Baltic J. Modern Computing, Vol. 11 (2023), No. 4, 686-702
https://doi.org/10.22364/bjmc.2023.11.4.08

Knowledge Discovery Frameworks and Characteristics

Martins JANSEVSKIS, Kaspars OSIS

Vidzeme University of Applied Sciences, Terbatas Street 10, Valmiera, Latvia

martins.jansevskis@va.lv, kaspars.osis@va.lv

ORCID 0009-0005-3435-723X, ORCID 0000-0002-6354-7009

Abstract. Knowledge is an abstract concept with no tangible connection to the physical world. Knowledge discovery refers to extracting knowledge from data and emphasizes the high-level application of specific techniques. The purpose of the knowledge discovery process is to extract knowledge from data. Knowledge discovery frameworks are a structured approach that combines various techniques and tools from different fields, such as data mining, machine learning, statistics, information visualization, and knowledge discovery process models, and includes underlying technologies to assist in the knowledge discovery systems development process. The primary goal of knowledge discovery systems is to identify patterns and relationships in the data that can be used to gain new insights and improve decision-making by applying a combination of tools, technologies, and techniques. The complexity of knowledge discovery systems can proliferate, making the use of knowledge discovery frameworks, design patterns, and process models important. This research summarizes the characteristics of knowledge discovery frameworks available for developing flexible and scalable knowledge discovery systems. The purpose is to determine if the characteristics indicate that the existing frameworks can support flexible and scalable knowledge discovery systems development according to modern design principles. In order to identify the knowledge discovery framework characteristics, authors apply structured literature research and identify the underlying knowledge discovery process models and characteristics of knowledge discovery frameworks.

Keywords. Knowledge discovery, Knowledge discovery frameworks, Knowledge discovery frameworks characteristics, Knowledge discovery process models, Selection of relevant studies approach, Feature collection of knowledge discovery frameworks

1. Introduction

Knowledge is an abstract concept with no tangible connection to the physical world. People have tried to define knowledge from Greek philosophers to knowledge management experts, but existing definitions still need to be clarified (Bolisani and Bratianu, 2018). Beginning with Plato and Aristotle, philosophers developed Epistemology as a theory of knowledge, trying to answer the fundamental question: What is knowledge? Defining knowledge and explaining its nature proved only possible with a convincing and generally accepted result (Neta and Pritchard, 2009). Most theories are integrated into two main perspectives: rationalism and empiricism. Both theories recognize that knowledge is based on true beliefs (Bolisani and Bratianu, 2018).

Knowledge discovery systems are developed by applying knowledge discovery process models. Almost three decades ago, knowledge discovery process models were developed in the late 1990s. In almost thirty years, knowledge discovery has evolved, and process models must be viewed in the context of the knowledge discovery framework and system in which the process is applied. Two of the first introduced process models were Knowledge Discovery in Databases (KDD) and Cross Industry Standard Process for Data Mining (CRISP-DM). Researchers Martínez-Plumed et al. (2021) and Rotondo and Quilligan (2020) argue that CRISP-DM remains the default standard for developing data acquisition and retrieval projects.

The knowledge discovery frameworks are designed to be applied to a business process to solve a specific objective. Knowledge discovery frameworks include the infrastructure definitions, modules, components, and potential technology stack. Knowledge discovery process models are an integral part of knowledge discovery frameworks as they describe the underlying process of knowledge discovery (see Fig. 1). Fig. 1 lists some of the knowledge discovery process models - KDD (Fayyad et al., 1996), CRISP-DM (Chapman et al., 2000), SEMMA (SAS, 2017), FMDS (Rollins, 2015), TDSP (Severtson, 2021) and RAMSYS (Moyle and Jorge, 2021) as well as the three in-depth analyzed knowledge discovery frameworks (see section 3).

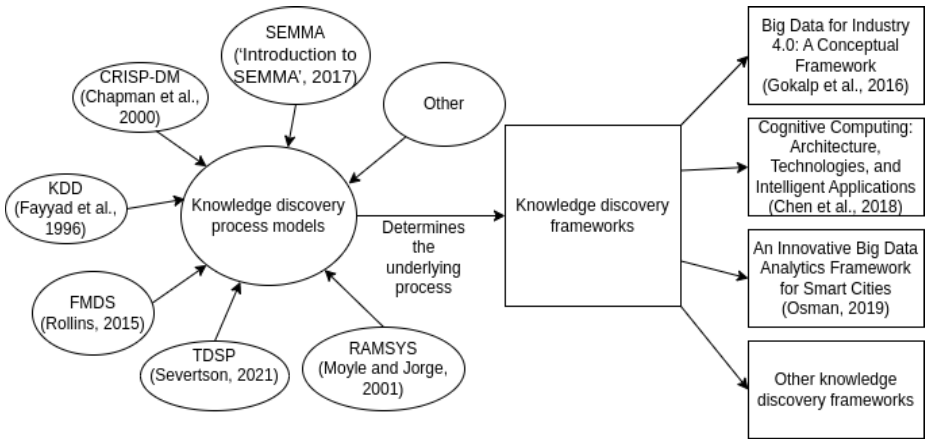

Figure 1. Knowledge discovery process models in context with knowledge discovery frameworks

The state of knowledge discovery process models and frameworks are identified by applying literature research in order to answer two questions - Q1: What design principles characterize knowledge discovery frameworks? Moreover, Q2: Do the characteristics indicate that the existing frameworks can support a flexible and scalable knowledge discovery system development according to modern design principles? To answer the questions, research articles addressing knowledge discovery frameworks were identified, selected, and analysed.

