SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2026
ORGANIZATION AND MANAGEMENT SERIES NO. 242

GRAPH DATABASE AS A TOOL FOR REPRESENTING DOMAIN KNOWLEDGE – A CASE STUDY

Kinga NOWAKOWSKA

Kielce University of Technology; kinga.nowakowska@autograf.pl, ORCID: 0009-0009-9372-1790

Purpose: To present graph databases as an effective tool for representing and exploring domain knowledge. A graph database was designed as an implementation of a semantic model to demonstrate the feasibility of mapping comprehensive relationships within a single, coherent database system. The case study examines dog breeding as an industry with a high degree of interconnection among genetic, health, and behavioural data.
Design/methodology/approach: A graph knowledge database model was designed and implemented in the Memgraph environment, drawing on prior studies. The implementation included identification of knowledge units and semantic relations, construction of a graph in the Cypher language, and experimental data mining through queries and result consistency analysis. The study is conceptual and application-oriented and is based on a case study methodology.
Findings: The proposed solutions confirmed that the graph model allows natural mapping of links in domain knowledge. The developed database enables adequate data storage and interactive exploration, including analyses of disease occurrence, links between breeds, and relationships between symptoms and treatments. The graph model facilitates inference and integration of various data types within a single system.
Research limitations/implications: The model is a prototype verified on a limited dataset (two dog breeds) and does not include complete genealogical or statistical data. Future research will extend the model with recommendation modules and machine learning integration.
Practical implications: The database may support decision-making in breeding and veterinary practice by enabling analysis of relationships between diseases, symptoms, and treatment methods.
Social implications: Improved knowledge representation in animal breeding may contribute to more responsible practices and better health management.
Originality/value: The study's innovative approach is the use of a graph model in zootechnics and its potential for decision support systems. The paper demonstrates the feasibility of applying graph database technology to dog breeding, a domain relatively underrepresented in the literature.
Keywords: graph databases, domain knowledge, knowledge representation, NoSQL, Memgraph.
Category of the paper: Research paper.

1. Introduction

The current development information technology development are connected with a continuous increase in diverse data – comprehensive, multidimensional and derived from a multitude of sources. Data collection itself is becoming increasingly important, as is its interpretation in the context of interdependencies. Relations between data units are a key element of analysis in numerous areas of science and economy (Groth, 2020). In such cases, classic relational databases, based on a rigid tabular structure, prove insufficient.

Non-relational systems have emerged in response to these limitations and are collectively referred to as NoSQL. (*Not only SQL*) – allowing for flexible modelling of data structures. One of their variations include graph databases, which allow for natural mapping of the links between objects (Sadalage, Fowler, 2012).

Graph databases, such as Memgraph, are increasingly widely used in analytical systems, where relationships among data are crucial. They are used, among other applications, in the analysis of social networks, recommendation systems, biological research, genetic relationships, and specialist areas such as animal breeding. The use of a graph structure enables effective modelling of dynamic and complex relationships that are difficult to capture with classical relational models (Robinson, 2015).

This article aims to analyse the possibility of using graph databases as an effective tool for representing and exploring domain knowledge. An example of the analysed are is the breeding of purebred dogs, in particular Amstaffs and German Shepherds. Knowledge in this area includes information on the origin of animals, their health status, characteristics, behaviour, genealogy and breeding relationships. These data do not exist in isolation – they form an extensive network of connections, which is ideal for mapping as graphs.

The current state of research on graph databases indicates their growing importance in the analysis of semantic knowledge. According to Hogan (2021) and Groth and Moreau (2020), knowledge graphs form the basis of modern information systems, enabling the integration of data from various sources and their processing in meaningful contexts. Studies by Angles and Gutierrez (2018) have shown that the graph model is particularly effective at mapping complex cause-and-effect relationships, while Franconi and Tessaris (2002) have highlighted its usefulness in conceptual and biological systems. In practical applications, Memgraph and Neo4j are examples of tools that combine real-time visualization and analysis of relations, as evidenced by Mansfeld (2020) and Robinson (2015). Building on these studies, this paper contributes by applying graph theory to model knowledge in dog breeding. This area had been poorly represented in the scientific literature.

