Identification of Data Quality Indicators for Data Governance with Metasynthesis and Fuzzy Delphi Approach

Sahar Bonyadi
PhD Candidate in Knowledge and Information Science; Department of Knowledge and Cumminucation Sciences; Islamic Azad University; Science and Research Branch; Tehran, Iran Email: sahar_bonyadi2005@yahoo.com

Nadjla Hariri*
PhD in Knowledge and Information Science; Professor; Department of Knowledge and Cumminucation Sciences; Islamic Azad University; Science and Research Branch; Tehran, Iran; Email: nadjlahariri@gmail.com

Seyed Mahdi Taheri
PhD in Knowledge and Information Science; Associate Professor; Department of Knowledge and Information Science; Allameh Tabataba'i University; Tehran, Iran Email: Taherismster@gmail.com

Roya Poornaghi
PhD in Knowledge and Information Science; Associate Professor; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: pournaghi@irandoc.ac.ir

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 40 | No. 3 | pp. 987-1020
Spring 2025
https://doi.org/10.22034/jipm.2024.716653

Received: 28, Jan. 2024 **Accepted: 06, Oct. 2024**

Abstract: Data is considered as an asset in organizations and its quality is an important principle to achieve productivity. For optimal management these organizational assets need a kind of governance, so that with its help, the data align with the goals of our constructive leadership. The purpose of this research is to identify the dimensions of data quality for data governance. To achieve this goal, a two-stage qualitative approach was used. In the first stage of the meta-combination method by searching for the key and keywords of data quality, Data management, data governance and data quality management in Air&Dec databases, Science Direct, Google Scholar, Springer, IEEE and ACM were conducted between the years (1995-2022) and 268 related articles were identified and in a detailed study and evaluation, 62 approved articles were decided. By reviewing and studying these articles fully, 8 concepts were identified. After applying the opinions of professors and three data experts, finally 55 components were extracted for the research question. In the second stage, fuzzy Delphi method was used to get the opinion of experts. For this purpose,

* Corresponding Author

the items needed to design a fuzzy Delphi questionnaire were provided from the output of metacombination, and this process was continued until the experts' opinions on the answers to the questions reached a consensus. 21 experts who had at least one research paper in the field of data quality were selected and finally 14 completed questionnaires were returned. In response to the questions of data quality dimensions for data governance, there are 46 indicators: completeness, timing, communication, accessibility, compliance with laws and standards, confidentiality, interpretability, redundancy (ability to add), reputation and reliability, ability traceability, value (value), simplicity, update, concept, regularity, linkability, referential integrity, uniqueness, purposefulness, auditability, accuracy, comparability, consistency, commonality, completeness, metadata compliance, acceptability, validity, conciseness, applicability (usability), believability, comprehensibility, reliability, reasonableness, consistency, retrievability, reproducibility, ability to display null values, appropriateness, clarity, added value, comprehensiveness, extensibility, patternability and variability were identified. More than 80% of the indicators obtained from the extracombination were accepted by the experts, so our system and data-driven businesses can use these indicators as a priority for measuring the quality of their data.

Keywords: Data Quality, Data Governance, Data Quality Dimensions, Meta Synthesis, Fuzzy Delphi

شناسایی شاخص‌های کیفیت داده برای حکمرانی داده با رویکرد فراترکیب و دلفی فازی

سحر بنیادی
دانشجوی دکتری علم اطلاعات و دانش‌شناسی؛ واحد علوم و تحقیقات؛ دانشگاه آزاد اسلامی؛ تهران، ایران sahar_bonyadi2005@yahoo.com

نجلا حریری
دکتری علم اطلاعات و دانش‌شناسی؛ استاد؛ واحد علوم و تحقیقات؛ دانشگاه آزاد اسلامی؛ تهران، ایران؛ پدیدآور رابط nadjlahariri@gmail.com

سید مهدی طاهری
دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛ دانشگاه علامه طباطبائی؛ تهران، ایران؛ Taherismster@gmail.com

رؤیا پورنقی
دکتری علم اطلاعات و دانش‌شناسی؛ دانشیار؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران pournaghi@irandoc.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۲/۱۱/۰۸ **پذیرش:** ۱۴۰۳/۰۷/۱۵ مقاله برای اصلاح به مدت ۱۱ روز نزد پدیدآوران بوده است.

چکیده: داده‌ها در سازمان‌ها به‌عنوان دارایی هستند و کیفیت آن‌ها، اصلی مهم برای رسیدن به بهره‌وری سازمان‌هاست. برای مدیریت بهینه این دارایی‌های سازمانی، به‌نوعی حکمرانی نیاز است تا به کمک آن، داده‌ها را با اهداف راهبردی سازمان همراستا کرد. هدف این پژوهش شناسایی ابعاد کیفیت داده برای حکمرانی داده است. برای تحقق این هدف از یک رویکرد دو مرحله‌ای کیفی استفاده شد. در مرحله اول، روش فراترکیب با جست‌وجوی کلیدواژه‌های کیفیت داده، مدیریت داده، حکمرانی داده و مدیریت کیفیت داده در پایگاه‌های ایرانداک، Google، Science Direct، IEEE، Springer، Scholar و ACM در بین سال‌های (۱۹۹۵-۲۰۲۲) انجام و ۲۶۸ مقاله مرتبط تشخیص داده شد و در مراحل مطالعه و ارزیابی دقیق ۶۲ مقاله مورد تأیید قرار گرفت. با بررسی و مطالعه کامل این مقالات ۸ مفهوم شناسایی شد. بعد از اعمال نظر اساتید و سه نفر از متخصصان داده سرانجام، ۵۵ مؤلفه برای سؤال پژوهش استخراج گردید. در مرحله دوم،

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 ۴۰ | شمارهٔ ۳ | ص‌ص ۹۸۷-۱۰۲۰
بهار ۱۴۰۴
https://doi.org/10.22034/jipm.2024.716653

به‌منظور دریافت نظر خبرگان از روش دلفی فازی استفاده شد. به همین منظور از خروجی فراترکیب، گویه‌های لازم برای طراحی پرسشنامه دلفی فازی تأمین شد و این فرایند تا جایی ادامه یافت که نظرات خبرگان در مورد پاسخ به سؤالات به اجماع رسید. ۲۱ نفر از خبرگان که حداقل دارای یک مقاله پژوهشی در حوزهٔ کیفیت داده بودند، انتخاب شدند و سرانجام، ۱۴ پرسشنامهٔ تکمیل‌شده برگردانیده شد. در پاسخ به سؤالاتِ ابعاد کیفیت داده برای حکمرانی داده، تعداد ۴۶ شاخص شامل کامل بودن، زمان‌مند بودن، ارتباط، دسترس‌پذیری، مطابقت با قوانین و استانداردها، محرمانه بودن، تفسیرپذیری، افزونگی (قابلیت اضافه کردن)، شهرت و اعتمادپذیری، قابلیت ردیابی، ارزش (مقدار)، سادگی، به‌روزرسانی، مفهوم، نظم‌پذیری، پیوندپذیری، یکپارچگی مرجع، یکتایی، عینیت، امکان ممیزی، صحت، مقایسه‌پذیری، ثبات، امنیت دسترسی، رایج و شایع بودن، تمامیت، انطباق فراداده، مقبولیت، اعتبار، موجز بودن، کاربردپذیری (قابلیت استفاده)، باورپذیری، قابلیت درک، قابلیت اطمینان، منطقی بودن، سازگاری، قابلیت بازیابی، تکرارپذیری، قابلیت نمایش مقادیر پوچ، تناسب، وضوح، ارزش افزوده، جامعیت، توسعه‌پذیری، الگوپذیری، و تغییرپذیری شناسایی شدند. بیش از ۸۰ درصد شاخص‌های حاصل از فراترکیب از نظر خبرگان پذیرفته شدند. بنابراین، سازمان‌ها و کسب‌وکارهای داده‌محور می‌توانند این شاخص‌ها را در اولویت سنجش کیفی داده‌های خود قرار دهند

کلیدواژه‌ها: کیفیت داده، حکمرانی داده، ابعاد کیفیت داده، فراترکیب، دلفی فازی

۱. مقدمه

تعریف داده کار مشکلی است و تنها از دیدگاه استفاده‌کنندگان می‌توان آن‌ها را از هم تشخیص داد (کرمی و اسفیدانی ۱۳۸۳، ۵۳). بعضی از تعاریف داده از دیدگاه صاحب‌نظرانِ مختلف ارائه شده است. واژه «داده» مناسب‌ترین واژه‌ای است که به واقعیات شکل‌نیافته و بدون ساختار و فراوانِ تولیدشده توسط کامپیوتر می‌توان اطلاق نمود که بر اعداد، نمودارها و دیگر نوشته‌ها دلالت می‌کند و به تنهایی معنا ندارد (رضاییان ۱۳۷۴، ۱۶). به بیان دیگر، داده‌ها حقایق و واقعیت‌های خام هستند و این اجزا در پایگاه‌های داده[1] ذخیره و مدیریت می‌شوند. به‌عنوان مثال ۲، ۱۵ و محمد نمونه‌ای از داده هستند و تا زمانی که پردازش نشوند، هیچ برداشتی از این سه داده صورت نمی‌گیرد. «ردمن»[2] معتقد است که داده‌ها عناصر اصلی اطلاعات هستند (ردمن ۱۳۸۱، ۸۳). داده‌ها در صورتی به اطلاعات تبدیل می‌شوند که افراد بخواهند برای درک بیشتر از آن‌ها استفاده کنند. به‌طور کلی،

1. data bases
2. Redman

داده‌ها عبارت‌اند از «اطلاعات، به‌ویژه حقایق یا اعداد جمع‌آوری‌شده برای بررسی و در نظر گرفتن و استفاده برای کمک به تصمیم‌گیری. اطلاعات به شکل الکترونیک می‌تواند توسط رایانه ذخیره و استفاده شود» (Hassenstein & Vanella 2022).

اطلاعات، داده‌های خلاصه‌ای هستند که گروه‌بندی، ذخیره، پالایش و سازماندهی شده‌اند تا معنادار شوند (رادینگ[1] ۱۳۸۳). اطلاعات زمانی ارزش پیدا می‌کنند که برای یک بعد خاص، یک فرد خاص، یک هدف خاص، و در زمان خاص گردآوری و آماده شوند. بنابراین، اطلاعاتی که برای یک مدیر، جنبه اطلاعاتی دارد، برای مدیر دیگر ممکن است اصلاً ارزشی نداشته باشد (اخوان آملی ۱۳۷۵، ۴۰)

اهمیت دادن به کیفیت داده در سازمان‌ها نشانگر دید عمیق مدیران به این مسئله و استقرار حاکمیت داده در یک سازمان داده‌محور است. امروزه، داده‌ها اغلب با سیستم‌های اطلاعاتی مرتبط هستند و طی انجام فرایندهای کاری که به‌وسیله سیستم‌های اطلاعاتی مدیریت می‌شوند، ایجاد می‌گردند. سودآوری بیشتر و پرهیز از هزینه‌های غیرضروری یکی از اهداف اصلی سازمان‌های تولیدکننده است و این مهم محقق نمی‌شود، مگر با تولید محصولات باکیفیت که با استفاده از سیستم‌های کارا و برنامه‌ریزی‌های صحیح قابل دستیابی هستند. لازمه تحقق این موارد داشتن داده‌های باکیفیت است. داده‌ای باکیفیت است که برای استفاده در فرایندها و تصمیم‌گیری‌ها مناسب باشد و الزامات مورد نیاز برای استفاده مورد نظر را برآورده کند. داده‌های با کیفیت بالا کلید تجزیه‌وتحلیل داده‌های قابل تفسیر و قابل اعتماد و مبنایی برای تصمیم‌گیری‌های داده‌محورِ معنادار است. در سناریوهای عملی، کیفیت داده به‌طور معمول، با پیش‌پردازش، پروفایل و پاکسازی داده‌ها برای کارهای بعدی مانند یکپارچه‌سازی داده یا تجزیه‌وتحلیل داده‌ها مرتبط است. با این حال، از منظر علمی، پژوهش‌های زیادی در مورد اندازه‌گیری (یعنی تشخیص) مسائل کیفیت داده منتشر شده و ابعاد و معیارهای مختلف کیفیت داده به‌طور کلی قابل اجرا مورد بحث قرار گرفته است (Ehrlinger & Wöß 2022).