This research provides a comparison of various knowledge discovery frameworks. The authors have laid the groundwork for developing a conceptual knowledge discovery framework proposal through this research. By leveraging the insights gained from this research, authors are working to develop a conceptual knowledge discovery framework from the response data of intelligent systems to support the development of flexible and scalable intelligent systems with knowledge discovery capabilities. This research underscores the need for a well-defined process and framework for knowledge discovery and highlights the need for continued research and innovation in this field.

2. Selection of Relevant Studies

To address the research questions, the authors adopted systematic literature research about the knowledge discovery frameworks and the underlying knowledge discovery process models to determine the information base for this research. The search strategy was divided into several iterations with the following goals:

- identify publications, authors, domains, and problem areas covered
- perform literature search in scientific databases (Scopus, Science Direct, and IEEE Xplore - three research databases according to Indeed (Indeed, 2023)) and create an overview of the obtained information;
- use the following terms - knowledge discovery frameworks and knowledge discovery process models;
- create the information base for this research.

The following criteria were utilized to summarize the acquired results: four search scenarios with the top ten results in each scenario. The search scenarios were the following four:

- the most cited publications since the keyword first appeared in the scientific databases - to acquire the most fundamental publications;
- the most cited publications in the last ten years - to acquire the most influential publications in the last ten years;
- researchers with the most publications in the field - to determine the most prolific researchers;
- the most active researchers in the last ten years - to determine the active researchers.

The four search scenarios provide initial literature search results of 240 sources, creating an initial base to grasp the topic of interest. The search results were reviewed, and additional topic-relevant publications were included from the references to expand the result set (see Table 1). The results of the search scenarios were processed manually as the quantitative collection of results needed to provide more information about the publication.

To assess the relevance of the publications, authors applied a categorization scale, which was used to sort the results according to their applicability to the topic. Publications with the highest categorization value were summarized in the results (see Table 1), including related publications that could provide additional insights. Table 1 shows a fragment of the literature research results - the results were used to identify the existing knowledge discovery process models and knowledge discovery frameworks. The literature research results provide the information base for this research. The publications present in the results and not referenced are not contributing factors to this research.

Table 1. Fragment of the literature research results.

Publication	Related topic relevant publications
Osman, A.M.S. A novel big data analytics framework for smart cities, 2019	• Al Nuaimi, E., Al Neyadi, H., Mohamed, N., Al-Jaroodi, J., Applications of big data to smart cities, 2015 • Chen, M., Mao, S., Liu, Y., Big data: A survey, 2014 • Singh, D., Reddy, C.K., A survey on platforms for big data analytics, 2015 • Cooper, H.M., Organizing knowledge syntheses: A taxonomy of literature reviews, 1988 • Khan, Z., Anjum, A., Soomro, K., Tahir, M.A., Towards cloud based big data analytics for smart future cities, 2015

It was taken into account that a subjective literature research categorization scale creates a possibility of overlooking fundamental studies. To reduce this possibility, the authors used the results of the literature research and expanded the result set with additional resources identified in the process.

3. Knowledge Discovery Process Models

The term knowledge discovery refers to acquiring knowledge in data and emphasizes the high-level application of specific data mining methods. In the field of knowledge discovery in the 1990s, researchers defined a multi-step process of instructing users of data mining tools in the process execution efforts (Skoda and Adam, 2020).

Knowledge discovery process models provide a systematic and structured approach for discovering insights and patterns in complex data sets. Process models assist researchers and practitioners in following a predefined process, from data preparation to model evaluation, to uncover knowledge from data. Knowledge discovery process models facilitate the study of data and enable the development of reliable and accurate insights that can assist decision-making and advance our understanding of the world (Fayyad et al., 1996). Knowledge discovery process models describe the underlying processes of knowledge discovery frameworks and are an integral part.

One of the first proposed knowledge discovery process models was published in 1996, known as the Knowledge Discovery in Databases (see Fig. 2) (Fayyad et al., 1996).