The work is conceptual and application-oriented and presents a case study of the design and implementation of a graph knowledge database using the Memgraph system. The article presents a data structure model, sample queries in Cypher, and an analysis of the practical possibilities for knowledge exploration.

The following research hypotheses were verified:

H1: The graph structure allows for a practical and intuitive exploration of knowledge based on semantic relations.

H2: As a graph platform, Memgraph may serve as a foundation for decision support systems in specialist areas, such as breeding purebred dogs.

The article combines a literature review, conceptual design, and experimental implementation of a graph model in the context of domain knowledge.

The subsequent sections of the article present:

- a review of NoSQL literature and technology,
- a description of the graph construction and analysis of the technology used,
- contextual knowledge obtained from the database under experimental queries,
- a discussion of results in the context of existing research,
- conclusions and recommendations on the possibility of further development of the system.

This work is based on a case study, as recommended by Yin (2018), in which a graph database was used to represent domain knowledge in dog breeding. The research methodology covers: an analysis of the literature on graph databases and representation of knowledge (Groth, Moreau, 2020; Hogan et al., 2021), a design of a semantic graph model based on domain-specific relations, the implementation of the Memgraph model using Cypher, and the operationalization of queries about real analytical problems in animal breeding.

2. Literary and technological review

With the development of information technology and the rapid growth in data volume, traditional relational databases (RDBMS) have faced significant limitations, particularly in scalability, flexibility, and the processing of complex data structures. The growing demand for systems capable of handling large, dynamic and heterogeneous sets of data has led to the development of a new class of databases – NoSQL systems. (Not only SQL) (Peng, 2018).

NoSQL databases differ from traditional relational databases in that they do not rely on rigid tabular structures. Instead, they use flexible data models that better reflect real-world relationships and enable efficient information processing in highly volatile environments. The main types of

NoSQL databases include:

1. Document databases – storing data in documents, often in formats such as JSON or BSON. Each document constitutes a separate unit of information, containing both the data and its structure. Examples of such databases include MongoDB and CouchDB. Their application is primarily in systems that require high structural flexibility, e.g., websites and e-commerce applications (Sadalage, 2012).
2. Column bases – they store data in columns instead of rows, which allows for high performance analysis of large data sets, in particular in analytical applications. Examples of such systems include Apache Cassandra and HBase, which are used, among other applications, for real-time analysis of transaction data.
3. Key-value databases – the simplest type of NoSQL databases where data is stored in pairs: a unique key and its corresponding value. Examples include Redis and DynamoDB. Such databases are characterized by exceptionally high efficiency for read and write operations and high scalability.
4. Graph databases – they store data as nodes and edges, which map the relations between objects. Each node represents a specific entity (e.g., a person, an organization, an animal), whereas edge describes their mutual relationship (e.g., "belongs to", "suffers from", "is a parent"). This model allows the natural representation of the comprehensive connections that occur in reality (Graph Foundation, 2021).

2.1. Development of NoSQL approach and graph concepts

The concept of non-relational databases is not new; it originated in the 1960s, when IBM developed the IMS (Information Management System), which operated on a hierarchical structure. The breakthrough came in the 70s, when Edgar Codd proposed a relational model which dominated the database market for multiple decades. The emergence of large-scale, dynamic data (Big Data) and distributed systems has necessitated more flexible solutions (Sadalage, Fowler, 2012).

NoSQL systems introduced four key features (Sadalage, Fowler, 2012):

- schema flexibility – the lack of need to define the data structure in advance,
- horizontal scalability – the ability to disperse data between multiple nodes of the system,
- support for unstructured data – the ability to integrate data from different sources in various formats,
- high performance – optimized support for reading and writing data,
- one of the critical design assumptions of NoSQL databases is the CAP theory (Hogan et al., 2021), which describes a compromise between three system properties (see Figure 1):
  - consistency,
  - availability,
  - partition tolerance.