یکی از موضوعات مهم در مدیریت داده‌ها، بحث کیفیت داده است. دادهٔ خوب و باکیفیت مبنای رسیدن به نتایج مطلوب از سیستم‌هاست. به بیان دیگر، اثربخشی مدیریت یک کسب‌وکار در سازمان به‌واسطهٔ کیفیت داده‌های آن سازمان به‌عنوان ماده خام

1. Radding

تصمیم‌گیری تعیین خواهد شد (Chaffey & White 2010). سیستم اطلاعاتی همراه با داده بی‌کیفیت، حتی اگر طراحی و پیاده‌سازی مناسبی داشته باشد، نمی‌تواند تصمیم‌گیران و کاربران سیستم‌های اطلاعاتی را به نتایج مطلوب مد نظر برساند. بنابراین، تمرکز بر روی موضوع کیفیت داده در سیستم‌های اطلاعاتی امری اجتناب‌ناپذیر است (خلیلی جعفرآباد ۱۳۹۶)

۲. مبانی نظری

منظور از کیفیت داده این است که خصوصیات داده تا چه اندازه می‌تواند خواسته‌های کیفی مورد انتظار را مرتفع سازد (خسروانجم ۱۳۹۲). اولین گام برای بهبود کیفیت داده‌ها و ارتقای آن‌ها توجه به معنای کیفیت داده است. سازمان در درجه اول باید بداند که کیفیت داده چیست تا در مراحل بعدی به فکر توسعه آن باشد (Wang & Strong 1996). تعریف کیفیت داده‌ها وابسته به هدف مورد نظر است و بنابراین واژه‌ای نیست که به‌صورت رسمی قابل تعریف باشد (نورزاد ۱۳۹۰). مفهوم کیفیت به‌طور کلی، امری پیچیده و در عین حال، چند بعدی است (غفاری، عادل و شباک ۱۳۹۷، ۵).

در پیشینه پژوهش از مفهوم «مناسب برای کاربرد» یا برآوردن نیازهای کاربر نهایی صحبت می‌شود. طبق تعریف کلی مدیریت کیفیت، کیفیت داده‌ها را به‌عنوان برآوردن نیازهای مشتریان به‌صورت منظم تعریف می‌کنند. بر خلاف باور عمومی، کیفیت لزوماً نبودِ خطا در حد صفر نیست (Hipp, Guntzer & Grimmer 2002). داده‌هایی دارای کیفیت بالا هستند که برای مقاصد مورد نظر در «کاربردها»، «تصمیم‌گیری‌ها» و «برنامه‌ریزی‌ها» مناسب باشند. از سوی دیگر، می‌توان داده‌های دارای کیفیت بالا را داده‌هایی دانست که به‌درستی بیانگر موجودیت‌های معادل خود در جهان واقع باشند (نورزاد ۱۳۹۰)

کیفیت داده‌ها با مورد استفادهٔ آن‌ها در فرایندها و تصمیم‌گیری‌ها ارتباط دارد. داده‌ای باکیفیت در نظر گرفته می‌شود که برای استفاده در فرایندها و تصمیم‌گیری‌ها مناسب است و الزامات مورد نیاز برای استفاده مورد نظر را برآورده می‌کند. (محجوب و سیمه‌ساز ۱۳۹۴)

به‌طور کلی، در تعریفی که از کیفیت داده ارائه شده، آن را متناسب با نیاز مورد نظر کاربر تعریف کرده‌اند. در این تعریف، مفهوم کیفیت داده باید با نیاز کاربران متناسب باشد (Tayi & Ballou 1998)

هر قدر داده‌های مورد نظر بتوانند اهداف بیشتری از نیاز کاربران را محقق نمایند، از سطح کیفیتی بالاتری برخوردار هستند. آنچه در این تعریف قابل تأمل است، آن است که کیفیت داده ارتباط نزدیکی با کاربر و نیاز او دارد (Larburu et al. 2015). اکثر پژوهش‌هایی که در زمینهٔ کیفیت داده و اطلاعات صورت گرفته، کیفیت به‌صورت چارچوبی چندبعدی و مجموعه‌ای از ویژگی‌ها در نظر گرفته شده است (Bizer & Cygania 2009). اما به‌طور کلی، یک مجموعه پذیرفته‌شده و عمومی از ابعاد کیفیت و معنای آن‌ها وجود ندارد (Lee & Haider 2011).

۳. تفاوت کیفیت اطلاعات و کیفیت داده

برای تمایز بین دو عبارت «داده» و «اطلاعات» باید در نظر داشت که داده مشاهدات پردازش‌نشده و خام است. اطلاعات، داده سازماندهی‌شده مبتنی بر روابط بین داده‌هاست (Park et al. 2010؛ Chen & Tseng 2011؛ Batini & Scannapieco 2016). بعضی‌ها مبحث کیفیت اطلاعات و داده را به شدت وابسته به دامنه و حیطه پژوهش می‌دانند. آن‌ها همچنین معتقدند که کیفیت داده در بستر عملیات تعریف می‌شود (Bizer & Cygania 2009)، اما اطلاعات می‌تواند برای یک کاربرد مناسب و دارای کیفیت بوده و در کاربردی دیگر نامناسب و بدون کیفیت باشد (Bonner 2010).

٤. بیان مسئله

کیفیت داده بر کیفیت تصمیم‌گیری تأثیر به‌سزایی دارد. ضعف کیفیت داده افزون بر کاهش کیفیت تصمیم‌ها، هزینه زیادی از جمله هزینهٔ اجرای مجدد کل فرایند و هزینه‌های فرصت‌ها و درآمدهای ازدست‌رفته به سازمان را تحمیل می‌کند. بنابراین، سیستم‌های اطلاعاتی فاقد کیفیتِ داده کاربرد مؤثری ندارد و به ایجاد ارزش برای بهره‌داران منجر نخواهد شد. در پژوهش‌های مرتبط با مدل‌های موفقیت سیستم‌های اطلاعاتی از کیفیت داده به‌عنوان یکی از عامل‌های اصلی یاد می‌شود. عقیده بر این است که عوامل موفقیت‌آمیز در ارزیابی سیستم‌های اطلاعاتی، کیفیت داده و رضایت‌مندی کاربران است (Petter, DeLone & McLean 2008). نتایج پژوهش‌های دیگر در این حوزه حاکی از آن است که رضایت از به‌کارگیری سیستم و رضایت کاربران به‌شدت به کیفیت داده وابسته است (عبدالوند و بوبه‌رژ ۱۳۹۳)

مطالعات وسیعی جهت شناخت ابعاد کیفیت داده و ارزیابی آن‌ها در نظام‌های اطلاعاتی صورت گرفته است. این مطالعات نشان می‌دهند که ارزیابی ابعاد مختلف کیفیت داده با استفاده از طراحی و تبیین مدل‌ها، شیوه‌ها، ابزار و چارچوب‌های مختلف انجام شده است (رحیمی ۱۳۹۵).

افزون بر اینکه بخش مهندسی سازمان باید از اجزای باکیفیت تشکیل شده باشد، بخش‌های پشتیبانی هم نیاز به کیفی بودن اجزای خود دارند. تمامی این موارد، سازمان‌ها را بر آن داشته است که از ابزار کیفیت داده‌ها به‌عنوان ابزاری که صحت، دقت و سرعت را در اجرای فرایندهای سازمانی دخیل می‌کنند، سود ببرند

پژوهش‌های مربوط به کیفیت داده، طبقه‌بندی کاملی از ابعاد کیفیت داده‌ها را فراهم می‌کنند. با این‌حال، برخی اختلافات در تعریف مهم‌ترین ابعاد با توجه به ماهیت تعریف‌شده از کیفیت وجود دارد. شش طبقه مهم از ابعاد کیفیت توسط Wand & Wang (1996)؛ Wang & Strong (1996)؛ Redman & Blanton (1996)؛ Jarke et al. (1995)؛ Bovee, Srivastava & Mak (2003)؛ Naumann (2002) ارائه شده است. با تجزیه‌وتحلیل این طبقه‌بندی، تعریف مجموعهٔ اولیه‌ای از ابعاد کیفیت داده‌ها امکان‌پذیر می‌شود که از جمله آن‌ها دقت، جامعیت، ثبات و به‌موقع بودن است و مورد تمرکز اکثر نویسندگان قرار گرفته است. با این‌حال، هیچ توافق کلی بر روی تعریف مجموعه‌ای از ابعاد کیفیت داده‌ها و یا معنای دقیق هر یک از ابعاد وجود ندارد (ارشادی، نصیری و شیرانی ۱۳۹۶). به‌دلیل کثرت ابعاد کیفیت داده و تنوع آن‌ها در استفاده، این پژوهش به بررسی شاخص‌های کیفیت داده از دیدگاه متخصصان ایرانی می‌پردازد.

۵. پیشینه پژوهش

کیفیت داده مفهومی چندبعدی است. به گفته دیگر، نمی‌توان یک مقدار محض را به‌عنوان کیفیت داده به آن منتسب کرد، بلکه بایستی کیفیت داده را از منظرهای مختلف مانند صحت، سازگاری، امنیت، باورپذیری، کامل بودن و دسترس‌پذیری و غیره بررسی نمود (Sumalatha & Ramasubbareddy 2010)

در این بخش، مطالعات مرتبط با حوزه کیفیت داده بررسی می‌شوند.

مطالعات داخلی

«محجوب و سیمه‌ساز» (۱۳۹۴) در پژوهشی با عنوان «کیفیت داده‌ها پیش‌نیاز مدیریت منابع سازمان» بحث می‌کنند که سازمان‌ها می‌توانند با مدیریت و برنامه‌ریزی منابع خود به بهره‌وری رسیده و هزینه‌ها را کاهش دهند. منابع اطلاعاتی به‌عنوان یکی از مهم‌ترین منابع سازمانی باید بتواند نیازهای سازمان را در برنامه‌ریزی‌ها و تصمیم‌سازی‌ها برآورده کند. کیفیت داده‌ها مبحثی است که به شیوه‌های بهبود و مدیریت منابع سازمان و منابع اطلاعاتی مرتبط با آن‌ها می‌پردازد. استاندارد ISO8000 با موضوع کیفیت داده‌ها و استاندارد ISO22745 با موضوع یکپارچگی و سیستم‌های اتوماسیون صنعتی در مقاله مذکور مورد بحث واقع شده است

«ارشادی، نصیری و شیرانی» در مطالعه‌ای به بررسی طراحی مدل کیفیت فراداده در سامانه ثبت پایان‌نامه/ رساله‌های دانش‌آموختگان داخل و خارج کشور در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) پرداختند. این محققان در مطالعه خود بحث می‌کنند که سامانه ثبت پایان‌نامه‌ها و رساله‌های دانش‌آموختگان کل کشور به‌عنوان یک سامانه حیاتی در ثبت و اشاعه مدارک علمی بوده و نقشی انکارناپذیر در توسعه پژوهش‌ها دارد. از این‌رو، سطح کیفیت مطلوب سامانه تأثیر ویژه‌ای در ارتقای کیفیت پژوهش‌های محققان خواهد داشت. نام، کد ملی، مقطع تحصیلی، رشته دانشگاهی، مشخصات استاد راهنما، مشخصات پیشنهاده و پایان‌نامه از مهم‌ترین اقلام اطلاعاتی به شکل داده در فیلدهای مربوط هستند. به‌منظور بهبود سطح کیفیت در این سامانه تعیین وضعیت مطلوب هر یک از اقلام اطلاعاتی در سامانه ثبت، اقدامی بنیادین محسوب می‌شود. در این راستا مدل‌های کیفیت داده از متون علمی مختلف استخراج و جمع‌بندی گردیده و وضعیت مطلوب کیفیت فراداده‌های مربوط به پیشنهاده پایان‌نامه‌ها و رساله‌ها ترسیم گردید. بررسی وضعیت موجود فرایند ثبت و اشاعه مدارک علمی، مشکلات کیفی مشاهده‌شده، سیاست‌ها و راهبردهای اشاعه از اقدامات اصلی بود. نقش هر یک از افراد درگیر در سامانه ثبت (نمایندگان دانشگاه‌ها، کارکنان واحد فراهم‌آوری) در قالب بازبینی فهرست‌های کنترل کیفی فراداده‌ها تبیین گردید. چارچوبی به‌منظور درجه‌بندی این عدم انطباق‌ها (اصلی و جزئی) و همچنین اقدامات لازم در صورت بروز آن‌ها تعیین شد. این اقدامات اصلاح فراداده، حذف آن، پذیرش و ارسال فراداده به نمایه‌ساز و ... را شامل می‌شود. کلیه مستندات تدوین‌شده در قالب طرح کیفیت فراداده سامانه ثبت ارائه شد (۱۳۹۶)