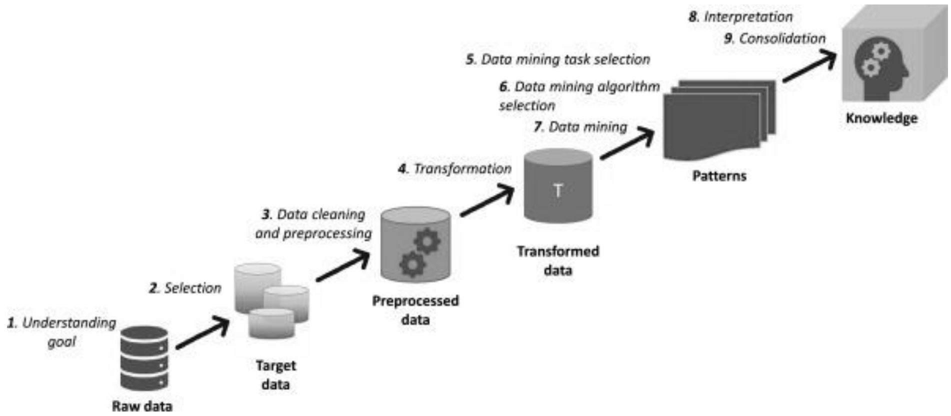

Figure 2. KDD process model (Fayyad et al., 1996)

The nine-step model developed by Fayyad et al. is considered one of the leading models in the knowledge discovery process (Osei-Bryson and Barclay, 2015). The KDD process is a simple methodology and a relatively easy-to-understand model for knowledge discovery.

Researchers Skoda and Adam describe two significant flaws in the KDD (Skoda and Adam, 2020). First - the defined levels must be more abstract, precise and formalized. The second drawback is the need for more description of the business aspects since the development of the model was a research-based approach.

Multiple knowledge discovery process models exist as different researchers and organizations have addressed the knowledge discovery realm. However, the process models have yet to be viewed in the context of the knowledge discovery frameworks. This section includes detailed information about KDD, CRISP-DM, Six-step and SEMMA process models. KDD is chosen as it was the first developed model, CRISP-DM as it is considered the default standard, the Four-step as it is from the early 1990s and one with the least amount of steps, and the Six-step model as the authors' subjective choice from the pool and SEMMA as it is developed by SAS, one of the largest manufacturers of statistical and business intelligence software.

3.1. Four-step model

The four-step model was proposed by researchers Berry and Linoff (Berry and Linoff, 1997). The model consists of the following steps (see Fig. 3): problem identification, problem analysis, execution of necessary actions and measuring results.

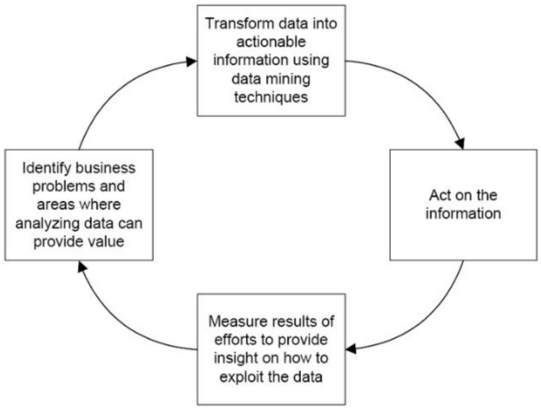

Figure 3. Four-step process model (Berry and Linoff, 1997)

Berry and Linoff (1997) specified 11 stages, further detailing the proposed four-step approach: (1) transform a business problem into a data mining problem, (2) select the appropriate data, (3) familiarize with the data, (4) create a set of models, (5) fix data problems, (6) transform data to make the information understandable, (7) make models, (8) evaluate models, (9) deploy models, (10) evaluate the results and (11) repeat the process from the beginning.

3.2. Six-step model

In 2005, researchers Cios and Kurgan proposed a six-step knowledge discovery process model (see Fig. 4). It includes understanding the scope of the problem, understanding the data, preparing the data, retrieving the data, analysing the results and applying the acquired knowledge (Cios and Kurgan, 2005).

The six-step model consists of the following steps (Cios and Kurgan, 2005):

- understanding the problem: Step involves collaborating with industry experts to define the problem and determine project goals, identify key individuals involved, and learn about current solutions to the problem;
- understanding the data: The step includes data sampling and appropriate data selection. The usefulness of the data concerning the objectives of knowledge discovery is tested;
- data preparation: The data preparation step determines how successful the knowledge retrieval process will be. Usually, this step takes about half of the total project time. This step makes decisions about which data will be used as input. This may include sampling data, performing correlation and significance tests, cleaning data such as checking data records for completeness and removing or correcting noise;
- data mining: It is considered to be the most crucial stage in the knowledge discovery process. This step involves the use of planned tools, many types of algorithms such as rough and fuzzy sets, Bayesian methods, evolutionary computing, neural networks, clustering, and pre-processing methods;

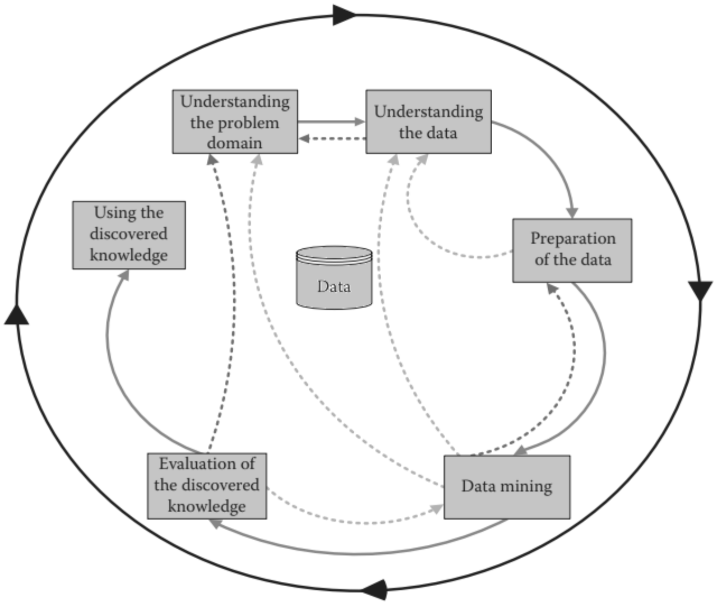

Figure 4. Six-step process model (Cios and Kurgan, 2005)