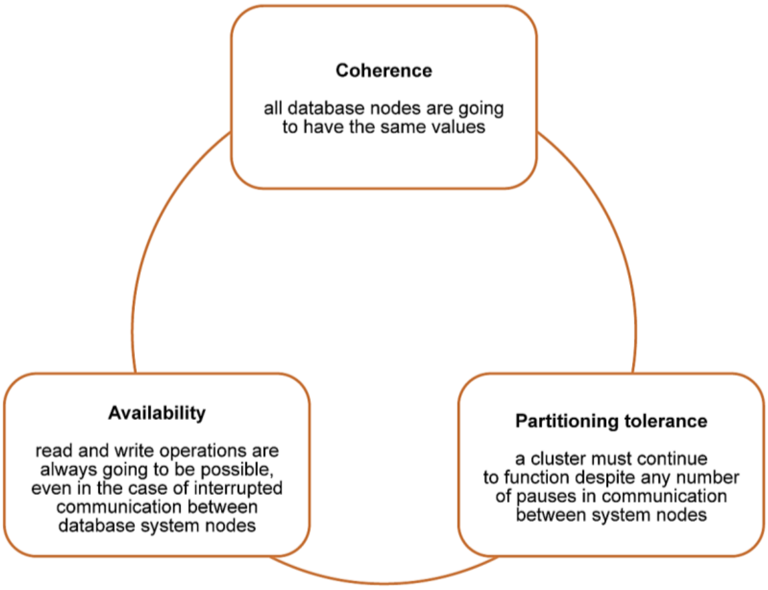

Figure 1 CAP theory diagram.

Source: Own elaboration.

CAP theory states that a database can provide only two of those three features simultaneously. In practical terms, this means that NoSQL systems often abandon complete consistency in favour of greater availability and resiliency (Microsoft, 2021).

2.2. Graph databases and their applications

Graph databases occupy a critical position among NoSQL databases, as they allow modeling relationships between objects in a manner that is closest to human understanding of relationships. Data are stored as nodes and edges, and each edge may contain attributes describing the nature of the relationship (Sadalage, Fowler, 2012).

In practice, graph databases are used, among others, to (Hogan et al., 2021):

- analyse social networks (e.g., relations between users),
- detect frauds,
- analyse biological and genetic links,
- represent expert knowledge in decision-making systems,
- manage relations in animal breeding and genetics.

As noted by Hogan (2021), a graph representation of knowledge enables not only data storage but also the mapping of the meaning and hierarchy of concepts, making it highly useful in semantic applications.

2.3. Memgraph – a tool for analysing and exploring graphs

Memgraph, a modern graph database system, was used for implementation. This system combines high performance (because it operates in the operating memory) with full support for ACID transactions. Its main advantages include:

- support for Cypher query language,
- real-time data analysis,
- integration with data streams,
- easy visualization of graphs,
- scalability and flexibility in modelling knowledge structures.

The use of this tool was justified by its intuitive syntax, high efficiency, and advanced analytical capabilities, which enable the exploration of complex relationships between objects (Angles, 2018; Memgraph Ltd., 2021). Memgraph, as an implementation tool, was identified following a comparative analysis of available graph engines, including Neo4j, ArangoDB, and OrientDB. The final decision was made based on its high performance in processing data in operational memory, which also ensures the portability of queries.

Memgraph – a modern graph engine supporting the Cypher query language, compliant with the principles of ACID transactions (Angles, 2018; Memgraph Ltd., 2021), which meets four basic requirements:

- **A** (atomicity) – each transaction is indivisible.
- **C** (consistency) – a transaction leads the database from one consistent state to another.
- **I** (isolation) – parallel transactions do not affect each other.
- **D** (durability) – once a transaction is approved, the data is durable, even in the event of a failure.