«خلیلی جعفرآباد» در مطالعه‌ای به بررسی تغییرات حوزه کیفیت داده با استفاده از تحلیل کلمات کلیدی پرداخت. این محقق در مطالعه خود بحث می‌کند که حوزه کیفیت داده از جمله حوزه‌های روبه‌رشد و مهم در حوزه سیستم‌های اطلاعاتی است. شناخت دقیق این حوزه از یک سو، و شناخت ویژگی‌های زیرحوزه‌های نوین این حوزه برای محققان از اهمیت بالایی برخوردار خواهد بود. شناخت ویژگی این حوزه و میزان بینِ رشته‌ای بودن آن‌ها به محققان برای تصمیم‌گیری در مورد روند پژوهش‌ها و انتخاب حوزه فعالیت کمک به‌سزایی خواهد کرد. برای شناسایی حوزه‌ها و بررسی میزان بین رشته‌ای بودن آن‌ها، در این مطالعه با استفاده از گراف همرخدادی کلمات و تحلیل ۹۰۰۰ مقاله، ویژگی‌های حوزه‌های نوین مطالعاتی مورد بررسی قرار گرفته است. بر اساس این مطالعه مشخص شد که این حوزه‌ها بیشتر بین رشته‌ای بوده و تمرکز آن بر روی ارتباط بین چندین حوزه مطالعاتی است. به بیان دیگر، این کلمات کلیدی بیشتر با کلمات کلیدی موجود در خوشه‌های دیگر ارتباط داشته‌اند تا با کلمات کلیدی که با آن‌ها در یک خوشه قرار گرفته‌اند. بر اساس این مطالعه جدیدترین حوزه، مرتبط با کلان داده است که مباحث یکپارچه‌سازی و داده گمشده در این حوزه از اولویت بالاتری نسبت به بقیه حوزه‌ها برخوردار است (۱۳۹۶)

«اشتریان اصفهانی، ارشادی و عزیزی» در پژوهشی به بررسی توسعهٔ شاخص‌های کیفیت داده به‌منظور ارزیابی سامانه‌های اطلاعاتی پژوهشی پرداختند. در این پژوهش پس از بررسی متون علمی مختلف، شاخص‌های کیفیت داده مانند دقت، صحت، جامعیت و به‌هنگام بودن برای ارزیابی سامانه‌های اطلاعاتی پژوهش‌ها توسعه داده شدند. سامانهٔ ملی ثبت پایان‌نامه/ رسالهٔ دانش‌آموختگان کل کشور به‌عنوان مطالعهٔ موردی انتخاب شد. نتایج نشان داد که پس از بهبود فرایند ثبت در این سامانهٔ ملی، شاخص‌های کیفیت داده وضعیت بهتری نشان می‌دهد. استفاده از فهرست‌های آمادهٔ کرکره‌ای به‌منظور افزایش خطاناپذیرسازی در ثبت داده‌ها، احراز هویت دانشجو و اساتید راهنما و مشاور به کمک کد ملی از جمله پیشنهادهای اجرایی هستند که در راستای بهبود کیفیت داده‌ها در سامانهٔ ثبت ارائه شدند (۱۳۹۸)

«سهرابی» و همکاران در پژوهشی با عنوان «بررسی شناخت و تحلیل سیستمی متدولوژی‌های کیفیت داده و ارائه یک چارچوب جامع» ضمن به‌کارگیری دیدگاه سیستمی و با استفاده از روش کدگذاری باز، کدهای مربوط به سه مقوله اصلی رویکرد

سیستمی شامل ورودی، فرایند و خروجی استخراج کرده و مفاهیم مشابه را در کدهای فرعی و در ادامه، کدهای فرعی را در کدهای اصلی دسته‌بندی کردند. ورودی‌های اصلی شامل زمینه و وضعیت سازمان، داده‌ها و منابع اطلاعاتی و ابعاد کیفیت داده هستند. همچنین، گام‌های متدولوژی‌های کیفیت داده در سه مرحله اصلی بازسازی وضعیت، ارزیابی/ اندازه‌گیری و ارتقا طبقه‌بندی شده‌اند. افزون بر این، خروجی‌های کیفیت داده در شش دسته کلی شامل فهرست فعالیت‌ها و تکنیک‌های مرتبط مشخص‌شده برای ارتقای کیفیت داده‌ها، فرایندهای کنترل‌شده یا بازطراحی‌شده، جریان‌ها و پایگاه‌های داده اندازه‌گیری یا ارتقا داده‌شده، نتایج ارائه‌شده از وضعیت کیفیت داده، سیاست‌ها یا قوانین کیفیت داده تصحیح‌شده، و هزینه‌ها و منفعت‌ها طبقه‌بندی شده‌اند. نتایج حاصل از این پژوهش می‌تواند ابزار مناسبی جهت شناخت متدولوژی‌های کیفیت داده موجود و همچنین ارزیابی نقاط ضعف و قوت متدولوژی‌های کیفیت داده باشد (۱۴۰۰)

«صالحی» و همکاران در پژوهشی با عنوان «ارائه یک چارچوب مفهومی برای پیش‌پردازش و بهبود کیفیت نگاره‌های رویداد در فرایندکاوی»، ۱۰۲ پژوهش مرتبط با حوزه کیفیت داده در فرایندکاوی را بررسی کردند. مهم‌ترین چالش‌های کیفیت داده در این زمینه پس از پالایش و یکپارچه‌سازی شامل «رویدادهای آشفته/ کم‌تکرار»، «رویدادهای پرت»، «رویدادهای ناهنجار»، «مقادیر گمشده»، «قالب زمانی نادرست»، «برچسب‌های زمانی مبهم»، «فعالیت‌های مترادف» و «اندازه و پیچیدگی» هستند. در ادامه این پژوهش، گام‌های اساسی برای پیش‌پردازش و پاک‌سازی مناسب داده‌ها تعیین شده‌اند که دربرگیرنده فعالیتهای «ترمیم»، «کشف ناهنجاری»، «پالایش» و «کاهش ابعاد» هستند. سپس، چارچوب مفهومی نهایی بر پایه مشکلات کیفیت داده و فعالیت‌های پاک‌سازی شناسایی‌شده ایجاد شده است. در پژوهش آن‌ها، برای بررسی عملکرد چارچوب پیشنهادی از چهار مجموعه داده استاندارد برگرفته از فرایندهای واقعی استفاده شده است. این داده‌ها در مرحله اول به‌صورت خام و در مرحله دوم پس از انجام پیش‌پردازش توسط چارچوب معرفی‌شده به چهار الگوریتم متداول کشف فرایند اعمال شده‌اند. نتایج نشان داد که پیش‌پردازش داده‌های ورودی به بهبود معیارهای کیفیت مدل استخراج‌شده از الگوریتم‌های کشف فرایند منجر می‌شود (۱۴۰۲)

مطالعات خارجی

«دای» و همکاران در پژوهشی چارچوب کنترل و مدل مدیریت کیفیت داده برای پشتیبانی تصمیم‌گیری در حوزه سلامت را پیشنهاد کرد که یک روش بازرسی کیفیت داده، فرایند و تشخیص اجرایی شدن برای داده‌های سلامت با انواع گوناگون بر پایه مدل پیشنهادی طراحی‌شده است (Dai et al. 2015).

«وو» و همکاران در پژوهشی به بررسی «اهمیت کیفیت داده‌ها: مطالعه موردی در مورد صحت برچسب داده‌ها برای پیش‌بینی گزارش اشکال امنیتی» پرداختند. نتایج در پژوهش آن‌ها نشان می‌دهد که مجموعه داده‌های تمیزشده منجر به بهبود عملکرد مدل‌های طبقه‌بندی می‌شوند (Wu et al. 2021).

«استجپاندیچک و کرول» در پژوهشی به بررسی مدیریت کیفیت داده برای قابلیت همکاری پرداختند. در این پژوهش نحوه مدیریت سیستماتیک الزامات کیفیت داده و پشتیبانی از یک مدیریت جامع کیفیت داده ارائه شده است. در این پژوهش پس از تعریف دیجیتال به‌عنوان بزرگراه داده، طبقه‌بندی کیفیت داده‌ها بر اساس ابعاد کیفیت داده‌ها و استانداردهای مرتبط با حوزه تولید مورد بحث قرار گرفته و معیارهای کیفیت داده‌ها بر اساس دستورالعمل‌های توسعه‌یافته توسط نهادهای هماهنگ‌کننده ملی و بین‌المللی از صنعت خودروسازی جهانی کشف شد (Stjepandić & Korol 2022).

«سید» و همکاران در پژوهشی به بررسی مسائل کیفیت داده‌های سلامت دیجیتال پرداختند. چارچوب DQ-DO سلامت دیجیتال شامل ۶ بُعد DQ، یعنی دسترس‌پذیری، دقت، کامل بودن، سازگاری، اعتبار متنی، و ارز است. روابط متقابل بین ابعاد DQ سلامت دیجیتال، با سازگاری که تأثیرگذارترین بُعد است، بر سایر ابعاد DQ سلامت دیجیتال تأثیر می‌گذارد. ۵ پیامد DQ سلامت دیجیتال یعنی نتایج بالینی، پزشک، مرتبط با پژوهش‌ها، فرایند کسب‌وکار و نتایج سازمانی، و روابط بین ابعاد DQ سلامت دیجیتال و نتایج DQ، با ابعاد سازگاری و دسترسی که بر همه نتایج DQ تأثیر می‌گذارد، رابطه دارند (Syed et al. 2023).

در بیشتر پژوهش‌های انجام‌شده، چه در داخل و چه در خارج از کشور، تمرکز محقق بر روی یک یا چند بُعد یا شاخص کیفیت داده بوده است و پژوهشی که بر روی تمامی ابعاد و یا بیشترین ابعاد کار کرده باشد، یافت نشد. «غفاری، آذر، و شباک» (۱۳۹۷) سه بُعدِ انسجام، تضمین درستی، و جامعیت، و « اشتریان اصفهانی، ارشادی و عزیزی» (۱۳۹۸)

چهار بُعدِ دقت، صحت، جامعیت، و به‌هنگام بودن را بررسی کرده‌اند. Peer et al. (2022) چهار جنبه توجه، درک، درستی، و قابلیت اطمینان را بررسی کردند. همچنین، Syed et al. (2023) دسترسی، دقت، کامل بودن، سازگاری، اعتبار متنی، و ارزش را به‌عنوان چارچوب کیفیت داده بررسی کردند

مرور منابع مختلف داخلی و خارجی نشان دادند که تاکنون پژوهشی به‌منظور شناسایی شاخص‌های کیفیت داده برای حکمرانی داده صورت نگرفته است. یافته‌های این پژوهش به‌منظور استقرار حاکمیت داده در سازمان‌های داده‌محور مناسب خواهد بود.

٦. روش پژوهش

پژوهش حاضر به لحاظ هدف، از نوع کاربردی است و از نظر روش، ماهیت توصیفی-پیمایشی دارد. به‌منظور دستیابی به هدف پژوهش، در گام اول با مطالعه منابع کتابخانه‌ای و ارزیابی نتایج مطالعات مرتبط با روش فراترکیب، ابعاد کیفیت داده در این مطالعات بررسی و شناسایی شد. در گام بعدی نیز با استفاده از پرسشنامه و بهره‌گیری از روش دلفی فازی، ابعاد مهم کیفیت داده از دیدگاه متخصصان داده (اعضای پنل دلفی) شناسایی شدند. جامعه آماری این پژوهش به تعداد ۲۱ نفر از پژوهشگران حوزه کیفیت داده بودند که حداقل دارای یک مقاله پژوهشی در این موضوع بودند. پرسشنامه در دو دور دلفی به ایمیل این پژوهشگران ارسال شد و سرانجام، تعداد ۱۴ پرسشنامهٔ تکمیل‌شده برگردانیده شد. در دور اول دلفی تعداد ۷ مورد از ابعاد کیفیت داده که از نظر خبرگان مناسب نبودند، حذف و در دور دوم هم تعداد ۲ مورد دیگر از ابعاد حذف شدند. در مرحله دوم اجماع نظر خبرگان به‌دست آمد. برای تجمیع دیدگاه خبرگان از روش میانگین فازی استفاده شد

فراترکیب

در این پژوهش از روش هفت-مرحله‌ای فراترکیب «سندلوسکی و باروسو»[1] که بیشترین استفاده را در روش فراترکیب دارد، استفاده شد. در ادامه، هفت گام این روش آمده است.

