و قطعات اولیه فروش، بین قطعات اولیه و محصول نهایی رابطه مونتاژ و در نهایت، بین شرکت‌های خریدار و محصول نهایی

رابطه خرید وجود دارد.

شکل ۴، یک نمونه کوچک به‌ازای یک محصول نهایی از این ساختار را نشان می‌دهد. گره‌های آبی شرکت‌های تولیدکننده مواد اولیه یا خریدار را نمایش می‌دهند، گره‌های سبز مواد اولیه همانند اسکلت، ریل و پرینز کمر بند است و گره زرد نیز محصول نهایی را نشان می‌دهد. روابط تولید، خرید و فروش نیز بین گره‌ها دیده می‌شود.



شکل ۴. فرایند تولید یک محصول به همراه روابط آن

در مرحله بعد با استفاده از این کلاس مشخصات گره‌های مرتبط با شرط در این کلاس قرار داده شده و اطلاعات مربوطه در یک جدول داده ریخته و در نهایت، در جدول قرار می‌گیرد.

۴-۱. شبیه‌سازی و ارزیابی

آزمایش انجام شده بر روی یک سرور با CPU ۲۲ هسته‌ای ۲/۲ گیگاهرتز، رم ۶۴ گیگابایت و هارد یک ترابایتی صورت گرفته است. در این آزمایش و پرس‌وجو هشت شرکت تأمین کننده و خریدار، ۶ میلیون محصول اولیه و ۲ میلیون محصول تولید شده داریم.

هدف در این مقاله یافتن محصولات با کیفیت خوب و هزینه متوسط است.

۴-۲. تحلیل داده‌ها

در آزمایش ما، پرس‌وجوهای فازی برای گزارش هزینه و کیفیت محصولات تولید شده استفاده می‌شود. نمونه‌هایی از مجموعه‌های فازی قیمت، کیفیت و خرید تعریف شده است. شکل ۵، روند فرایند غیرفازی‌سازی را نشان می‌دهد. اصطلاح‌های فازی هزینه به‌عنوان یکی از متغیرهای زبانی ما به چهار متغیر زبانی تقسیم شده که Usual، Cheap، Expensive و Lux نام‌گذاری شده است. همچنین، اصطلاح‌های زبانی کیفیت عبارت است از: Normal، Low، Middle، Good، High و عبارت فازی برای خرید تقسیم نیز با مقادیر Low، Normal، High تعریف می‌شوند. اصطلاح‌های فازی توسط مشاور در نرم‌افزار همانند شکل ۵ (الف) تعریف شده است. همچنین، توابع عضویت این سه اصطلاح در شکل ۵ (ب) نشان داده شده است.

در یکی از آزمایش‌ها می‌خواهیم مقدار مرکز ثقل (CoP)^۱ را برای نتیجه یک پرس‌وجوی فازی بیابیم. با فرض یک پرس‌وجوی فازی برای محصولات با کیفیت پایین و هزینه معمولی، بسیاری از حالت‌های مختلف ممکن است اتفاق بیفتند. به‌عنوان مثال، یکی از این حالات می‌تواند پرس‌وجویی باشد که می‌خواهند محصولی با کیفیت ۳۵ (کیفیت آن می‌تواند بین ۰ تا ۱۰۰ باشد) و قیمت ۷۰۰۰۰۰ را پیدا کنند و قوانین انتخاب‌شده آن مطرح شده است. در نهایت، نمودارهای مشتق‌شده از طریق انتخاب‌شدن قوانین بر روی کل داده‌ها در پایگاه داده گراف مشخص شده (همانند شکل ۵ (ج)) و روش استفاده از روش مرکز ثقل برای محاسبه مقدار دقیق نشان داده شده است. نتیجه نهایی برای این پرس‌وجو یک محصول با شباهت ۷۷ درصد به درخواست مشتری است و به این معناست که سیستم، این محصول را به‌عنوان بهترین گزینه مناسب برای آنچه که مشتری می‌خواهد، یافته است که در شکل ۵ (د) نشان داده شده است.