This principle guarantees the correctness of the operation and the reliable operation of the database (Sadalge, 2012).

The advantages of Memgraph include:

- high efficiency due to operation in the operating memory,
- complete support for transactions,
- the ability to perform real-time data analysis,
- integration with data streams,
- simple visualization of query results.

Memgraph allows the creation of dynamic knowledge structures in which data is not stored linearly but rather as links between nodes. This feature makes it particularly useful for modeling areas in which relationships play a dominant role, as in breeding purebred dogs.

In the scientific context, graph databases are part of a broader concept of knowledge graphs, which, as emphasized by Groth and Moreau (2020), enable the integration of data from various sources into a coherent semantic structure.

A graph database was designed as an implementation of a semantic model to demonstrate the feasibility of mapping comprehensive relationships of meaning within a single, coherent database system.

3. Methods

3.1. Project concept

A prototype graph knowledge database, Canis, was created using the Memgraph framework. The area of application chosen for the case study was the breeding of purebred dogs. This example was used for illustrative purposes as a model depicting how graph databases may support knowledge mining and inference processes in domains with complex information structures. The project aimed to create a data structure that not only stores information about objects and their properties but primarily maps their complex interrelationships.

The following methodological assumptions were adopted in the design process.

- the base must be semantic in nature, i.e., relations between nodes must not only combine the data technically, but also reflect their logical meaning (e.g., *a dog suffers from a disease*, *the disease manifests itself as a symptom*);
- the structure must be extensible and scalable, allowing for easy addition of new types of nodes (e.g., new breeds, diseases, care elements);
- the database is designed to support knowledge exploration through queries in Cypher language, including analytical and comparative queries;
- the project aims to reflect realistic data, based on generally known characteristics of exemplary breeds of Amstaffs and German Shepherds, their typical diseases, and methods of care.

The study employed qualitative methods to analyze the structure of domain knowledge and the design methods used in knowledge engineering (Sowa, 2000; Brachman, Levesque, 2004). The data were sourced from publications on dog diseases, breed patterns, and breeding recommendations, official FCI breed standards, and sample descriptions of veterinary cases. These data have been transformed into unified knowledge units (graph nodes). The criteria for model validation included: semantic consistency of relations, ambiguity of relations between entities, and the correctness of results obtained via test queries. The research hypothesis was verified by applying a set of diagnostic, comparative, and exploratory queries, consistent with the analytical triangulation method recommended in studies on knowledge graphs (Hogan et al., 2021).

The research methodology included three stages:

1. design of the knowledge graph structure, identifying the key knowledge units and relationships,
2. implementation of the model in the Memgraph environment,
3. experimental exploration – covering query tests and analysis of results consistency.

The selection of dog breeding as an application domain was justified by the availability of semantic relationships (genetic, health, behavioural), thus making this area an excellent test case for verifying research hypotheses.

3.2. Knowledge graph structure design

The basic unit of the knowledge base is a node representing a particular real entity – in this case, a dog. This node is the central point of the graph, and the relations from other nodes stem from it.

The project defines the following main types of nodes:

- *dog*– contains information on a particular specimen, such as race, name, body weight, colour, growth, and character traits,
- *disease* – represents a disease entity, e.g., hip dysplasia, food allergy,
- *symptom* – describes a clinical symptom of the disease, e.g., limping, mobility issues, falls,
- *treatment* – describes a therapeutic method, e.g., rehabilitation, surgery, supplementation,
- *care* – applies to care treatments, e.g., combing, bathing, and cleaning the ears,
- *nutrition* and *dog food* – contains information on diet, type of food, and feeding frequency.

Relations between nodes reflect logical dependencies typical of the breeding industry:

- *(Dog)-[:SUFFERS_FROM]→(Disease).*
- *(Disease)-[:SYMPTOM]→(Symptom).*
- *(Disease)-[:TREATMENT]→(Treatment).*
- *(Dog)-[:CARE]→(Care).*
- *(Dog)-[:Feeding]→(Food).*

These relations are semantic in nature; each carries meaning and enables interpretation of data within the field, not only in technical terms.