- evaluation of acquired knowledge. The step includes understanding and checking the results - whether the acquired knowledge is new and relevant, interpretation of the results by industry experts and checking the impact of the acquired knowledge;
- application of acquired knowledge. The step involves planning where and how the discovered knowledge will be used.

3.3. CRISP-DM model

Another of the early knowledge discovery process models is CRISP-DM (see Fig. 5) - an industry and instrument-neutral process model developed in late 1996 by three then-market leaders: Daimler, SPSS and NCR (Osei-Bryson and Barclay, 2015). In 1997, a consortium was established to formalize the experience of various organizations engaged in knowledge discovery. One of the key results is the focus on creating a non-proprietary and freely available model to assist in knowledge discovery projects (Osei-Bryson and Barclay, 2015).

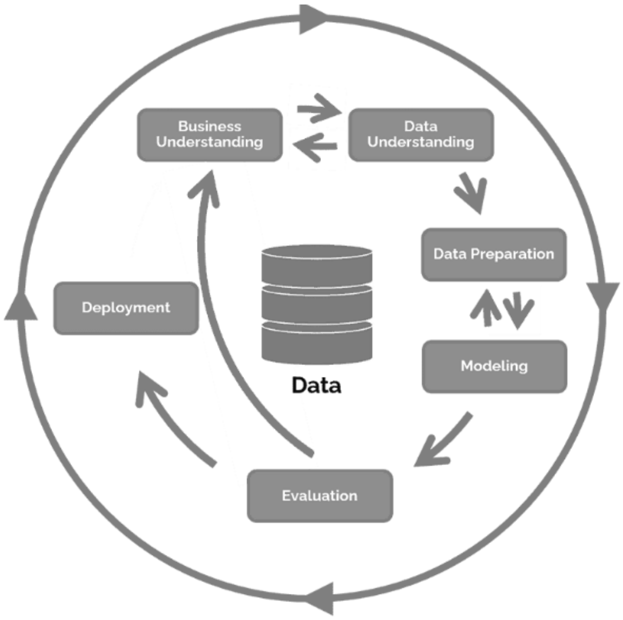

Figure 5. CRISP-DM process model (Chapman et al., 2000)

CRISP-DM (see Fig. 5) describes the life cycle of a knowledge discovery project in the form of six stages - understanding business goals, understanding data, preparing data, modeling, evaluating and deploying.

The CRISP-DM model also includes several feedback loops to emphasize that some steps must be revised to take advantage of the new information or knowledge acquired in the next step (Chapman et al., 2000).

Based on multiple surveys, researchers F. Martinez-Plumed et al. (2021) and Rotondo and Quilligan (2020) argue that CRISP-DM remains the default standard for designing data mining and knowledge retrieval projects.

3.4. SEMMA model

The name SEMMA denotes the sequential actions of the model (see Fig. 6): collection of data samples (Sample), data research (Explore), transformation (Modify), modelling (Model) and evaluation (Assess) (Azevedo and Santos, 2008).

The SEMMA model is a list of sequential actions developed by the SAS Institute, one of the largest statistical and business intelligence software manufacturers. The SEMMA model is considered a general data mining methodology, and the SAS Institute claims it is a logical set of functional tools for one of their products, SAS Enterprise Miner, to perform basic data mining tasks (SAS, 2017).

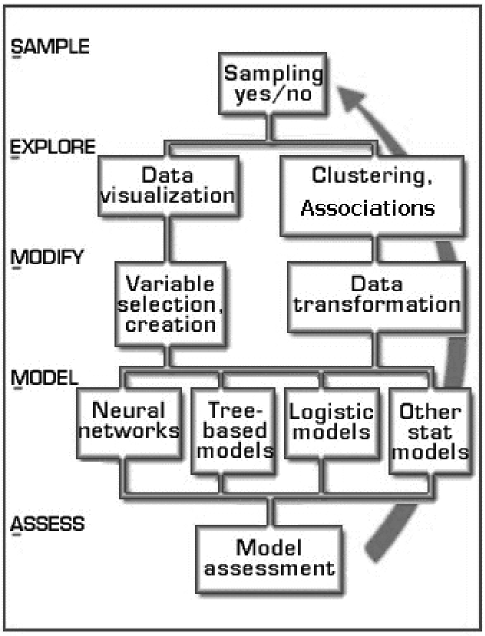

Figure 6. SEMMA process model (SAS, 2017)

3.5. Knowledge discovery process models evolution

The industry has evolved significantly, and data science is present in almost every organization. This section aims to identify and describe knowledge discovery process models without comparing them or specifying favoured and creating subjective conclusions. The development of different knowledge discovery process models and methodologies is depicted in Fig 7.