گام اول: تنظیم سؤال پژوهش

چه چیزی: اولین گام در فراترکیب مشخص کردن چه چیزی است. به‌منظور مفهوم‌سازی

1. Sandelowski and Barroso

در این پژوهش، سؤال اصلی عبارت از این است که شاخص‌های کیفیت داده برای حکمرانی داده کدام‌اند؟

گام دوم: بررسی نظام‌مند متون

طبق مدل «ساندلوسکی و باروسو»، کلیدواژه‌هایی برای جست‌وجوی نظام‌مند تعریف شد. کلیدواژه‌های جست‌وجوشده در حوزه منابع انگلیسی حکمرانی داده، کیفیت داده، مدیریت کیفیت داده، و ارزیابی کیفیت داده بودند. در این پژوهش پایگاه‌های داده، نشریه‌ها و موتورهای جست‌وجوی مختلفی بین سال‌های ۱۹۹۵ تا سال ۲۰۲۲ بررسی شد. در جست‌وجو و بررسی پایگاه‌های داده، نشریه‌ها و موتورهای جست‌وجوی مختلف و با استفاده از واژه‌های کلیدی مورد نظر، ۲۶۸ منبع یافت شد

گام سوم: انتخاب منابع مناسب

در ابتدای فرایند جست‌وجو باید مشخص کنیم که آیا مقالات یافت‌شده مناسب با سؤال پژوهش است یا خیر؟ به‌منظور رسیدن به هدف، مقالاتِ منتخب چندین بار بازبینی شده و در هر بازبینی تعدادی از آنها حذف شدند. در بازبینی، پارامترهای مختلفی مانند عنوان، چکیده، محتوا، روش و جزئیاتی از قبیل نام نویسنده، شماره و سال انتشار و منابع در نظر گرفته شده و بر اساس آن‌ها مقالات نامرتبط حذف شدند. در این پژوهش پس از چندین بار بررسی و بازبینی تعداد ۶۲ سند مناسب و مرتبط برای بررسی نهایی باقی ماند

گام چهارم: استخراج نتایج از متون

در این مرحله اطلاعات متون به‌صورت نظام‌مند آورده شد؛ به این صورت که اطلاعات مربوط به مقالات در یک جدول شامل سه ستون درج گردید. در ستون اول نام خانوادگی نویسنده و سال انتشار، در ستون بعدی عنوان و در ستون آخر شاخص‌ها و مؤلفه‌های استخراج‌شده مربوط به کیفیت داده آورده شد.

گام پنجم: تجزیه‌وتحلیل یافته‌ها

گام پنجم فراترکیب را، بررسی موضوعی یاد می‌کنند؛ به‌طوری که در این مرحله موضوعات شناسایی‌شده در یک طبقه‌بندی قرار می‌گیرند. به همین منظور، ابتدا تمام عوامل شناسایی‌شده به‌عنوان یک کد در نظر گرفته شده و سپس، با در نظر گرفتن

مفاهیم هر یک از مؤلفه‌ها، آن‌ها در یک مفهوم مشابه دسته‌بندی و به این ترتیب، مفاهیم پژوهش مشخص شدند.

گام ششم: کنترل کدهای استخراجی

در این مرحله برای کنترل مفاهیم استخراجی از مقایسه نظر پژوهشگر با یک خبره استفاده شد؛ به این صورت که مفاهیم و مؤلفه‌های استخراجی با نظر یک متخصص داده مقایسه و نهایی شدند.

گام هفتم: ارائه یافته‌ها

در گام آخر، بر اساس بازنگری عمیق پژوهش‌های پیشین و بررسی مجدد کدهای استخراج‌شده، ترکیب یافته‌ها و مؤلفه‌های کیفیت داده ارائه شد. جدول ۱، شاخص‌های حاصل از فراترکیب را نشان می‌دهد

جدول ۱. شاخص‌های کیفیت داده برای حکمرانی داده حاصل از روش فراترکیب

ردیف	مؤلفه	تعریف	منبع
۱	صحت (accuracy)	میزان نزدیک بودن مقادیر داده به مقادیر واقعی	CDDQ 2019; Eurostat 2015; Wang & Strong 1996; English 1999; DAMA-UK 2013; DAMA 2017; Eurostat 2015; Redman & Blanton 1996; Wagner & et al. 2020; Batini et al. 2009; Bharati, Chaudhury 2004; Bizer & Cyganiak 2009; Chen 2010; Chen & Tseng 2011; Chung 2006; Du & Zhou 2012; Freitas et al. 2011; Haider & Lee, 2012; Hazen et al. 2014; Jin et al. 2009; Kim & Han 2011; Kim et al. 2009; Lederer et al. 2000; Lee & Haider 2011; Lee, Kim & Kim 2007; Lee & Chung 2009; Lee et al. 2002; Li et al. 2005; Lin 2006; Michnik & Lo 2009; Negash, Ryan & Igbaria 2003; Sebastian-Coleman 2012; Yeganeh, Sadiq & Sharaf 2014; Serhani et al. 2016; Taleb & Serhani 2017; Ardagna et al. 2018; CDDQ 2019; English 1999; Van Nederpelt 2009; Redman & Blanton 1996; ISO 25012; Wagner et al. 2020; Serhani et al. 2016

ردیف	مؤلفه	تعریف	منبع
۲	به‌موقع بودن یا زمان‌مند بودن (timeliness)	داده‌ها در کوتاه‌ترین و سریع‌ترین زمان در دسترس استفاده‌کنندگان باشند.	CDDQ 2019; Eurostat 2015; Wang, & Strong 1996; English 1999; DAMA-UK 2013; DAMA 2017; Redman & Blanton 1996; Wagner & et al 2020; Batini et al. 2009; Bharati & Chaudhury 2004; Bizer & Cyganiak 2009; Chen 2010; Chen & Tseng 2011; Chien & Tsaur 2007; Chung 2006; Freitas et al. 2011; Haider & Lee 2012; Hazen et al. 2014; Ifinedo et al. 2010; Jin et al. 2009; Kim & Han 2011; Kim et al. 2009; Lederer et al., 2000; Lee, Kim & Kim 2007; Lee & Chung 2009; Lee et al. 2002; Lee & Haider 2011; Li et al. 2005; Michnik & Lo 2009; Negash, Ryan & Igbaria 2003; Sebastian-Coleman 2012; Yeganeh, Sadiq & Sharaf 2014; Serhani et al. 2016; Taleb & Serhani, 2017; Eurostat 2014; Ardagna et al. 2018
۳	کامل بودن (completeness)	درجه‌ای که همه ویژگی‌های مورد نیاز در مجموعه داده وجود دارد.	ISO 25012; Batini et al. 2009; Bharati & Chaudhury 2004; Bizer & Cyganiak 2009; Chen 2010; Chen & Tseng 2011; Chien & Tsaur 2007; Chung 2006; Du & Zhou 2012; Freitas et al. 2011; Goel & Chengalur-Smith 2010; Haider & Lee 2012; Hazen et al. 2014; Jin et al. 2009; Kim & Han 2011; Kwon, Lee & Shin 2014; Lee, Chung, 2009; Lee et al. 2002; Lee & Haider 2011; Michnik & Lo 2009; Negash et al. 2003; Sebastian-Coleman 2012; Sung & You 2007; Yeganeh, Sadiq & Sharaf 2014; Serhani et al. 2016; Taleb & Serhani 2017
۴	دسترس‌پذیری (accessibility)	سهولت بررسی و بازیابی داده	CDDQ 2019; Eurostat 2015; Wang 1996; English 1999; ISO 25012; Chen & Tseng 2011; Chien & Tsaur 2007; Chung 2006; Freitas et al. 2011; Goel & Chengalur-Smith 2010; Haider & Lee 2012; Hazen et al. 2014; Lee & Haider 2011; Lee et al. 2002; Lin 2006; Michnik & Lo 2009; Salaün & Flores 2001; Yeganeh, Sadiq & Sharaf 2014; Redman & Blanton 1996

ردیف	مؤلفه	تعریف	منبع
۵	ثبات (consistency)	میزان رعایت مقادیر داده در ویژگی‌ها که در یک رکورد، در یک فایل داده، بین فایل‌های داده، در یک رکورد در نقاط مختلف زمانی با یک قاعده مطابقت داشته باشد	CDDQ 2019; Eurostat 2015; Wang & Strong 1996; English 1999; DAMA-UK 2013; DAMA 2017; Redman & Blanton 1996; Wagner et al. 2020; Batini et al. 2009; Bizer & Cyganiak 2009; Chen & Tseng 2011; Chien & Tsaur 2007; Chung 2006; Du & Zhou 2012; Freitas et al. 2011; Haider & Lee 2012; Hazen et al. 2014; Kwon et al. 2014; Lee & Haider 2011; Michnik & Lo 2009; Sebastian-Coleman 2012; Serhani et al. 2016; Taleb & Serhani 2017; Ardagna et al. 2018
۶	ربط (relevance)	درجه‌ای که ترکیب مجموعه داده‌ها، نیازهای مصرف‌کننده داده را برآورده می‌کند.	Eurostat 2015; Wang & Strong 1996; Redman & Blanton 1996; Bharati & Chaudhury 2004; Bizer & Cyganiak 2009; Boritz 2005; Chen 2010; Chen & Tseng 2011; Chien & Tsaur 2007; Chung 2006; Goel & Chengalur-Smith 2010; Haider & Lee 2012; Ifinedo et al. 2010; Jin et al., 2009; Kim et al. 2009; Lederer et al. 2000;. Lee & Chung 2009; Lee & Haider 2011; Michnik & Lo 2009; Negash et al. 2003; Salaün & Flores 2001; Song & Zahedi 2007; Sung & You 2007
۷	کاربردپذیری (قابلیت استفاده) (usability)	میزان اثربخشی داده برای استفاده‌کنندگان داده	English 1999; ISO 25012; DAMA-UK 2013; Redman & Blanton 1999; Boritz 2005; Chen & Tseng 2011; Chung 2006; Freitas et al. 2011; Haider & Lee 2012; Hsieh et al. 2010; Jin et al. 2009; Lederer et al. 2000; Lee et al. 2002; Michnik & Lo 2009; Song & Zahedi 2007
۸	اعتبار (credibility)	درجه‌ای که ارزش داده‌ها توسط مصرف‌کنندگان، به‌عنوان داده‌های واقعی و باورپذیر تلقی شود.	ISO 25012; DAMA 2017; DAMA-UK 2013; English 1999; Wagner & et al 2020; Bizer & Cyganiak 2009; Boritz 2005; Chen 2010; Haider& Lee 2012; Hsieh et al. 2010; Kim et al. 2009; Lin 2006; Salaün & Flores 2001; Song & Zahedi 2007

ردیف	مؤلفه	تعریف	منبع
۹	قابلیت درک (understandability)	درجه‌ای که داده به‌سادگی قابل درک است.	Chen & Tseng 2011; Chung 2006; Haider & Lee 2012; Hsieh et al. 2010; Ifinedo et al. 2010; Lee et al. 2002; Lee & Haider 2011; Salaün & Flores 2001; Chen 2010; Song & Zahedi 2007; Sung & You 2007
۱۰	عینیت (objectivity)	میزان ایجاد مقادیر داده به شیوه‌ای بی‌طرفانه	Wang 1996; Bizer & Cyganiak 2009; Chen & Tseng 2011; Chung 2006; Freitas et al. 2011; Haider & Lee 2012; Lee et al. 2002; Lee & Haider 2011; Michnik & Lo 2009
۱۱	مقبولیت (plausibility)	میزان مطابقت ارزش داده‌ها با دانش دنیای واقعی	Bizer & Cyganiak 2009; Chen & Tseng 2011; Chung 2006; Hazen et al. 2014; Lee et al. 2002; Lee & Haider 2011; Li et al. 2005; Michnik & Lo 2009
۱۲	قابلیت نمایش مقادیر پوچ (ability to represent null values)	درجه‌ای که یک قالب اجازه می‌دهد تا مقادیر پوچ در یک ویژگی مشخص شود.	Redman & Blanton 1996; Chen 2010; Chen & Tseng 2011; Chung 2006; Goel & Chengalur-Smith 2010; Kim & Han 2011; Lee et al. 2007; Lee & Haider 2011
۱۳	امنیت دسترسی (access security)	میزان دسترسی به مجموعه‌های داده محدود شده است.	Wang 1996; Chen & Tseng 2011; Freitas et al. 2011; Haider & Lee 2012; Lee et al. 2002; Lee & Haider 2011; Michnik & Lo 2009; Sung & You 2007
۱۴	شهرت (اعتبار) (reputation)	درجه‌ای که داده‌ها از نظر منبع یا محتوای خود مورد اعتماد هستند یا بسیار مورد توجه قرار می‌گیرند.	Wang 1996; Chen & Tseng 2011; Chung 2006; Freitas et al. 2011; Haider & Lee 2012; Lee et al. 2002; Lee & Haider 2011; Michnik & Lo 2009
۱۵	تفسیرپذیری (interpretability)	میزان استفاده از داده‌ها به زبان‌ها و واحدهای اندازه‌گیری مناسب	Redman & Blanton 1996; Wang 1996; Chen 2010; Chung 2006; Freitas et al. 2011; Haider & Lee 2012; Lee et al. 2002; Lee & Haider 2011; Michnik & Lo 2009
۱۶	موجز بودن (to be brief)	درجه‌ای که داده تلخیص شده است.	Bizer & Cyganiak 2009; Chen 2010; Chen & Tseng 2011; Chung 2006; Freitas et al. 2011; Haider & Lee 2012; Lee et al. 2002; Michnik & Lo 2009