Genealogical relations are planned for future development of the database, such as *IS_PARENT* and DERIVED_FROM, which will enable analysis of trait and disease inheritance across genetic lines. The database structure has been designed to support expansion with additional data types, such as genealogical relations (IS_PARENT, DERIVED_FROM) or statistical indicators of disease frequency.

3.3. Implementation – exemplary Cypher queries

The database creation began with the definition of nodes and relations in Cypher. The following code snippets present examples of commands for creating nodes and relationships in the Canis graph knowledge base (figure 2).

```
CREATE (p:Dog {
  race: "Amstaff,"
  weight: "30 kg,"
  colour: "brindle,"
  height: "45 cm"
})
CREATE (cz:Care {type: "Combing"})
CREATE (p)-[:CARE {
  frequency: "once per week,"
  tools: "rubber brush, glove, comb."
}]->(cz)
```

Nodes and relations on diseases have been created similarly, as illustrated in Figure 2:

```
CREATE (ch:Disease {name: "Hip dysplasia"})
CREATE (obj:Symptom {symptom: "Limping"})
CREATE (p)-[:SUFFERS_FROM]->(ch)
CREATE (obj)-[:SYMPTOM]->(ch)
```

Figure 2. Concept of nodes and their mutual connections in the Canis database.

Source: Own elaboration.

The structure is designed to allow flexible expansion of the database; adding a new dog, disease, or symptom does not require modifying the diagram. This makes the model scalable and dynamic, a key advantage of graph databases.

3.4. Principles of knowledge exploration

The developed model allows the formulation of queries of varying degrees of detail. Users – breeders, veterinarians, students – can explore data intuitively and flexibly, thus obtaining answers to practical and analytical questions.

Sample queries:

- Which dog breeds are registered in the database?
  MATCH (r:Dog) RETURN DISTINCT r.breed
- Which diseases occur in specific breeds?
  MATCH (p:Dog)-[:SUFFERS_FROM]->(c:Disease) RETURN p.breed, c.name
- Which symptoms are associated with a specific disease?
  MATCH (c:Disease)<-[:SYMPTOM]-(o:Symptom) RETURN c.name, o.symptom
- What diseases are common among the two breeds (Amstaffs and German Shepherds)?
  MATCH (:Dog {breed: “Amstaff”})-[:SUFFERS_FROM]->(c:Disease)<-[:SUFFERS_FROM]-(:Dog {breed: “German Shepherd”})
  RETURN DISTINCT c.name

4. Results

4.1. Characteristics of the developed database

Exploring the *Canis* database. A graph domain knowledge base was created in the Memgraph environment as a result of the project, including data on selected dog breeds: Amstaffs and German Shepherds. Each Dog node has been associated with a set of properties describing its physical characteristics (body weight, height, colour) and with other nodes representing components of the domain knowledge, such as diseases, symptoms, treatments, care, and nutrition.