The arrows in the Fig. 7 indicate that CRISP-DM incorporates principles and ideas from several process models while forming the basis for later-developed knowledge discovery process models. New models of knowledge discovery processes have evolved as extensions of CRISP-DM while demonstrating how the model can be modernized without fundamentally changing it. Knowledge discovery process models RAMSYS (Moyle and Jorge, 2001), ASUM-DM (IBM, 2005), CASP-DM (Martínez-Plumed et al., 2017) and HACE (Xindong et al., 2014) are some examples of how a default standard model can be extended and applied for a specific field. The large volume of data, as well as the experimental and exploratory nature of data science projects, calls for less defined, lighter and more flexible methodologies; therefore, several IT companies have introduced new knowledge discovery process models for data science projects, such as IBM FMDS (Rollins, 2015) and Microsoft TDSP (Severtson, 2021).

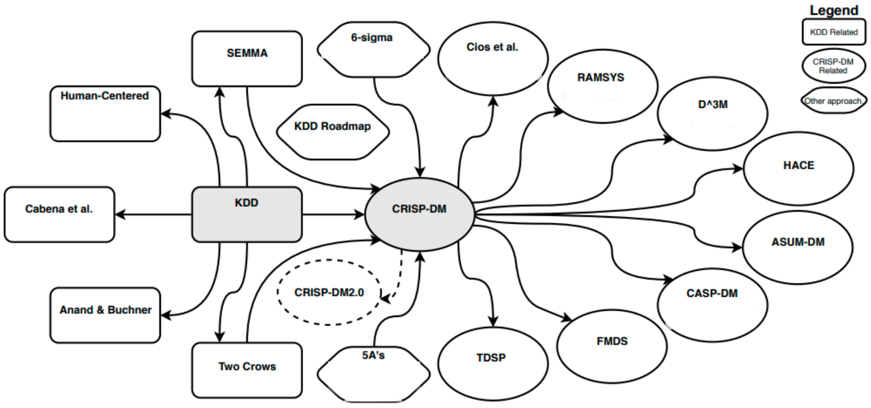

Figure 7. Knowledge discovery process models interconnection (Martinez-Plumed et al., 2021)

The knowledge discovery process has changed significantly since the creation of the CRISP-DM process, and an important area where the CRISP-DM model does not work well enough is data-driven products, where the product is the data and the knowledge acquired from it (Martinez-Plumed et al., 2021).

Data can have multiple use cases in a realm unrelated to the domain in which the data was obtained. This means that the knowledge discovery process needs to be viewed in the context of the knowledge discovery framework in which the process is applied. The knowledge discovery process models provide a systematic and structured approach to discovering insights and patterns in complex data sets. Therefore, a knowledge discovery framework has to support multiple process model approaches.

4. Knowledge Discovery Frameworks

Practical knowledge discovery is mostly a closed problem-solving process that includes a series of purposeful activities: problem definition, framework, and model development. Knowledge discovery is designed to serve business purposes that can be seamlessly linked or integrated with business processes and systems. These and similar challenges have affected how knowledge discovery process models are applied to business (Cao et al., 2010).

However, the knowledge discovery frameworks are designed to be applied to a business process to solve a specific challenge. This chapter examines several knowledge discovery frameworks. The frameworks have been chosen so that each is for a different domain. The frameworks are innovative, and their presenting publications have been used in related research. The following criteria are established to ensure framework selection - (1) the knowledge discovery framework is not older than six years, (2) the publication representing the framework, according to Scopus Field-Weighted Citation Impact, is rated with a value of at least nine, which means that this publication has added value in the domain.

Based on the criteria, three frameworks are chosen: "An Innovative Big Data Analytics Framework for Smart Cities" (Osman, 2019), "Cognitive Computing:

Architecture, Technologies, and Intelligent Applications" (Chen et al., 2018) and "Big Data for Industry 4.0: A Conceptual Framework" (Gokalp et al., 2016).

4.1. Industry 4.0 conceptual framework

According to Gokalp et al., the use and installation of big data analytics platforms require significant expertise in data science and IT in general due to their complex infrastructure and programming models. Otherwise, it could hinder the adoption of big data technologies in Industry 4.0. Thus, from the users' ability to adopt, big data platforms require a programming model that provides higher-level abstractions. The conceptual framework proposes a visual and data flow-based architectural framework that abstracts developers from the complexity of data processing platforms (Gokalp et al., 2016). The conceptual framework Gokal can be seen in Fig. 8. It includes the following modules - big data application design, data pre-processing, shared infrastructure and distribution of results.