1. center of point

کیفیت

Period Name	Function Type	A	B	C	D
Low	RTrapezoidal			20	40
Mid	Trapezoidal	30	40	50	60
Good	Triangular		55	75	90
High	LTrapezoidal			85	95

هزینه

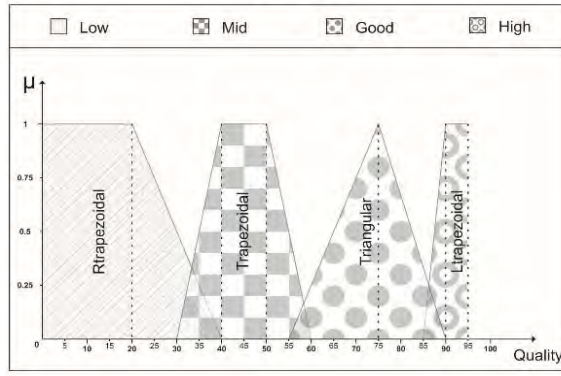
Period Name	Function Type	A	B	C	D
Cheap	RTrapezoidal			100000	400000
Usual	Trapezoidal	360000	500000	660000	860000
Expensive	Triangular		800000	1000000	1200000
LUX	LTrapezoidal			1100000	2000000

خرید

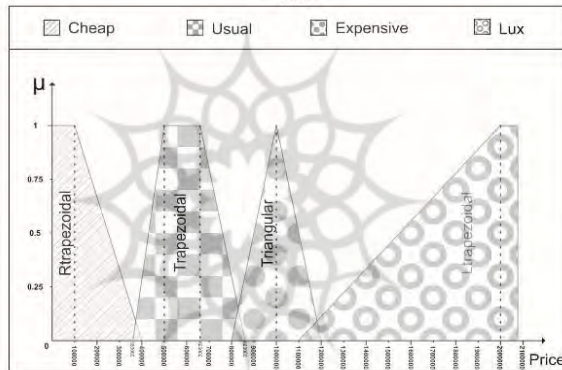
Period Name	Function Type	A	B	C	D
Low	RTrian			0	40
Normal	Triangular		30	50	80
High	LTrian			70	100

(الف)