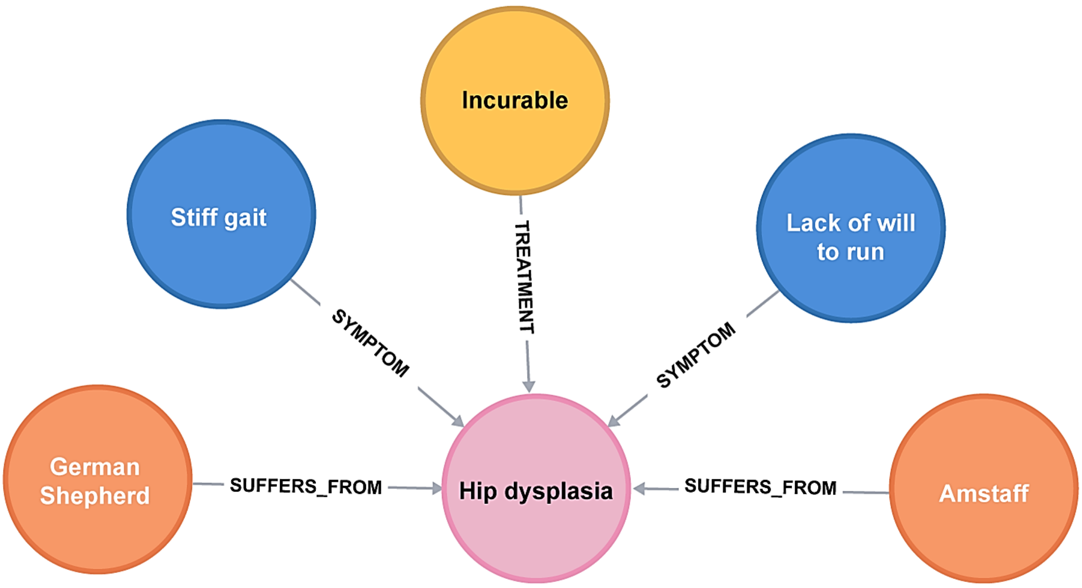

Figure 3. Representation of the case study graph database – detail.

Source: Own elaboration.

The developed graph enables movement across varying levels of information, from a single individual to general relations. For instance, Figure 3 presents a condition shared by both breeds, i.e. *hip dysplasia*. Symptoms may include limping and mobility issues. The treatment methods include Surgery and Exercise. The MATCH command generates the diagram.

4.2. Results of exploratory inquiries

This chapter presents the results of exploratory Cypher queries, illustrating how knowledge can be obtained from the Canis graph database and the practical analytical possibilities enabled by the applied data structure.

Example 1. Summary of dog breeds registered in the database

Query:

MATCH (r:Dog) RETURN DISTINCT r.breed

Result:

Amstaff

German Shepherd

The results indicate that two dog species, representing different breeding lines, were introduced into the test version of the database. The query enables users to identify available breeds and serves as a starting point for further data exploration.

Example 2. A list of diseases attributed to specific breeds

MATCH (p:Dog)-[:SUFFERS_FROM]->(c:Disease) RETURN p.breed, c.name

Result:

Breed	Disease
Amstaff	Hip dysplasia
German Shepherd	Hip dysplasia
German Shepherd	Food allergy

The resulting list indicates the links between disease units and dog breeds recorded in the database. The result provides a basis for further analysis of the relationship between genetic factors and disease frequency.

Example 3. Identification of dogs suffering from a selected disease entity

Query:

MATCH (p:Dog)-[:SUFFERS_FROM]->(c:Disease {name: “Hip dysplasia”})

RETURN p.name, p.breed

Result:

Name	Breed
Rex	Amstaff
Luna	German Shepherd

This query allows the identification of individuals affected by a particular disease. The data may serve as the basis for further analyses, e.g., epidemiological or breeding analyses.

Example 4. A list of disease symptoms and treatment methods

Query:

MATCH (p:Dog)-[:SUFFERS_FROM]->(c:Disease {name: “Hip Dysplasia”})<-[rel]-(info)

WHERE type(rel) IN [“SYMPTOM”, “TREATMENT”].

RETURN type(rel), info

Result:

Relationship type	Name/description
SYMPTOM	Limping, mobility issues, falls
TREATMENT	Rehabilitation, surgery, supplementation
SYMPTOM	Limping, mobility issues, falls

The results reflect how knowledge units are linked in the graph—each disease has a set of relationships describing its symptoms and potential treatments.

Example 5. Diseases common to the two breeds

Query:

```
MATCH (:Dog {breed: "Amstaff"})-[:SUFFERS_FROM]->(c:Disease)<-
[:SUFFERS_FROM]-(:Dog {breed: "German Shepherd"})
RETURN DISTINCT c.name
```

Result:

Hip dysplasia

The query results indicate that hip dysplasia occurs in both breeds under analysis. This allows further research into its genetic condition.