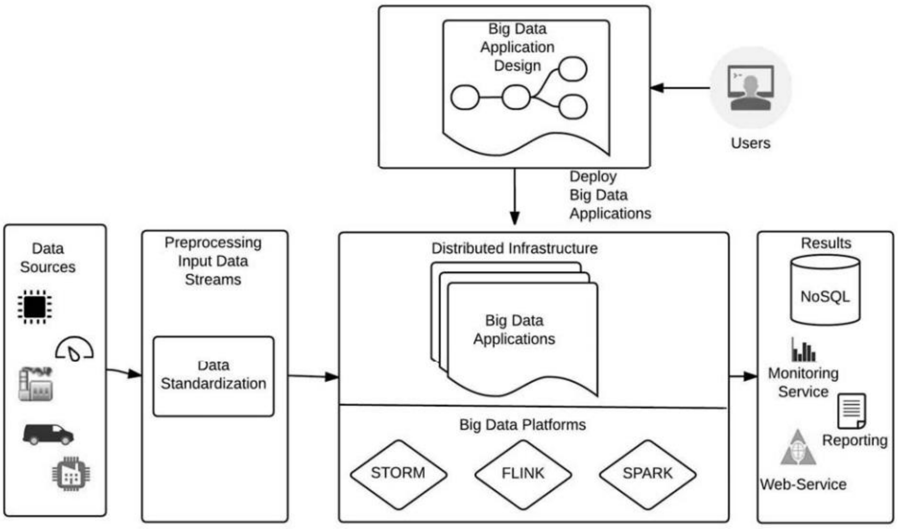

Figure 8. The conceptual framework proposed by Gokalp et al. (Gokalp et al., 2016)

In the conceptual framework, the application development module allows system engineers to develop applications of their choice. Applications are represented as directed graphs, where vertices represent data mining, machine learning algorithms, and programming constructs, while edges represent data flows corresponding to intermediate results. Programming nodes can connect and process data in a single standard from different sources and be integrated with other programming nodes. Application logic can be created by connecting programming nodes without worrying about their internal operations and interfaces (Gokalp et al., 2016).

Input data in the conceptual framework are generated in different formats, so data diversity is a significant challenge. The input data pre-processing module is essential - ensuring data transformation into a uniform format for further processing. Framework applications require high-performance and scalable infrastructure, which creates the need for a distributed infrastructure. Extensive data application requirements vary between use cases, so no one-size-fits-all big data solution exists. The conceptual framework aims to support multiple big data platforms such as Flink, Spark and Storm. Considering the logic and use cases of the developed applications, the framework can offer the most suitable big data platform for application development. The obtained results can be sent to interested parties in several ways. Each distribution channel is defined as a programming node, and end users can choose multiple distribution channels. The framework does not exclude the possibility of delivering results to external entities, using web services for data visualization or monitoring purposes (Gokalp et al., 2016).

The conceptual framework presents the components of the potential knowledge discovery system. However, it does not detail the basic principles of the modules' operation. The specific framework strictly preserves the conceptual approach.

4.2. Cognitive computing framework

Chen et al. describe cognitive computing as a subfield of artificial intelligence, which more accurately simulates human thought processes using self-learning algorithms, knowledge discovery, pattern recognition and natural language processing. These artificial environments rely on deep learning algorithms and neural networks to process information by comparing it to a training dataset (Chen et al., 2018).

Chen et al. believe that challenges related to human-machine interaction, voice recognition, and computer vision can be solved with the support of technologies such as 5G, robotics, deep learning, the Internet of Things, and cloud computing infrastructure. Supported application domains may include health monitoring, cognitive healthcare, smart city, intelligent transportation, and scientific experiments. Each layer of the framework architecture (see Fig. 9) is associated with corresponding technological challenges and technical requirements of the framework (Chen et al., 2018).

In the proposed framework (see Fig. 9), the Internet of Things obtains information about objects using, for example, RFID and wireless sensors, satellite and Wi-Fi positioning, and fingerprints. Using means of communication, the Internet of Things ensures information dissemination, sharing and integration, as well as information analysis and processing. Decision-making processes use intelligent technologies such as cloud computing, machine learning and knowledge discovery. Cognitive computing, suggested by the researchers Chen et al., could provide tools with better energy efficiency for data perception and extraction from the Internet of Things.

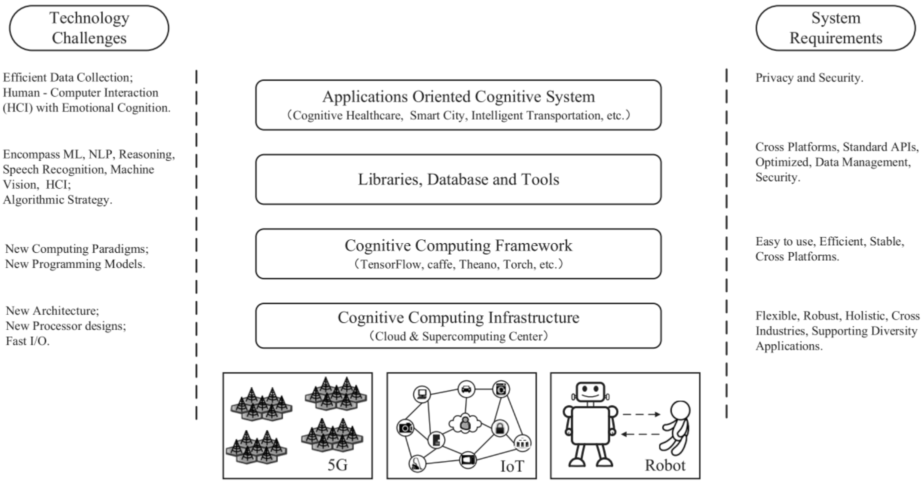

Figure 9. Architecture of a Cognitive Computing Framework (Chen et al., 2018)