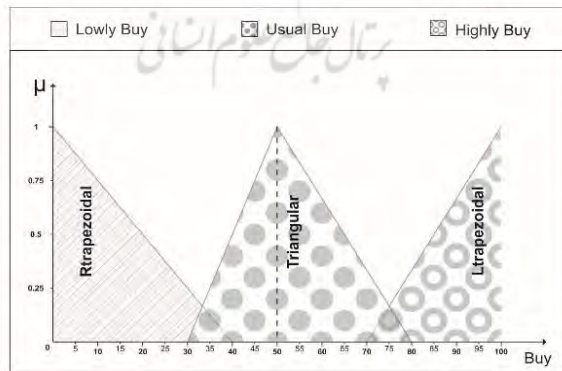
کیفیت



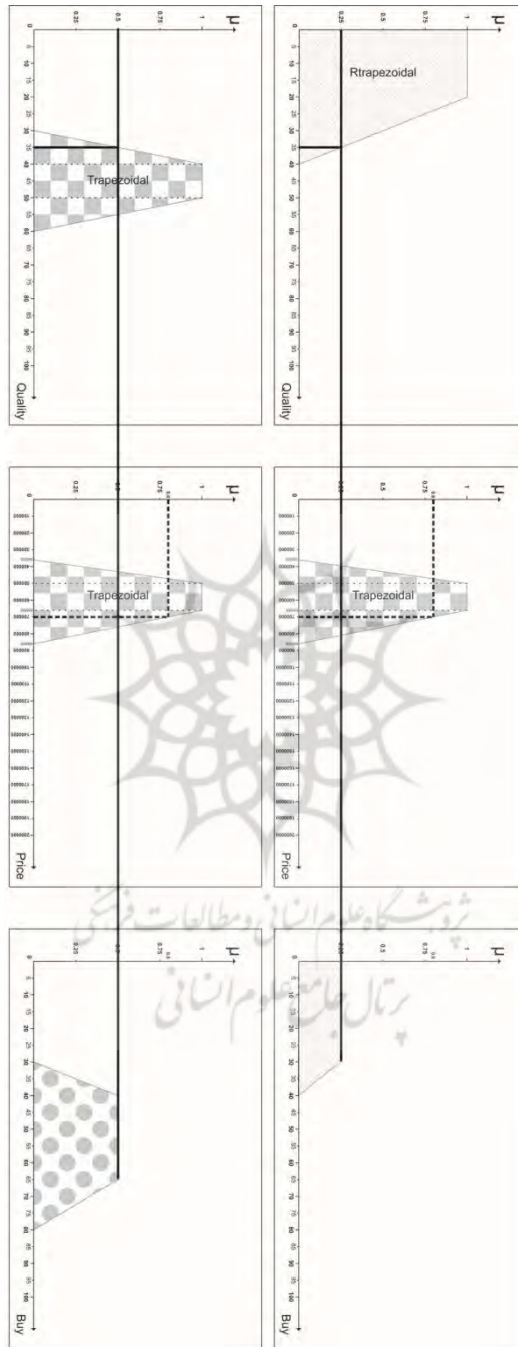
هزینه



پژوهشگاه علم‌الانسان و مطالعات فرهنگی
 خرید



(ب)



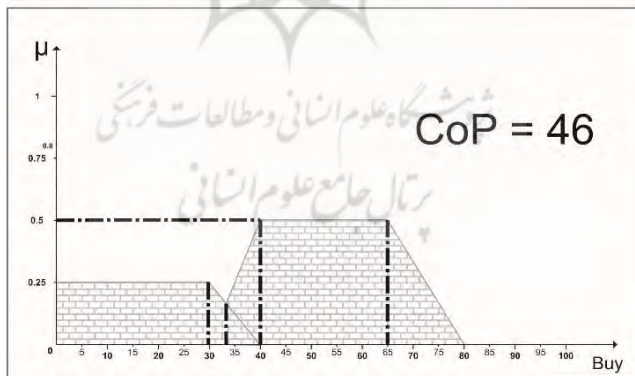
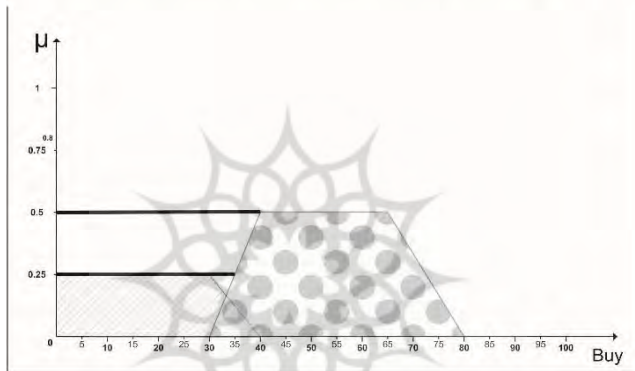
(ج)

انتخاب سطر از پایگاه داده

	Product Name	Quality	Price
▶	DENA JCL01IL0010001	35	700000

سناریوها

F1	F1Deg	F2	F2Deg	Rule	F3Type
Low	0.25	Usual	0.8	Low	RTrian
Mid	0.5	Usual	0.8	Normal	Triangular



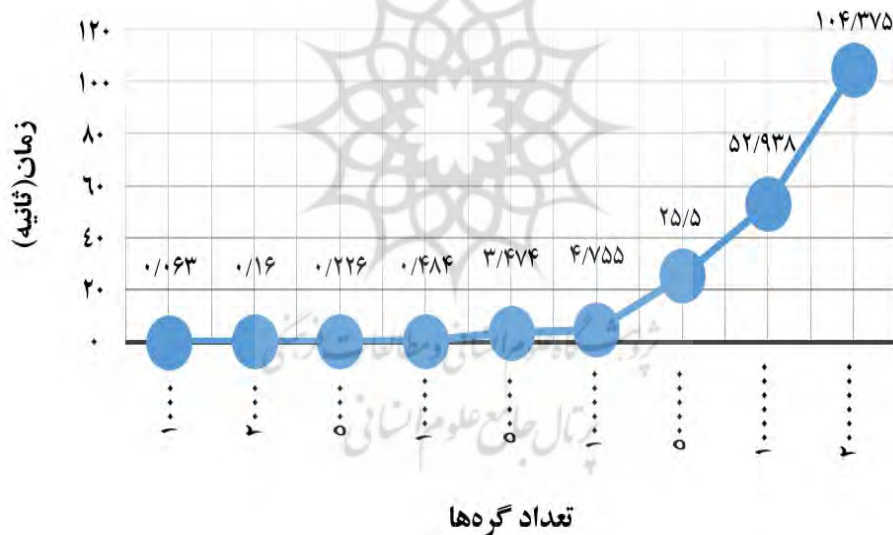
(د)