ردیف	مؤلفه	تعریف	منبع
۱۷	مقایسه‌پذیری(comparability)	درجه‌ای که ارزش دادههایی که دو جمعیت را نشان می‌دهند دارای تعریف یکسانی هستند و به یک روش اندازه‌گیری می‌شوند.	Eurostat 2014; CDDQ 2019; Eurostat 2015; Wang 1996; English 1999; DAMA-UK 2013; DAMA 2017; Eurostat 2015; Redman & Blanton 1996; Wagner & et al. 2020
۱۸	یکتا بودن (uniqueness)	میزان وقوع اجسام (در دنیای واقعی) فقط یک‌بار به‌عنوان یک رکورد در یک فایل داده	DAMA 2017; DAMA-UK 2013; English 1999; Wagner & et al. 2020
۱۹	مطابقت با قوانین و استانداردها (vompliance with laws, regulations, or standards)	میزان مطابقت داده‌ها با قوانین، مقررات و استانداردها	Wang 1996
۲۰	محرمانه بودن (confidentiality)	میزان افشای داده‌ها باید به مصرف‌کنندگان مجاز داده محدود شود.	DAMA-UK 2013; Eurostat 2014; ISO 25012
۲۱	پیوندپذیری (Linkability)	درجه‌ای که می‌توان رکوردهای (موجودیت‌های) یک فایل داده را به‌درستی با رکوردهای یک فایل داده دیگر مرتبط کرد.	Daas 2010; Freitas et al. 2011; Hsieh, Kuo & Yang, Lin 2010
۲۲	رایج و شایع بودن (currency)	میزان به روز بودن ارزش داده‌ها	DAMA 2017; CDDQ 2019; Redman 1996; ISO 25012
۲۳	وضوح (clarity)	سهولتی که مصرف‌کنندگان داده می‌توانند فراداده یک مجموعه را درک کنند.	Eurostat 2014; Redman 1996; English 1999

ردیف	مؤلفه	تعریف	منبع
۲۴	ارزش (value)	میزان استفاده از داده‌ها و مزایای آن‌ها	DAMA-UK 2013; Wang 1996
۲۵	تمامیت (integrity)	میزان عدم از دست دادن ارزش داده	DAMA 2017; CDDQ 2019; English 1999
۲۶	معادل‌سازی (equivalence)	میزان برابر بودن ویژگی‌های ذخیره‌شده در مجموعه داده‌های متعدد از نظر مفهومی	English 1999
۲۷	جداگانه بودن (granularity)	درجه‌ای که یک ویژگی واحد در صفات تقسیم می‌شود.	Redman & Blanton 1996
۲۸	تأخیر (latency)	فاصله زمانی بین زمان ایجاد داده‌ها و زمانی که برای استفاده در دسترس هستند.	DAMA 2017
۲۹	طبیعی بودن (naturalness)	میزان همسویی ترکیب مجموعه داده‌ها با اشیای دنیای واقعی که آن‌ها نشان می‌دهند.	Redman & Blanton 1996
۳۰	انسجام (coherence)	میزان ترکیب مجموعه داده‌ها	Eurostat 2015
۳۱	قابلیت اطمینان (روایی) (reliability)	میزان نزدیک بودن مقدار داده اولیه به مقدار داده بعدی.	Eurostat 2015
۳۲	قابلیت حمل (portability)	میزان نصب، جایگزینی یا انتقال داده‌ها از یک سیستم به سیستم دیگر با حفظ کیفیت موجود	Redman & Blanton 1996; ISO 25012

ردیف	مؤلفه	تعریف	منبع
۳۳	منطقی بودن (reasonability)	میزان مطابقت الگوی داده با انتظارات	DAMA 2017
۳۴	قابلیت بازیابی (recoverability)	میزان حفظ مجموعه داده‌ها در صورت بروز حادثه	ISO 25012
۳۵	افزونگی (قابلیت اضافه کردن) (redundancy)	میزان اضافه و ذخیره داده‌های منطقی یکسان بیش از یک بار	Redman 1996; Early 2011
۳۶	یکپارچگی مرجع (referential integrity)	داده‌ها به‌صورت یکنواخت ذخیره و استفاده می‌شوند.	Ardagna et al. 2018
۳۷	انطباق فراداده (metadata compliance)	میزان مطابقت ارزش داده‌ها با تعریف، مشخصات قالب و دامنه ارزش	Daas 2010; English 1999
۳۸	تکرارپذیری (reproducibility)	میزان بازآفرینی یک مجموعه داده با مقادیر داده یکسان	CDDQ 2019
۳۹	دوره نگهداری (retention period)	دوره‌ای که مجموعه داده‌ها در دسترس هستند تا زمانی که می‌توانند یا باید حذف شوند.	English 1999
۴۰	قابلیت ردیابی (traceability)	میزان دسترسی به اصل داده‌ها	Wang 1996; ISO 25012
۴۱	تنوع (variety)	میزان در دسترس بودن داده‌ها از منابع داده‌ای مختلف	Wang 1996
۴۲	نوسان (volatility)	میزان تغییر مقادیر داده در طول زمان	DAMA 2017
۴۳	تناسب (appropriateness)	میزان مناسب بودن قالب برای استفاده	Wang 1996; Redman & Blanton 1996;

روش دلفی

تکنیک دلفی یکی از روش‌های کسب دانش گروهی است که بیش از نیم قرن مورد استفاده بوده است. اساس محاسبات تکنیک دلفی بر دیدگاه خبرگان استوار است. بنابراین، هرگونه خطا و ناسازگاری در سنجش دیدگاه خبرگان، نتیجه نهایی به‌دست‌آمده از محاسبات را تحت تأثیر قرار می‌دهد. در رویکردهای دلفی سنتی، اگرچه از شایستگی‌ها و توانایی‌های ذهنی خبرگان برای انجام مقایسات استفاده می‌شود، اما کمی کردن دیدگاه خبرگان به روش سنتی، امکان انعکاس کامل سبک تفکر انسانی را ندارد (Habibi, Firouzi & Sarafrazi 2015)

تکنیک دلفی فازی

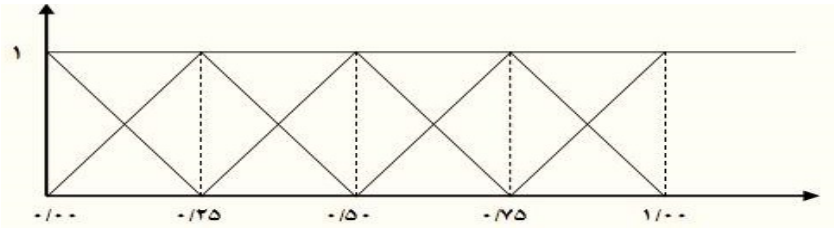
برای تعیین میزان اهمیت شاخص‌ها و غربال مهم‌ترین شاخص‌های شناسایی‌شده می‌توان از تکنیک دلفی با رویکرد فازی استفاده کرد. یکی از عمده‌ترین مزیت‌های تکنیک دلفی فازی نسبت به تکنیک دلفی سنتی برای غربال شاخص‌ها آن است که می‌توان از یک دوره برای تلخیص و غربال آیتم‌ها استفاده کرد. الگوریتم اجرای تکنیک دلفی فازی شامل گام‌های زیر است

- ◇ شناسایی طیف مناسب برای فازی‌سازی عبارات کلامی؛
- ◇ تجمیع فازی مقادیر فازی‌شده؛
- ◇ فازی‌زدایی مقادیر؛
- ◇ انتخاب شدت آستانه و غربال معیارها.

در الگوریتم اجرای تکنیک دلفی فازی برای غربالگری، نخست باید طیف فازی مناسبی برای فازی‌سازی عبارات کلامی پاسخ‌دهندگان توسعه داد. برای این منظور می‌توان از روش‌های توسعه طیف فازی استفاده کرد یا از طیف‌های فازی متداول برای این منظور استفاده کرد (Habibi, Firouzi & Sarafrazi 2015). برای نمونه طیف فازی مثلثی برای مقیاس پنج درجه‌ای لیکرت در بیان اهمیت شاخص‌ها به‌صورت زیر است

جدول ۲. اعداد فازی مثلثی معادل طیف لیکرت پنج درجه‌ای

خیلی بااهمیت	بااهمیت	متوسط	بی‌اهمیت	خیلی بی‌اهمیت
(۱، ۱، ۰/۷۵)	(۱، ۰/۷۵، ۰/۵)	(۰/۷۵، ۰/۵، ۰/۲۵)	(۰/۵، ۰/۲۵، ۰)	(۰/۲۵، ۰، ۰)



شکل ۱. اعداد فازی مثلثی معادل طیف لیکرت پنج درجه‌ای

پس از انتخاب یا توسعه طیف فازی مناسب، دیدگاه خبرگان گردآوری و به‌صورت فازی ثبت می‌شود. در گام دوم باید به تجمیع دیدگاه خبرگان پرداخت. راه‌های متعددی برای تجمیع فازی دیدگاه خبرگان پیشنهاد شده است. اگر دیدگاه هر کارشناس به‌صورت عدد فازی مثلثی (l, m, u) نمایش داده شود، ساده‌ترین روش محاسبه میانگین فازی دیدگاه خبرگان چنین است

$$F_{AVE} = \frac{\sum l}{n}, \frac{\sum m}{n}, \frac{\sum u}{n}$$

به‌جای استفاده از میانگین فازی، روش‌های متنوع دیگری نیز برای تجمیع[1] دیدگاه خبرگان به‌کار می‌روند. این روش‌های تجمیع در واقع، روش‌هایی تجربی هستند که توسط پژوهشگران مختلف ارائه شده‌اند. برای نمونه، یک روش مرسوم برای تجمیع مجموعه‌ای از اعداد فازی مثلثی را کمینه l و میانگین m و بیشینه u در نظر گرفته‌اند (Hsu, Lee & Kreng 2010)

از رابطه ساده $\frac{l+m+u}{3}$ برای فازی‌زدایی و قطعی‌سازی میانگین دیدگاه‌ها سود برده شده است. افزون بر این، آستانه تحمل نیز مقدار ۰/۶ در نظر گرفته شده است. نتایج حاصل از مراحل فوق به‌صورت زیر خلاصه شده است:

جدول ۳. نتایج فازی‌زدایی مقادیر تجمیع‌شده خبرگان

معیار	میانگین دیدگاه	مقدار قطعی	نتیجه
کامل بودن	(۰/۸۳۹، ۰/۷۱۴، ۰/۵۰)	۰/۶۸۴	پذیرش
معادل‌سازی	(۱، ۰/۹۸۲، ۰/۶۷۸)	۰/۸۸۶	پذیرش
جداگانه بودن	(۰/۸۰۳، ۰/۵۷۱، ۰/۳۲۱)	۰/۵۶۵	رد