Example 6. Potential disease based on symptoms

Query:

```
MATCH (c:Disease)<-[:SYMPTOM]-(o:Symptom)
WHERE o.symptom IN [
  "Mobility issues,"
  "Limping,"
  "Falls,"
  "Difficulty maintaining balance and turning"]
RETURN DISTINCT c.name
```

Result:

Hip dysplasia
Arthritis
Pelvic limb injury

This type of query allows for the identification of potential disease units based on defined symptoms. It is an example of the application of the database in veterinary diagnostic processes.

4.3. Graph result structure

The actions performed yielded a complete knowledge graph structure, including nodes of the following types: Dog, Disease, Symptom, Treatment, Care, and Feed, connected by appropriate semantic relations.

Figure 4 presents a general representation of the relationship between knowledge units in the *Canis* database.

Figure 4. A representation of a portion of the Canis graph database.

Source: Own elaboration.

5. Discussion

Exploring the *Canis* database graph not only allows the storage of information on dogs and their characteristics, but also enables the acquisition of knowledge dynamically through queries based on inter-node relationships. The database may serve as a support for decision-making in breeding and veterinary medicine. Similar conclusions were drawn in the research by Angles and Gutierrez (2018), in which graph models of knowledge enabled the analysis of cause-and-effect relationships in biological systems. The applied model has potential for implementation and education. It may serve as the basis for systems supporting breeding, clinical, or didactic decisions. In the future, the database could be integrated with expert systems or machine learning algorithms, enabling automatic inference from the relational structure.

The application of Memgraph database in this project confirmed the practical advantages of the environment – e.g., the ability to execute queries in Cypher, support for ACID transactions, and integration with data streams. As pointed out by Mansfeld (2020), such features make graph databases handy for analysing dynamic data, where relationships are more important than values themselves.

In the context of the subject literature, it should be noted that the presented model is consistent with the stream of research on the ontological representation of domain knowledge. Brachman (2004) and Sowa (2000) point out that effective representation of knowledge requires tools that allow for the association of facts with their semantic meaning. The graph knowledge base presented in the article implements the assumption.

Compared to other works, the project stands out for its attempt to apply the concept of graph modelling of knowledge in the area of animal husbandry, which has been relatively poorly represented in computer science literature to date. Typically, graph databases are used in social network analysis, natural language processing, and biological data analysis (Hogan, 2021). This paper demonstrates that the same concept can be successfully applied in the zootechnical context for modeling and exploring breeding data.

The results confirm H1, which states that exploring knowledge using Cypher queries proved intuitive and practical, consistent with observations by Robinson (2015) and Angles (2018). Compared to previous studies on expert systems in medicine (Franconi, Tessaris, 2002), the presented model shows greater structural flexibility while maintaining relational semantics. The results support H2: the Memgraph database could serve as the foundation for future decision support systems. In the context of dog breeding, such a system can provide breeders with information on genetic links, behavioural characteristics, and health predispositions, thereby supporting breeding selection and planning. For veterinarians, the system enables analysis of the relationships among diseases, symptoms, and treatment modalities, supporting diagnostics, monitoring animal health, and planning therapies. The graph structure enables quick, intuitive link searches and the generation of recommendations based on complex relationships among data.

It should be noted that the developed model is a prototype and has been verified on a limited dataset (two breeds of dogs). Therefore, its full implementation would require expanding the database to include additional cases and integrating it with real breeding and veterinary records. In addition, the visual graph analysis, as specified in the original project assumptions, was not fully implemented due to constraints of the presentation environment.

The results obtained indicate that graph databases are an effective tool for representing and analyzing domain knowledge. In the context of the presented case study – breeding purebred dogs – the application of a graph structure allows for the combination of data from various areas: health, genealogical, and behavioural. Unlike traditional relational databases, which store information in tables with a fixed schema, a graph approach allows for more natural and semantically accurate representation of relations among individuals (Robinson, 2015; Groth, Moreau, 2020).