شکل ۵. غیرفازی سازی و به دست آوردن مرکز ثقل با استفاده از رویکرد پیشنهادی (الف) عبارات فازی برای کیفیت، هزینه و خرید، (ب) نمودار فازی کیفیت، هزینه و خرید، (ج) قوانین fire شده برای کیفیت و هزینه، و (د) جزئیات CoP

۴-۳. ارزیابی و تحلیل نتایج

برای ارزیابی رویکرد پیشنهادی، زمان اجرای رویکرد ارائه شده در موارد مختلف اندازه‌گیری شده است. برآورد فرایند غیرفازی‌سازی و کل زمان لازم برای یافتن نتایج برای یک پرس‌وجوی فازی، ارزیابی‌های گزارش شده در این بخش است. از آنجا که بر اساس بررسی ما، هیچ راه حل دیگری وجود ندارد که پرس‌وجو را با شرایط فازی پذیرفته و نتایج را به صورت دقیق و تفکیک شده نمایش دهد، فقط نتایج پیشنهاد شده را گزارش می‌کنیم. ابتدا، برای بررسی اثر افزایش تعداد گره‌ها بر برآورد زمان غیرفازی‌سازی برای رویکرد پیشنهادی ما، آزمایش انجام شده است که مدت زمان پردازش برای تمام محصولات در پایگاه داده ما را محاسبه می‌کند.

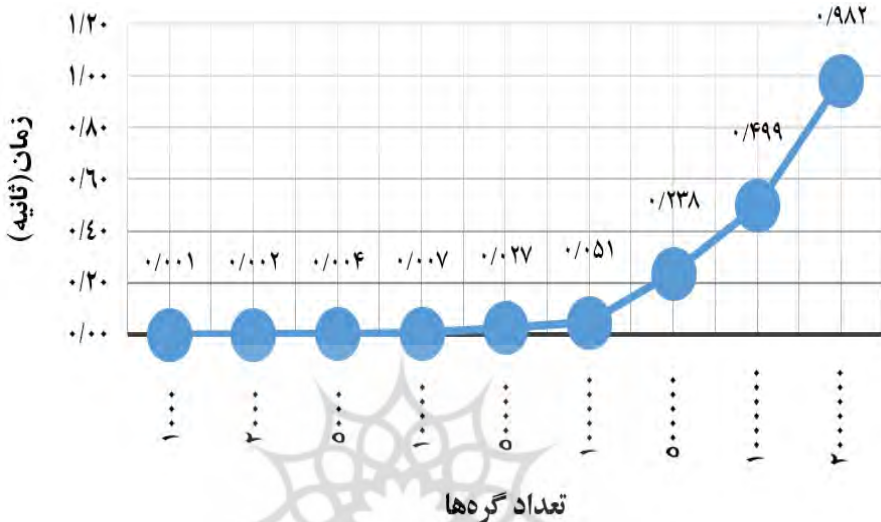
نتایج حاصل از زمان اجرا مرحله غیرفازی‌سازی بر اساس تعداد گره‌های متغیر در پایگاه داده از ۱۰۰۰ تا ۲ میلیون در شکل ۶، نشان داده شده است.



شکل ۶. نتایج زمانی مرحله غیرفازی‌سازی

آزمایش بعدی، ارزیابی کل زمان مورد نیاز برای پیدا کردن نتایج یک پرس‌وجوی فازی در رویکرد پیشنهادی ماست. شکل ۷، زمان مورد نیاز برای پرس‌وجو را با دو اصطلاح فازی نشان می‌دهد که دو فیلتر فازی را بر روی محصولات در پایگاه داده گراف Neo4j بر اساس تعداد گره‌های ۱۰۰۰ تا ۲,۰۰۰,۰۰۰ اعمال می‌کند. بدیهی است که زمان

لازم برای هر پرس و جو با افزایش تعداد گره‌ها افزایش می‌یابد. اما، زمان لازم برای رویکرد پیشنهادی ما غیرمنتظره نیست و با وجود افزایش پیچیدگی منطق فازی، قابل قبول است.