1. aggregation

معیار	میانگین دیدگاه	مقدار قطعی	نتیجه
زمان‌بندی	(۰/۸۷۵، ۰/۶۷۸، ۰/۴۶۶)	۰/۶۷۳	پذیرش
انسجام	(۰/۹۲۸، ۰/۷۵۰، ۰/۵۰)	۰/۷۲۶	پذیرش
طبیعی بودن	(۰/۶۴۲، ۰/۵۸۹، ۰/۲۶۷)	۰/۴۹۹	رد
ارتباط	(۰/۹۲۸، ۰/۷۶۷، ۰/۵۱۷)	۰/۸۷۰	پذیرش
دسترس‌پذیری	(۰/۹۸۲، ۰/۸۳۹، ۰/۵۸۹)	۰/۸۰۳	پذیرش
مطابقت با قوانین و استانداردها	(۰/۹۲۸، ۰/۷۸۵، ۰/۵۳۸)	۰/۷۵۰	پذیرش
محرمانه بودن	(۰/۷۸۵، ۰/۵۸۹، ۰/۴۹۲)	۰/۶۲۲	پذیرش
تفسیرپذیری	(۰/۸۲۱، ۰/۶۶۰، ۰/۴۴۶)	۰/۶۴۲	پذیرش
تأخیر	(۰/۶۲۵، ۰/۳۹۲، ۰/۲۱۴)	۰/۴۱۰	رد
افزونگی(قابلیت اضافه کردن)	(۰/۸۹۲، ۰/۷۳۱، ۰/۵۰)	۰/۷۰۷	پذیرش
شهرت و اعتمادپذیری	(۰/۹۱۰، ۰/۶۶۰، ۰/۴۱۰)	۰/۶۶۰	پذیرش
قابلیت ردیابی	(۰/۹۲۸، ۰/۸۰۳، ۰/۵۵۳)	۰/۷۶۱	پذیرش
ارزش	(۰/۹۴۶، ۰/۷۳۱، ۰/۵۵۳)	۰/۷۴۳	پذیرش
تنوع	(۰/۶۶۰، ۰/۴۱۰، ۰/۲۱۴)	۰/۴۲۸	رد
پیوندپذیری	(۰/۹۲۸، ۰/۸۳۹، ۰/۵۸۹)	۰/۷۸۵	پذیرش
یکپارچگی مرجع	(۰/۹۴۶، ۰/۸۲۱، ۰/۵۷۱)	۰/۷۷۹	پذیرش
یکتایی	(۱، ۰/۹۴۶، ۰/۷۳۱)	۰/۸۹۲	پذیرش
عینیت	(۰/۹۸۲، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۱۵	پذیرش
صحت	(۱، ۰/۸۲۱، ۰/۵۷۱)	۰/۷۹۷	پذیرش
مقایسه‌پذیری	(۰/۹۴۶، ۰/۷۵۰، ۰/۵۰)	۰/۷۳۲	پذیرش
ثبات	(۱، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۲۱	پذیرش
رایج و شایع بودن	(۰/۹۸۲، ۰/۸۹۲، ۰/۶۴۲)	۰/۸۳۸	پذیرش
تمامیت	(۰/۹۴۶، ۰/۷۵۰، ۰/۵۰)	۰/۷۲۶	پذیرش
انطباق فراداده	(۰/۹۱۰، ۰/۷۵۰، ۰/۴۸۲)	۰/۷۱۴	پذیرش
مقبولیت	(۰/۹۴۶، ۰/۷۸۵، ۰/۵۷۱)	۰/۷۶۷	پذیرش
اعتبار	(۰/۷۸۵، ۰/۵۸۹، ۰/۴۹۲)	۰/۶۲۲	پذیرش
نوسان	(۰/۶۶۰، ۰/۴۱۰، ۰/۲۱۴)	۰/۴۲۸	رد

معیار	میانگین دیدگاه	مقدار قطعی	نتیجه
موجز بودن	(۰/۹۲۸، ۰/۷۶۷، ۰/۵۳۸)	۰/۷۴۴	پذیرش
کاربردپذیری (قابلیت استفاده)	(۰/۶۹۶، ۰/۴۸۲، ۰/۵۸۶)	۰/۷۸۸	پذیرش
قابلیت اطمینان (روایی)	(۰/۸۷۵، ۰/۷۳۱، ۰/۵۰)	۰/۷۰۲	پذیرش
قابلیت درک	(۰/۹۶۴، ۰/۸۵۷، ۰/۶۲۵)	۰/۸۱۵	پذیرش
منطقی بودن	(۱، ۰/۹۴۶، ۰/۶۹۶)	۰/۸۸۰	پذیرش
امنیت	(۱، ۰/۸۷۵، ۰/۶۷۸)	۰/۸۵۱	پذیرش
قابلیت بازیابی	(۰/۹۴۶، ۰/۸۳۹، ۰/۶۲۵)	۰/۸۰۳	پذیرش
تکرارپذیری	(۱، ۰/۷۸۵، ۰/۵۸۹)	۰/۷۹۱	پذیرش
دوره نگهداری	(۰/۶۴۲، ۰/۴۲۸، ۰/۲۵۰)	۰/۴۴	رد
قابلیت نمایش مقادیر پوچ	(۰/۸۹۲، ۰/۷۶۷، ۰/۵۳۸)	۰/۷۳۲	پذیرش
تناسب	(۰/۹۸۲، ۰/۹۱۰، ۰/۶۶۰)	۰/۸۵۰	پذیرش
قابلیت حمل	(۰/۸۰۳، ۰/۶۲۵، ۰/۳۵۷)	۰/۵۹۵	رد
وضوح	(۰/۸۳۹، ۰/۶۶۰، ۰/۴۱۰)	۰/۸۰۳	پذیرش

با استناد به نتایج فازی‌زدایی مقادیر تجمیع‌شده خبرگان، آستانه تحمل معیارهای «استناد به نتایج فازی‌زدایی مقادیر تجمیع‌شده خبرگان، جداگانه بون، طبیعی بودن، تأخیر، تنوع، نوسان، دوره نگهداری، قابلیت حمل)، از مقدار در نظر گرفته‌شده ۰/۶ پایین‌تر است. لذا، این معیارها از مجموع معیارهای کیفیت داده دیدگاه خبرگان کنار گذاشته می‌شود؛ چرا که از دیدگاه کارشناسان این حوزه، معیارهای ذکرشده نقش تعیین‌کننده‌ای ندارند. نکته مهم در این راستا این است که در کاربرد تکنیک دلفی برای غربالگری، با فازی‌زدایی مقادیر و انتخاب شدت آستانه در همان مرحله نخست می‌توان به نتیجه دست یافت. اما برای کاربرد دلفی در پیش‌بینی دیدگاه هر کارشناس با میانگین دیدگاه‌ها محاسبه شده و چرخه دلفی تکرار می‌شود. این فرایند زمانی که دو میانگین متوالی ... $A_{ave}, B_{ave}, C_{ave}$، به‌صورتی روشن به هم نزدیک شوند، متوقف می‌شود (Habibi, Firouzi & Sarafrazi 2015). با اقتباس از این دیدگاه، می‌توان گفت که در پژوهش حاضر، تا حدود زیادی میانگین‌های متوالی به هم نزدیک هستند و در همین مرحلهٔ نخست، نتیجه مورد نظر حاصل شده است. گواه این ادعا را نیز می‌توان کم بودن معیارهای حذف‌شده در نظر گرفت.

مرحله دوم دلفی فازی

در این مرحله، میزان اختلاف نظر هر خبره با میانگین نظرات اعضای پانل خبرگان با استفاده از رابطه ۳، محاسبه شد. سپس، پرسشنامه دیگری به همراه نظر قبلی هر خبره و میزان اختلاف نظر وی با میانگین نظرات اعضای پانل در اختیار آن‌ها قرار گرفت. با توجه به نظرات ارائه‌شده در مرحله اول و مقایسه آن با نتایج مرحله دوم با استفاده از رابطه، چنانچه اختلاف نظر خبرگان دو مرحله از حد آستانه ۲ کمتر باشد، فرایند نظرسنجی متوقف می‌شود

جدول ٤. نتایج فازی‌زدایی مقادیر تجمیع‌شده خبرگان

معیار	میانگین دیدگاه	مقدار قطعی	نتیجه
کامل بودن	(۰/۹۸۲، ۰/۹۱۰، ۰/۶۶۰)	۰/۸۵۰	پذیرش
معادل‌سازی	(۰/۶۴۲، ۰/۴۱۰، ۰/۱۹۶)	۰/۴۱۶	رد
زمان‌بندی	(۰/۹۸۲، ۰/۶۷۸، ۰/۵۳۸)	۰/۷۳۲	پذیرش
انسجام	(۰/۸۰۳، ۰/۵۷۱، ۰/۳۲۱)	۰/۵۶۵	رد
ارتباط	(۰/۸۵۷، ۰/۷۱۴، ۰/۴۶۶)	۰/۶۷۹	پذیرش
دسترس‌پذیری	(۰/۹۴۶، ۰/۸۲۱، ۰/۵۷۱)	۰/۷۷۹	پذیرش
مطابقت با قوانین و استانداردها	(۰/۹۲۸، ۰/۷۶۷، ۰/۵۱۷)	۰/۷۳۷	پذیرش
محرمانه بودن	(۰/۸۷۵، ۰/۶۷۸، ۰/۴۲۸)	۰/۶۶۰	پذیرش
تفسیرپذیری	(۰/۸۰۳، ۰/۵۵۳، ۰/۴۴۶)	۰/۶۰۰	پذیرش
افزونگی(قابلیت اضافه کردن)	(۰/۸۹۲، ۰/۷۴۱، ۰/۶۰۷)	۰/۷۴۶	پذیرش
شهرت و اعتمادپذیری	(۱، ۰/۷۸۵، ۰/۵۸۹)	۰/۷۹۱	پذیرش
قابلیت ردیابی	(۰/۸۷۵، ۰/۶۹۶، ۰/۴۴۶)	۰/۶۷۳	پذیرش
ارزش	(۰/۹۱۰، ۰/۷۴۱، ۰/۴۸۲)	۰/۷۱۱	پذیرش
پیوندپذیری	(۰/۹۱۰، ۰/۷۱۴، ۰/۴۸۲)	۰/۷۰۲	پذیرش
یکپارچگی مرجع	(۰/۹۱۰، ۰/۷۱۴، ۰/۴۶۶)	۰/۶۹۶	پذیرش
یکتایی	(۰/۸۹۲، ۰/۷۴۱، ۰/۶۲۵)	۰/۷۵۲	پذیرش
عینیت	(۰/۹۴۶، ۰/۷۸۵، ۰/۵۳۸)	۰/۷۵۶	پذیرش
صحت	(۰/۹۸۲، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۱۵	پذیرش

معیار	میانگین دیدگاه	مقدار قطعی	نتیجه
مقایسه‌پذیری	(۰/۸۷۵، ۰/۷۶۷، ۰/۵۱۷)	۰/۷۱۹	پذیرش
ثبات	(۱، ۰/۹۱۰، ۰/۶۶۰)	۰/۸۵۶	پذیرش
رایج و شایع بودن	(۰/۸۷۵، ۰/۶۴۲، ۰/۴۴۶)	۰/۶۵۴	پذیرش
تمامیت	(۰/۹۱۰، ۰/۷۱۴، ۰/۴۸۲)	۰/۷۰۲	پذیرش
انطباق فراداده	(۰/۹۱۰، ۰/۷۱۴، ۰/۴۶۶)	۰/۶۹۶	پذیرش
مقبولیت	(۱، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۲۱	پذیرش
اعتبار	(۰/۹۱۰، ۰/۷۱۴، ۰/۴۸۲)	۰/۷۰۲	پذیرش
موجز بودن	(۱، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۲۱	پذیرش
کاربردپذیری (قابلیت استفاده)	(۰/۹۶۴، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۰۹	پذیرش
قابلیت اطمینان (روایی)	(۰/۹۲۸، ۰/۷۵۰، ۰/۵۰)	۰/۷۲۶	پذیرش
قابلیت درک	(۱، ۰/۸۷۵، ۰/۶۲۵)	۰/۸۳۳	پذیرش
منطقی بودن	(۱، ۰/۸۷۵، ۰/۶۲۵)	۰/۸۳۳	پذیرش
امنیت	(۱، ۰/۸۵۷، ۰/۶۰۷)	۰/۸۲۱	پذیرش
قابلیت بازیابی	(۰/۸۰۳، ۰/۶۰۷، ۰/۴۱۰)	۰/۶۰۶	پذیرش
تکرارپذیری	(۰/۸۹۲، ۰/۸۲۱، ۰/۵۷۱)	۰/۷۶۱	پذیرش
قابلیت نمایش مقادیر پوچ	(۰/۸۷۵، ۰/۷۱۴، ۰/۴۶۶)	۰/۶۸۶	پذیرش
تناسب	(۱، ۰/۹۴۶، ۰/۶۹۶)	۰/۸۸۰	پذیرش
وضوح	(۰/۹۶۴، ۰/۷۸۵، ۰/۵۳۸)	۰/۷۶۲	پذیرش

با استناد به نتایج فازی‌زدایی مقادیر تجمیع‌شده خبرگان، آستانه تحمل معیارهای «تناسب»، از مقدار در نظر گرفته‌شده ۰/۶ پایین‌تر است. لذا، این معیارها از مجموع معیارهای بررسی مدل مدیریت کیفیت داده برای حکمرانی داده از دیدگاه خبرگان کنار گذاشته می‌شود؛ چرا که از دیدگاه کارشناسان این حوزه معیارهای ذکرشده نقش تعیین‌کننده‌ای ندارند.