One of the differences between big data analytics and cognitive computing is the amount of data. According to researchers Chen et al., big data analysis is not necessarily cognitive computing. Big data focuses on extracting knowledge from a large amount of data. In contrast, with a large amount of data, it is possible to guarantee the accuracy and reliability of the forecast. On the other hand, cognitive computing does not rely on large amounts of data but on cognition and judgment, similar to the human brain. Chen et al. believe that cloud computing virtualizes computing, data storage and communication lanes, reducing the cost of deploying software services and supporting industrialization and applications for promoting cognitive computing. In addition, cloud computing power provides dynamic, flexible, virtual, shared, and efficient computing resource services for cognitive computing solutions (Chen et al., 2018).

Although researchers Chen et al. present the potential architecture of the framework, explanations need to be provided about the applicable technologies within the framework or their placements, data storage and inter-module communication. The framework proposed by the researchers maintains the conceptual level without providing a more detailed explanation of the potential application areas.

4.3. Framework for smart cities

Researcher Osman (2019) believes that in order to gain valuable insight into the development of city-level intelligent information services, the generated data sets from different city domains should be integrated and analyzed, and this process is called big data analysis or big data value chain. The publication "Innovative Big Data Analytics Framework for Smart Cities" presents a framework for intelligent cities called "Smart City Data Analytics Panel – SCDAP" (Osman, 2019).

The conceptual architecture of the proposed SCDAP framework can be seen in Fig. 10. It consists of a three-layer architecture: platform layer, security layer and data processing layer (Osman, 2019).

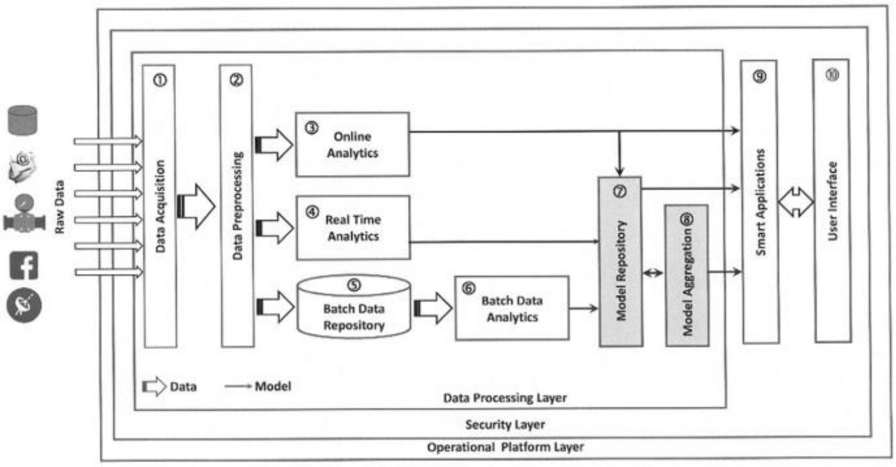

Figure 10. Architecture of the SCDAP framework (Osman, 2019)

As explained by researcher Osman, the framework layers have the following functionality: data mining, data pre-processing, online analytics, real-time analytics, batch data warehouse, model management, model aggregation, intelligent applications and user interface. Osman (2019) offers the conceptual SCDAP framework's approbation using Apache Hadoop and TensorFlow, Apache Hadoop and Anaconda. Six design conditions are put forward in SCDAP - a layered design approach, standardized data acquisition, real-time and historical data analytics, iterative and sequential data processing, model management, and model aggregation.

Osman's publication examines the knowledge discovery framework in the context of smart cities, looking at potential functionality, security and design requirements for the framework. The proposed SCDAP framework includes a contextual model and a potential implementation with a determined set of technologies.

4.4. Knowledge discovery framework comparison

To compare the selected knowledge discovery frameworks (see Table 2), the authors summarized the features that the researchers described in their proposals. Authors are summarizing the features provided in the publications and are not justifying, asserting or defending the criteria as this research aims to identify knowledge discovery framework characteristics. The collected features are not considered definitive elements characterizing knowledge discovery frameworks.

Table 2. The feature collection of knowledge discovery frameworks.

Feature	Industry 4.0 conceptual framework	Cognitive computing framework	Framework for smart cities
Has a description of the implementation	X	X	✓
Modular design	✓	✓	✓
Developers abstracted from the complexity of data processing platforms	✓	X	✓
Distributed infrastructure	✓	X	✓
Data pre-processing module	✓	X	✓
Results processing and distribution module	✓	✓	✓
Connectable programming nodes	✓	X	X
High performance scalable infrastructure	✓	X	✓
Support for big data platforms	X	X	✓
Potential technologies and their application are presented	X	X	✓
Online analytics module	X	X	✓

Table 2 shows that all of the compared frameworks are based on modular design and include results processing and distribution modules - both of the characteristics can be considered a requirement for a framework. The high-performance, scalable infrastructure, developers' and practitioners' abstraction from the complexity of data processing platforms, distributed infrastructure and pre-processing modules are also present in two of three frameworks and can be considered requirements. The rest of the summarized features require additional research to determine if they are required for a knowledge discovery framework.