شکل ۷. نتایج زمانی پرس و جو برای محصولی با کیفیت خوب و قیمت متوسط

۴-۴. جمع‌بندی

از آنجا که مشتریان خواستار بازدهی بیشتری هستند تا پرس و جوی خود را با شرایط فازی و متغیرهای زبانی بپرسند، ترجیح بر این است که از منطق فازی استفاده کنیم. پیچیدگی تحمیل شده و تأثیر منطق فازی بر افزایش زمان برای بهبود کارایی و پذیرش رویکرد ما توسط مشتری قابل پذیرش است. از این رو، با وجود این افزایش زمان و بهبود به دست آمده در بهره‌وری، پیشنهاد ما به عنوان یک راه‌حل قابل قبول برای استفاده در داده‌های بزرگ و به‌طور خاص پایگاه داده‌های گراف Neo4j مطرح است.

علاوه بر این، اگرچه راه‌حل‌هایی وجود دارند که منطق فازی را مانند رویکرد ارائه شده در پژوهش (Castellort and Martin (2018 اجرا می‌کنند، اما آن‌ها مشتریان را مجبور می‌کنند که دانش برنامه‌نویسی بالایی داشته باشند. این است که کاربر این راه‌حل‌ها را کمتر به کار برده و پذیرش کمتری خواهند داشت؛ در حالی که در رویکرد پیشنهادی، متغیرهای فازی توسط مشاور ارائه شده و بنابراین، عدم نیاز به سطح بالایی از

مهارت‌های برنامه‌نویسی در مقایسه با نمونه موجود یکی دیگر از مزایای رویکرد پیشنهادی ما محسوب می‌شود.

5. نتیجه‌گیری و کارهای آتی

توانایی پایگاه داده‌های گراف NoSQL برای ذخیره و مدیریت داده‌های بزرگ منجر به استفاده از آن‌ها در حوزه‌های گوناگون مانند صنعت، تجارت الکترونیک و شبکه‌های اجتماعی شده است. کاربران می‌توانند مزایای عبارات فازی و مبهم در نمایش داده‌های خود را برای بهبود تجزیه و تحلیل و مدیریت داده‌ها مورد استفاده قرار دهند. اگرچه چندین راه‌حل برای بیان پرسش‌های فازی ارائه شده است، اما نتایج پرس‌وجوها به صورت فازی ارائه می‌شود. به همین دلیل، یک رویکرد مبتنی بر فازی در این مقاله ارائه شده است که فاکتورهای فازی را پذیرفته و نتایج را به صورت دقیق به مشتریان ارائه می‌دهد.

اصطلاح‌های فازی و قوانین روش پیشنهادی توسط زبان C# اجرا شده است. همچنین، با توجه به داده‌های ذخیره‌شده در پایگاه داده، تمام حالت‌های ممکن که با ترکیب قوانین فازی اتفاق می‌افتد، در نظر گرفته شده است. به این ترتیب، در زمان پردازش درخواست‌های مشتری، جست‌وجو برای یافتن بهترین گزینه با شرایط درخواست مشتری صورت می‌گیرد و نتایج بر اساس درصد تشابه ارائه می‌شود. در نتیجه، بر اساس ارزیابی‌های انجام‌شده، افزایش پیچیدگی استفاده از منطق فازی در رویکرد پیشنهادی، تأثیرات غیرقابل قبول را بر زمان عملیات و بهبودپذیری کارایی ارائه نمی‌دهد. رویکرد پیشنهادی می‌تواند با اجرای سایر توابع عضویت فازی، روش‌های تعریف‌شده، و ارائه گزینه‌های مختلف در یک چارچوب کاربرپسند برای پردازش پرس‌وجوهای فازی بهبود یابد. علاوه بر این، یک سیستم خبره می‌تواند بر اساس موتور فازی اجراشده برای تسهیل توسعه سیستم‌های پیشنهاددهنده طراحی شود.