۷. نتیجه‌گیری

نتایج روش فراترکیب

در این پژوهش شناسایی شاخص‌های کیفیت داده برای حکمرانی داده مورد سؤال بود. با

استفاده از روش فراترکیب، ۲۶۸ مقاله در پایگاه‌های ایرانداک، Science Direct، Google Scholar، Springer، IEEE و ACM یافت شد و طی چندین مرحله غربالگری، مقالات غیرمرتبط حذف شدند. سرانجام، تعداد ۶۲ مقاله باقی ماند که به‌طور دقیق مورد مطالعه و بررسی قرار گرفتند. آنچه که از متن مقالات منتخب استخراج گردید، به‌عنوان کد در نظر گرفته شدند و با طبقه‌بندی کدها در گروه‌های مشابه، ۸ مفهوم ویژگی مجموعه داده‌ای با ۳ مؤلفه، داده با ۱۲مؤلفه، فایل داده با ۵ مؤلفه، ارزش داده با ۱۳مؤلفه، الگوی داده با ۲ مؤلفه، مجموعه داده با ۳ مؤلفه، قالب‌بندی با ۲ مؤلفه و فراداده با ۶ مؤلفه شناسایی شدند.

نتایج روش دلفی‌فازی

در این بخش شاخص‌های شناسایی‌شده کیفیت داده برای حکمرانی داده استخراج‌شده از روش فراترکیب وارد فرایند دلفی‌فازی گردید. به این منظور، جدول شاخص‌ها در قالب پرسشنامه دلفی فازی در اختیار خبرگان قرار گرفت. در مرحله اول دلفی فازی پس از جمع‌آوری نظرات خبرگان و تجزیه‌وتحلیل آن‌ها در این مرحله تعداد ۷ مؤلفه حذف شدند. در مرحله دوم دلفی فازی، پرسشنامه جدیدِ حاصل از مرحله اول دلفی فازی در اختیار خبرگان قرار گرفت و در این مرحله تعداد ۲ مورد از مؤلفه‌ها حذف شدند. نظر به دیدگاه‌های ارائه‌شده در مرحله اول و مقایسه آن با نتایج مرحله دوم، در صورتی که اختلاف بین دو مرحله کمتر از حد آستانه (۰/۲) باشد، در این صورت فرایند نظرسنجی متوقف می‌شود که با توجه به نتایج جدول ۴، هیچ‌کدام از سؤالات کمتر از ۰/۲ نبود و هیچ سؤالی حذف نگردید

جدول ۵. میزان اختلاف دیدگاه خبرگان در نظرسنجی مرحله اول و دوم

سؤالات	مرحله اول	مرحله دوم	اختلاف
کامل بودن	۰/۶۸۴	۰/۸۵۰	۰/۱۶۶
زمان‌بندی	۰/۶۷۳	۰/۷۳۲	۰/۰۵۹
ربط	۰/۸۷۰	۰/۶۷۹	۰/۱۹۱
دسترس‌پذیری	۰/۸۰۳	۰/۷۷۹	۰/۰۲۴
مطابقت با قوانین و استانداردها	۰/۷۵۰	۰/۷۳۷	۰/۰۱۳

سؤالات	مرحله اول	مرحله دوم	اختلاف
محرمانه بودن	۰/۶۲۲	۰/۶۶۰	۰/۰۳۸
تفسیرپذیری	۰/۶۴۲	۰/۶۰۰	۰/۰۴۲
افزونگی(قابلیت اضافه کردن)	۰/۷۰۷	۰/۷۴۶	۰/۰۳۹
شهرت و اعتمادپذیری	۰/۶۶۰	۰/۷۹۱	۰/۱۳۱
قابلیت ردیابی	۰/۷۶۱	۰/۶۷۳	۰/۰۸۸
ارزش	۰/۷۴۳	۰/۷۱۱	۰/۰۳۲
پیوندپذیری	۰/۷۸۵	۰/۷۰۲	۰/۰۸۵
یکپارچگی مرجع	۰/۷۷۹	۰/۶۹۶	۰/۰۸۳
یکتایی	۰/۸۹۲	۰/۷۵۲	۰/۱۴
عینیت	۰/۸۱۵	۰/۷۵۶	۰/۰۵۹
صحت	۰/۷۹۷	۰/۸۱۵	۰/۰۱۸
مقایسه‌پذیری	۰/۷۳۲	۰/۷۱۹	۰/۰۱۳
ثبات	۰/۸۲۱	۰/۸۵۶	۰/۰۳۵
رایج و شایع بودن	۰/۸۳۸	۰/۶۵۴	۰/۱۸۴
تمامیت	۰/۷۲۶	۰/۷۰۲	۰/۰۲۴
انطباق فراداده	۰/۷۱۴	۰/۶۹۶	۰/۰۱۸
مقبولیت	۰/۷۶۷	۰/۸۲۱	۰/۰۵۴
اعتبار	۰/۶۲۲	۰/۷۰۲	۰/۰۸
موجز بودن	۰/۷۴۴	۰/۸۲۱	۰/۰۷۷
کاربردپذیری (قابلیت استفاده)	۰/۷۸۸	۰/۸۰۹	۰/۰۲۱
قابلیت اطمینان (روایی)	۰/۷۰۲	۰/۷۲۶	۰/۰۲۴
قابلیت درک	۰/۸۱۵	۰/۸۳۳	۰/۰۱۸
منطقی بودن	۰/۸۸۰	۰/۸۳۳	۰/۰۴۷
امنیت	۰/۸۵۱	۰/۸۲۱	۰/۰۳
قابلیت بازیابی	۰/۸۰۳	۰/۶۰۶	۰/۱۹۷
تکرارپذیری	۰/۷۹۱	۰/۷۶۱	۰/۰۳
قابلیت نمایش مقادیر پوچ	۰/۷۳۲	۰/۶۸۶	۰/۰۴۶

سؤالات	مرحله اول	مرحله دوم	اختلاف
تناسب	۰/۸۵۰	۰/۸۸۰	۰/۰۳
وضوح	۰/۸۰۳	۰/۷۶۲	۰/۰۴۱

۸. پیشنهادها

۴۴. شناسایی شاخص‌های بهبود کیفیت داده از دیدگاه کارشناسان داده در سازمان‌ها و کسب‌وکارهای داده‌محور؛

۵۴. شناسایی شاخص‌های مدیریت کیفیت داده در کلان‌داده‌ها؛

۴۶. امکان‌سنجی کاربرد شاخص‌های مدیریت کیفیت داده برای حکمرانی داده در فضای ابری؛

۴۷. امکان‌سنجی کاربرد شاخص‌های مدیریت کیفیت داده برای حکمرانی داده در هوش مصنوعی؛

۴۸. امکان‌سنجی کاربرد شاخص‌های استخراج‌شده پژوهش در کلان‌داده‌ها.

فهرست منابع

اخوان آملی، رامین. ۱۳۷۵. نقش اطلاعات در شناخت عوامل داخلی، خارجی و محیطی سیستم. *ماهنامه تدبیر* ۶۶: ۳۸-۴۱.

ارشادی، محمدجواد، جلال‌الدین نصیری، و فرهاد شیرانی. ۶۹۳۱. طراحی مدل کیفیت فراداده در سامانه ثبت پایان‌نامه، رساله‌های دانش‌آموختگان داخل و خارج کشور. طرح پژوهشی. وزارت علوم، تحقیقات و فناوری. پژوهشگاه علوم و فناوری اطلاعات ایران

اشتریان اصفهانی، آیناز، محمدجواد ارشادی، و امیر عزیزی. ۱۳۹۸. توسعه شاخص‌های کیفیت داده به‌منظور ارزیابی سامانه‌های اطلاعاتی تحقیقاتی: یک مطالعه موردی. *فصلنامه علمی-پژوهشی مدیریت استاندارد و کیفیت* ۹ (بهار): ۶۰-۷۴.

خسروانجم، داود، علی‌اصغر انواری رستمی، رسول چاوشینی، مسعود احمدزاده. ۱۳۹۲. توسعه مدل‌های AHP فازی برای ارزیابی تأثیر قابلیت‌های IT و ابعاد کیفیت داده‌ها. *فصلنامه مدیریت صنعتی دانشگاه آزاد اسلامی واحد سنندج* ۸ (۲۵): ۱۱۶-۱۰۵.

خلیلی جعفرآباد، احمد. ۱۳۹۶. بررسی تغییرات حوزه کیفیت داده با استفاده از تحلیل کلمات کلیدی. *دو-فصلنامه علمی-پژوهشی مدیریت اطلاعات* ۲ (۳): ۱۲۱-۱۳۸.

رحیمی، علی‌رضا . ۱۳۹۵. بررسی تحولات پژوهش‌های حوزه ارزیابی کیفیت داده‌ها و اطلاعات در نظام‌های اطلاعاتی از سال ۲۰۰۰ تا نیمه نخست سال ۲۰۱۵. *پژوهشنامه پردازش و مدیریت اطلاعات* http://Jipm.irandoc.ac.ir (دسترسی در ۱۳۹۶/۴/۰۸).

ردمن، توماس سی. ۱۳۸۱. داده چیست یا دادهها چه هستند؟ ترجمه محمدحسین دیانی. *کتابداری و اطلاع‌رسانی* ۵ (۲۰): ۸۱-۱۱۰.

رضاییان، علی. ۱۳۷۴. سیستم‌های اطلاعاتی مدیریت. *فصلنامه تحول اداری سازمان امور اداری و استخدامی کشور* ۱۰ و ۱۱: ۱۶-۲۶.

سهرابی، بابک، حمیدرضا یزدانی، محمدجواد ارشادی، و سوده دوروش. ۱۴۰۰. شناخت و تحلیل سیستمی متدولوژی‌های کیفیت داده و ارائه یک چارچوب جامع (با استفاده از روش فراترکیب). *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۶ (۳): ۷۳۷-۷۶۶.

صالحی، احمد، محمد محمد اقدسی، توکتم خطیبی، و مجید شیخ محمدی. ۱۴۰۲. ارائه یک چارچوب مفهومی برای پیش‌پردازش و بهبود کیفیت نگاره‌های رویداد در فرایندکاوی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۸ (۳): ۹۴۵-۹۷۹.

عبدالوند، ندا، و آتنا بوبه رژ. ۱۳۹۵. رویکردی سیستماتیک به چالش کیفیت داده در راهبرد مشتری‌محور در صنعت بانکداری. *فصلنامه علمی-پژوهشی تحقیقات بازاریابی نوین* ۱ (۶): ۱۷۷-۱۹۶.

غفاری، مسعود، عادل آذر، و اشکان شباک. ۱۳۹۷. نگاشت علّی مدیریت محصول آماری با رویکرد کیفیت داده. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳ (۳۳): ۱۰۴۱-۱۰۶۴.

محجوب، عباس، و حمیدرضا سیمه‌ساز. ۱۳۹۴. کیفیت داده‌ها پیشنیاز مدیریت منابع سازمان. *فصلنامه مدیریت استاندارد و کیفیت* ۵ (۱۷): ۵۳-۶۲.

نورزاد، عبدالرحمان. ۱۳۹۰. بهبود کیفیت داده‌ها در کامل بودن داده با استفاده از قوانین وابستگی. پایان‌نامه کارشناسی ارشد علوم کامپیوتر- مهندسی نرم‌افزار. دانشگاه پیام نور مرکز تهران. دانشکده فنی و مهندسی

References

Ardagna, Danilo, Cinzia Cappiello, Walter Samá, & Monica Vitali. 2018. Context-aware data quality assessment for big data. *Future Generation Computer Systems* 89: 548-562.

Batini, C., C. Cappiello, C. Francalanci, & A. Maurino. 2009. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)* 41 (3):16.

Batini, C. & M. Scannapiec. 2016. In *book: Data and Information Quality Dimensions, Principles and Techniques* (pp.403-419). Cham: Springer International Publishing.

Bharati, P., & A. Chaudhury. 2004. An empirical investigation of decision-making satisfaction in web-based decision support systems. *Decision support systems* 37 (2): 187-197.

Bizer, C., & R. Cyganiak. 2009. Quality-driven information filtering using the WIQA policy framework. *Journal of Web Semantics* 7 (1): 1-10.

Bonner, J. M. 2010. Customer interactivity and new product performance: Moderating effects of product newness and product embeddedness. *Industrial marketing management* 39 (3): 485-492.

Boritz, J. E. 2005. IS practitioners' views on core concepts of information integrity. *International Journal of Accounting Information Systems* 6 (4): 260-279.

Bovee, M., R. P. Srivastava, & B. Mak. 2003. A conceptual framework and belief-function approach to assessing overall information quality. *International journal of intelligent systems* 18 (1): 51-74.

CDDQ (List of Conformed Dimensions of Data Quality. 2019. Retrieved from https://dimensionsofdataquality.com/alldimensions (accessed Mar 5, 2022)

Chaffey, D., & G. White. 2010. *Business information management: improving performance using information systems*. Canada; Pearson Education

Chen, C. C., & Y. D. Tseng. 2011. Quality evaluation of product reviews using an information quality framework. *Decision Support Systems* 50 (4): 755-768.