The most complete information about the knowledge discovery framework is presented for the innovative intelligent cities' solution, where potential technologies are provided. The proposed knowledge discovery frameworks offer innovative approaches. However, they do not provide detailed information about the technologies, system architecture, data extraction and distribution scenarios, or the implementation of the frameworks in a particular industry.

As determined in section 3, the knowledge discovery process models provide a systematic and structured approach for discovering insights and patterns in complex data sets, and a knowledge discovery framework has to support multiple process model approaches; however, the analysed frameworks do not consider underlying process models. From the summary, the characteristics do not indicate that the existing frameworks can provide a flexible and scalable knowledge discovery system according to modern design principles.

5. Conclusions and Discussion

The knowledge discovery process models from the 90s are still applied within organizations' projects, and multiple knowledge discovery framework proposals exist for various domains. However, none of them consider underlying process models as impacting factors. As determined in the framework feature summary, the prerequisite and technological requirements for knowledge discovery frameworks remain open for discussions and future research. The characteristics do not indicate that the existing frameworks can provide a flexible and scalable knowledge discovery system according to modern design principles.

The authors have laid the groundwork for developing a conceptual knowledge discovery framework proposal through this research. This research underscores the need for a well-defined process and framework for knowledge discovery and highlights the need for continued research and innovation in this field.

References

Azevedo A., Santos M. (2008). KDD, SEMMA and CRISP-DM: A Parallel Overview. IADS-DM.

Berry, M. J. A., Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support.* Wiley.

Bolisani, E., Bratianu, C. (2018). 'The Elusive Definition of Knowledge', in Bolisani, E. and Bratianu, C., *Emergent Knowledge Strategies*. *Cham: Springer International Publishing (Knowledge Management and Organizational Learning)*, pp. 1–22.

Cao, L., Zhao, Y., Zhang, H., Luo, D., Zhang, C., Park, E. K. "Flexible Frameworks for Actionable Knowledge Discovery," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 9, pp. 1299-1312, Sept. 2010

Cios, K. J., Kurgan, L. A. (2005). Trends in Data Mining and Knowledge Discovery. In N. R. Pal and L. Jain (Eds.), Advanced Techniques in Knowledge Discovery and Data Mining (pp. 1–26). Springer London. https://doi.org/10.1007/1-84628-183-0_1

Chapman, P., Julian, C., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. h*ttps://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf*

Chen, M., Herrera, F., Hwang, K. (2018). Cognitive Computing: Architecture, Technologies and Intelligent Applications. IEEE Access, 6, 19774–19783.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework. KDD'96: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 82–88.*

Gokalp, M. O., Kayabay, K., Akyol, M. A., Eren, P. E., Kocyigit, A. (2016). Big Data for Industry 4.0: A Conceptual Framework. 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 431–434.

IBM (2005). Analytics solutions unified method. *ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf*

Indeed (2023). 23 research databases for professional and academic use. *Retrieved April 29, 2023, from https://www.indeed.com/career-advice/career-development/research-databases*

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Flach, P., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J. (2017). CASP-DM: Context Aware Standard Process for Data Mining. *ArXiv:1709.09003 [Cs]. http://arxiv.org/abs/1709.09003*

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061.*

Moyle, S., Jorge, A. (2001). RAMSYS-A methodology for supporting rapid remote collaborative data mining projects. *https://www.researchgate.net/publication/247329752_RAMSYS-A_methodology_for_supporting_rapid_remote_collaborative_data_mining_projects*

Neta, R., Pritchard, D. (2009). Arguing about knowledge. London ; *New York: Routledge (Arguing about philosophy).*

Osei-Bryson, K.-M., Barclay, C. (Eds.). (2015). Knowledge discovery process and methods to enhance organizational performance. *CRC Press, Taylor & Francis Group.*

Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. Future Generation Computer Systems, 91, 620–633.

Rollins, J.B. (2015). Foundational Methodology for Data Science. IBM Analytics. *https://tdwi.org/~/media/64511A895D86457E964174EDC5C4C7B1.PDF*

Rotondo, A., Quilligan, F. (2020) 'Evolution Paths for Knowledge Discovery and Data Mining Process Models', *SN Computer Science, 1(2), p. 109.*

SAS (2017). Introduction to SEMMA. SAS Help Center. https://documentation.sas.com/doc/en/emref/14.3/n061bzurmej4j3n1jnj8bbjjm1a2.htm

Severtson, R. B. (2021, December 15). What is the Team Data Science Process? *https://docs.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle*

Skoda, P., Adam, F. (Eds.). (2020). Knowledge discovery in big data from astronomy and earth observation: *AstroGeoInformatics. Elsevier.*

*X*indong, W., Xingquan, Z., Gong-Qing, Wu., Wei, D. (2014). Data mining with big data. I*EEE Transactions on Knowledge and Data Engineering, 26(1), 97–107.*

Received December 6, 2022, revised October 31, 2023, accepted December 7, 2023