References

- Angles, R., and C. Gutierrez. 2008. Survey of graph database models. *ACM Computing Surveys (CSUR)* 40 (1): 1.
- Bosc, P., and O. Pivert. 1995. SQL: a relational database language for fuzzy querying. *IEEE transactions on Fuzzy Systems* 3 (1): 1-17.
- Castellort, A., and A. Laurent. 2014. Fuzzy queries over NoSQL graph databases: perspectives for extending the cypher language. International Conference on Information Processing and

- Management of Uncertainty in Knowledge-Based Systems, Springer. Montpellier, France.
- Castellort, A., and T. Martin. 2018. Handling scalable approximate queries over NoSQL graph databases: Cypherf and the Fuzzy4S framework. *Fuzzy Sets and Systems* 348: 21-49.
- Cheng, J., Z., Ma and L. Yan. 2010. f-SPARQL: a flexible extension of SPARQL. International Conference on Database and Expert Systems Applications, Springer. Bilbao, Spain.
- Fowler, A. 2015. *NoSQL for dummies*. USA: John Wiley & Sons.
- Galindo, J., J. M. Medina, O. Pons, and J. C. Cubero. 1998. A server for fuzzy SQL queries. International Conference on Flexible Query Answering Systems, Springer. Roskilde, Denmark.
- Kruse, R., J. E. Gebhardt, and F. Klowon. 1994. *Foundations of fuzzy systems*. USA: John Wiley & Sons, Inc.
- Laney, D. 2001. 3D data management: Controlling data volume, velocity and variety. *META group research note* 6 (70): 1.
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. 2011. Big data: The next frontier for innovation, competition, and productivity. Technical report (Vol. 7), McKinsey Global Institute.
- Mogharreban, N. and, L. Dilalla. 2006. Comparison of defuzzification techniques for analysis of non-interval data. Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American, IEEE. Montreal, Canada.
- Pivert, O., O. Slama and, V. Thion. 2019. Expression and efficient evaluation of fuzzy quantified structural queries to fuzzy graph databases. *Fuzzy Sets and Systems* 366: 3-17.
- Saade, J. J. and, H. B. Diab 2004. Defuzzification methods and new techniques for fuzzy controllers. *Iranian Journal of Electrical and Computer Engineering (IJECE)* 3 (2): 161-174.
- Smets, P. 1997. Imperfect information: Imprecision and uncertainty. Uncertainty management in information systems. Boston: Springer: 225-254.
- Takahashi, Y. 1991. A fuzzy query language for relational databases. *IEEE Transactions on Systems, Man, and Cybernetics* 21 (6): 1576-1579.
- Vicknair, C., M. Macias, Z. Zhao, X. Nan, Y. Chen, and D. Wilkins. 2010. A comparison of a graph database and a relational database: a data provenance perspective. Proceedings of the 48th annual Southeast regional conference, ACM. Oxford Mississippi.
- Wang, C. 2015. A study of membership functions on mamdani-type fuzzy inference system for industrial decision-making. Master of Science thesis, Department of Mechanical Engineering, Lehigh University.
- Yager, R. R. 2014. Social network database querying based on computing with words. Flexible approaches in data, information and knowledge management. Cham: Springer: 241-257.
- Zadrozny, S. and, J. Kacprzyk. 1998. Implementing fuzzy querying via the internet/WWW: Java applets, activeX controls and cookies. International Conference on Flexible Query Answering Systems, Springer. Roskilde, Denmark.
- Zimmermann, H.-J. 1985. Applications of fuzzy set theory to mathematical programming. *Information sciences* 36 (1, 2): 29-58.

زهرا احمدی

متولد ۱۳۷۰، دارای مدرک تحصیلی کارشناسی ارشد در رشته علوم کامپیوتر از دانشگاه علامه طباطبایی است. بررسی پایگاه داده‌های غیررابطه‌ای از جمله علایق پژوهشی وی است.



فرزام متین‌فر

متولد ۱۳۶۱، دارای مدرک تحصیلی دکتری در رشته مهندسی کامپیوتر نرم‌افزار از دانشگاه اصفهان است. ایشان هم‌اکنون استادیار گروه ریاضی و علوم رایانه دانشگاه علامه طباطبایی است. از جمله علایق پژوهشی وی وب معنایی، پایگاه داده و تجارت الکترونیک است.



فرشته آزادی پرند

متولد ۱۳۵۷، دارای مدرک تحصیلی دکتری در رشته مهندسی کامپیوتر از دانشگاه علم و صنعت ایران است. ایشان هم‌اکنون استادیار گروه ریاضی و علوم رایانه دانشگاه علامه طباطبایی است. از جمله علایق پژوهشی وی پایگاه داده، نظریه بازی‌ها و سیستم‌های توزیع شده است. پایگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