Chen, C. W. 2010. Impact of quality antecedents on taxpayer satisfaction with online tax-filing systems—An empirical study. *Information & Management* 47 (5-6): 308-315.

Chengalur-Smith, I. N., D. P. Ballou, & H. L. Pazer. 1999. The impact of data quality information on decision making: an exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering* 11 (6): 853-864.

Chien, S. W., & S. M. Tsaur. 2007. Investigating the success of ERP systems: Case studies in three Taiwanese high-tech industries. *Computers in industry* 58 (8-9): 783-793.

Chung, W. 2006. Studying information seeking on the non-English Web: An experiment on a Spanish business Web portal. *International Journal of Human-Computer Studies* 64 (9): 811-829.

Dai, T., H. Hu, Y. Wan, Q. Chen, & Y. Wang. 2015. A data quality management and control framework and model for health decision support. In 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) (pp. 1792-1796). Zhangjiajie, China.

DAMA (Data Management Association International) 2017. DAMA-DMBOK. Data Management Body of Knowledge. 2nd Edition. New Jersey: Technics Publications LLC.

DAMA (Data Management Association International) 2008. The DAMA dictionary of data management. Denver, Colorado: Technics publications.

DAMA-UK (Data Management Association International) 2013. The six primary dimensions for data quality assessment: Defining Data Quality Dimensions. https://www.sbctc.edu/resources/documents/colleges-staff/commissions-councils/dgc/data-quality-deminsions.pdf (accessed Oct. 8, 2020).

Earley, S. 2011. *The DAMA Dictionary of Data Management.* (2nd). NJ: Technics Publications LLC.

English, L. P. 1999. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. Hoboken, NJ: Wiley.

Ehrlinger, L., & W. Wöß. 2022. A survey of data quality measurement and monitoring tools. *Frontiers in Big Data* 5: 850611.

Eurostat. 2015. ESS Handbook for Quality Reports. Brussels, Belgium: Eurostat.

Freitas, A., T. Knap, S. O'Riain, & E. Curry. 2011. W3P: Building an OPM based provenance model for the Web. *Future Generation Computer Systems* 27 (6): 766-774.

Freitas, A., T. Knap, S. O'Riain, & E. Curry. 2011. W3P: Building an OPM based provenance model for the Web. *Future Generation Computer Systems* 27 (6): 766-774.

Goel, S., & I. N. Chengalur-Smith. 2010. Metrics for characterizing the form of security policies. *The Journal of Strategic Information Systems* 19 (4): 281-295.

Haider, A., & S. H. Lee. 2012. Using six sigma for continuous improvement of asset management information quality. In Internatial Conference on Information Resources Management

(CONF-IRM), Retrieved from https://aisel.aisnet.org/confirm2012/ (accessed Feb 25, 2022).

Hassenstein, M. J., & P. Vanella. 2022. Data quality—concepts and problems. *Encyclopedia* 2 (1): 498-510

Hazen, B. T., C. A. Boone, J. D. Ezell, & L. A. Jones-Farmer. 2014. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics* 154: 72-80.

Hipp, J., U. Guntzer, & U. Grimmer. 2002. Data quality mining. In DMKD2001 Workshop on Research Issues in Data Mining and Knowledge Discovery DMKD2001. Santa Barbara, CA, USA.

Hsieh, C. C., P. L. Kuo, S. C. Yang, & S. H. Lin. 2010. Assessing blog-user satisfaction using the expectation and disconfirmation approach. *Computers in Human Behavior* 26 (6): 1434-1444.

Hsu, Yu-Lung, Cheng-Haw Lee, & V.B. Kreng. 2010. The application of Fuzzy Delphi Method and Fuzzy AHP in lubricant regenerative technology selection. *Expert Systems with Applications* 37 (1): 419-425.

Ifinedo, P., B. Rapp, A. Ifinedo, & K. Sundberg. 2010. Relationships among ERP post-implementation success constructs: An analysis at the organizational level. *Computers in Human Behavior* 26 (5): 1136-1148

ISO 25012. (n.d.). Retrieved from https://iso25000.com/index.php/en/iso-25000-standards/iso-25012 (accessed Feb.20, 2022)

Jin, X. L., C. M. Cheung, M. K. Lee, & H. P. Chen. 2009. How to keep members using the information in a computer-supported social network. *Computers in Human Behavior* 25 (5): 1172-1181.

Jarke, M., R. Gallersdörfer, M. A. Jeusfeld, M. Staudt, & S. Eherer. 1995. ConceptBase—a deductive object base for meta data management. *Journal of Intelligent Information Systems* 4 (2): 167-192.

Kim, B., & I. Han. 2011. The role of utilitarian and hedonic values and their antecedents in a mobile data service environment. *Expert Systems with Applications* 38 (3): 2311-2318.

Kim, C., E. Oh, N. Shin, & M. Chae. 2009. An empirical investigation of factors affecting ubiquitous computing use and U-business value. *International Journal of Information Management* 29 (6): 436-448.

Kim, W., B. J. Choi, E. K. Hong, S. K. Kim, & D. Lee. 2003. A taxonomy of dirty data. *Data mining and knowledge discovery* 7 (1): 81-99.

Kim, Y. J., R. Kishore, & G. L. Sanders. 2005. From DQ to EQ: understanding data quality in the context of e-business systems. *Communications of the ACM* 48 (10): 75-81.

Kwon, O., N. Lee, & B. Shin. 2014. Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management* 34 (3): 387-394.

Larburu, N., R. Bults, M. van Sinderen, & H. Hermens. 2015. Quality-of-data management for telemedicine systems. *Procedia Computer Science* 63: 451-458.

Lederer, A. L., D. J. Maupin, M. P. Sena, & Y. Zhuang. 2000. The technology acceptance model and the World Wide Web. *Decision support systems* 29 (3): 269-282.

Lee, H., J. Kim, & J. Kim. 2007. Determinants of success for application service provider: An empirical test in small businesses. *International journal of human-computer studies* 65 (9): 796-815.

Lee, K. C., & N. Chung. 2009. Understanding factors affecting trust in and satisfaction with mobile banking in Korea: A modified DeLone and McLean's model perspective. *Interacting with computers* 21 (5-6): 385-392.

Lee, S. H., & A. Haider. 2011. A Framework for Information Quality Assessment Using Six Sigma Approach. Communications of the IBIMA. https://ibimapublishing.com/journals/communications-of-the-ibima (accessed Mar 6, 2022).

Lee, Y. W., D. M. Strong, B. K. Kahn, & R. Y. Wang. AIMQ: a methodology for information quality assessment. *Information & management* 40 (2): 133-146.

Li, S., S. S. Rao, T. S. Ragu-Nathan, & B. Ragu-Nathan. 2005. Development and validation of a measurement instrument for studying supply chain management practices. *Journal of operations management* 23 (6): 618-641.

Lin, A. 2006. The acceptance and use of a business-to-business information system. *International Journal of Information Management* 26 (5): 386-400.

Michnik, J., & M. C. Lo. 2009. The assessment of the information quality with the aid of multiple criteria analysis. *European Journal of Operational Research* 195 (3): 850-856.

Negash, S., T. Ryan, & M. Igbaria. 2003. Quality and effectiveness in web-based customer support systems. *Information & management* 40 (8): 757-768.

Naumann, F. 2002. *Quality-driven query answering for integrated information systems* (Vol. 2261). Berlin: Springer: Springer.

Park, J., J. Kim, & J. Koh. 2010. Determinants of continuous usage intention in web analytics services. *Electronic Commerce Research and Applications* 9 (1): 61-72.

Peer, E., D. Rothschild, A. Gordon, Z. Evernden, & E. Damer. 2022. Data quality of platforms and panels for online behavioral research. *Behavior research methods*. 54 (4): 1643-1662.

Petter, Stacie, William DeLone & Ephraim McLean. 2008. Measuring information systems success: models, dimensions, measures, and interrelationships. *European Journal of Information Systems* 17: 236–263.

Redman, T. C., & A. Blanton. 1996. *Data quality for the information age*. Norwood: Artech House.

Salaün, Y., & K. Flores. 2001. Information quality: Meeting the needs of the consumer. *International Journal of Information Management* 21 (1): 21-37.

Sandelowski, M., and J. Barroso. 2007. *Handbook for synthesizing qualitative research*. New York, NY: Springer.

Sebastian-Coleman, L. 2012. *Measuring data quality for ongoing improvement: a data quality assessment framework*. Amsterdam: Morgan Kaufmann.

Serhani, M. A., H. T. El Kassabi, I. Taleb, & A. Nujum. 2016. An hybrid approach to quality evaluate across big data value chain. In 2016 IEEE International Congress on Big Data (Big Data Congress) (pp. 418-425). IEEE.

Song, J. 2007. Trust in health infomediaries. *Decision support systems* 43 (2): 390-407.

_____. and Fatemeh “Mariam” Zahedi. 2007. Trust in health infomediaries. *Decision support systems* 43 (2): 390-407.

Stjepandić, J., & W. Korol. 2022. Data quality management for interoperability. DigiTwin: An Approach for Production Process Optimization in a Built Environment, 135-153. https://link.springer.com/chapter/10.1007/978-3-030-77539-1_7. (accessed Sept.25, 2021).

Sung, T. J., & M. You. 2007. A method for establishing an online design audit platform. *Design Studies* 28 (2): 195-211.

Syed, R., R. Eden, T. Makasi, I. Chukwudi, A. Mamudu, M. Kamalpour, ... & T. Myers. 2023. Digital Health Data Quality Issues: Systematic Review. *Journal of Medical Internet Research* 25: e42615.

Taleb, I., & M. A. Serhani. 2017. Big Data pre-processing: closing the data quality enforcement loop. In 2017 IEEE International Congress on Big Data (BigData Congress) (pp. 498-501). IEEE. Boston, USA.

Taleb, I., H. T. El Kassabi, M. A. Serhani, R. Dssouli, & C. Bouhaddioui. 2016, July. Big data quality: A quality dimensions evaluation. In 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld) (pp. 759-765). Toulouse, France.

Tayi, G. K., & D. P. Ballou. 1998. Examining data quality. *Communications of the ACM* 41 (2): 54-57.

Wagner, T., N. R. Lottig, M. L. Bartley, E. M. Hanks, E. M. Schliep, N. B. Wikle, ... & J. Zhou. 2020. Increasing accuracy of lake nutrient predictions in thousands of lakes by leveraging water clarity data. *Limnology and Oceanography Letters* 5 (2): 228-235.

Wand, Y., & R. Y. Wang. 1996. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM* 39 (11): 86-95.

Wang, R. Y. 1998. A product perspective on total data quality management. *Communications of the ACM* 41 (2): 58-66.

_____, & D. M. Strong. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 12 (4): 5-33.

Wu, X., W. Zheng, X. Xia, & D. Lo. 2021. Data quality matters: A case study on data label correctness for security bug report prediction. *IEEE Transactions on Software Engineering* 48 (7): 2541-2556.

Yeganeh, N. K., S. Sadiq, M. A. & Sharaf. 2014. A framework for data quality aware query systems. *Information Systems* 46: 24-44.

سحر بنیادی

دانشجوی دکتری علم اطلاعات و دانش‌شناسی گرایش ذخیره و بازیابی اطلاعات دانشگاه آزاد اسلامی واحد علوم و تحقیقات و مدیر کتابخانه دانشگاه پیام نور اردبیل است.

مدیریت داده، مدیریت کیفیت داده، ذخیره و بازیابی اطلاعات از جمله علایق پژوهشی وی است.

نجلا حریری

عضو هیئت علمی دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران و استاد تمام گروه علم اطلاعات و دانش‌شناسی است.

رفتار اطلاع‌یابی، سازماندهی اطلاعات، مدیریت اطلاعات و پایگاه‌های اطلاعاتی از جمله علایق پژوهشی وی است.

سید مهدی طاهری

دارای مدرک دکتری در رشته علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون رئیس کتابخانه مرکزی و مرکز اسناد و استادیار دانشگاه علامه طباطبائی است. سازماندهی اطلاعات، مدیریت فراداده و کتابخانه دیجیتال از جمله علایق پژوهشی وی است.

رؤیا پورنقی

دارای مدرک تحصیلی دکتری در رشته علم اطلاعات و دانش‌شناسی است. ایشان هم‌اکنون دانشیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. سیستم‌های اطلاعات مکانی، دروازه‌های اطلاعاتی، علم‌سنجی، حق مؤلف، سواد اطلاعاتی و اثر پژوهش از علایق پژوهشی وی است.