The project results align with the observations of other authors on graph knowledge modelling. Hogan (2021) emphasizes that knowledge graphs enable a transition from static information storage to dynamic association, thereby increasing the analytical value of the data. Similarly, Memgraph, used in this work, enabled interactive exploration of the relations among units of knowledge.

Similar applications have been reported in the bioinformatics literature (Franconi, Tessaris, 2002), in which graphs are used to represent metabolic networks and gene-to-gene relationships. This project uses a similar approach: the relationships among diseases, symptoms, and treatment modalities constitute a logical network, enabling multidimensional knowledge analysis.

6. Conclusions and summary

This article aimed to examine the feasibility of using graph databases to represent and explore domain knowledge. Based on the case study concerning the breeding of purebred dogs, it was shown that the graph data model enables a natural and semantically accurate representation of reality.

The results of the conducted analyses confirmed that the graph databases – in particular the Memgraph environment – allow:

- represent multidimensional relationships between knowledge units,
- integrate data at various structural levels (e.g., race, disease, symptoms, treatment, nutrition),
- real-time exploration of connections through Cypher queries,
- easy scaling and expansion of the database without the need to modify its schema,
- building a coherent semantic model reflecting the domain logic.

From a practical perspective, the designed graph knowledge base is a functional tool for collecting and analyzing breeding information, supporting decision-making processes in health prevention, planning crossbreeding, and genetic risk assessment. This system may also be used in veterinary and zoological education as an example of the implementation of the ontological knowledge model. From a scientific point of view, the project shows several important innovative features:

1. Application of the graph knowledge base in the dog breeding domain, which has not been described in detail in the literature.
2. Integration of many aspects of knowledge (health, care, nutrition, genetics) into a single, coherent semantic model.

3. The potential to extend the model with statistical and genealogical data, allowing for the analysis of the inheritance of diseases.
4. The possibility of using the base as a component of the recommendation system – e.g., in advisory services related to breeding and veterinary care.

The research carried out brings new value to the literature of the subject through:

1. Application of graph database technology to the domain of animal husbandry, which is a new field of its use.
2. Integration of multiple types of information – health, care, behavioural, and genealogical – into a single, coherent system.
3. Demonstration of the potential of a practical graph model of knowledge, which can provide the basis for future expert systems or decision support tools.
4. Proposing a universal graph structure which may be adopted in other domains, such as biology, medicine, or education.

It should be emphasized that the presented model is a prototype and requires further verification on larger real datasets. The database contained a limited number of units and lacked complete genealogical data or veterinary records.

The research conducted confirmed that graph databases are among the most promising solutions for modern knowledge management systems. Their ability to combine information in a relational context and to continuously expand their scope make them suitable for use in the life sciences, medicine, education, or the agricultural industry.

The model used in the work, despite the limitations in terms of its scale, proves that even in a small thematic domain – such as breeding purebred dogs – it is possible to effectively develop a coherent knowledge graph which not only collects data, but also supports their interpretation and practical application.

The results are consistent with Robinson's observations (2015), which emphasize that graph data models make it easier to navigate complex relationship structures than relational databases. Similarly, Hogan et al. (2021) indicate that graph databases facilitate the representation of knowledge in which semantic relations are as relevant as the objects themselves, as confirmed by the results of this study on the possibility of a complete interconnection among dogs, diseases, symptoms, and treatments. The results of our study are therefore consistent with the trends reported in the literature on the subject.

In conclusion, the discussion confirms that the use of graph databases, such as Memgraph, is a modern and effective tool for processing domain knowledge. The project adds scientific value, demonstrating the feasibility of practical mapping of complex semantic relations in animal husbandry.

In result, the developed system may be considered a prototype of an intelligent graph knowledge base, providing a starting point for further research on the use of graph data models in the automation of cognitive and decision-making processes.

The application of the model in the zootechnical domain is a novel approach that has not been the subject of extensive analysis to date. Our research shows that the graph structure can be extended to include a statistical and genealogical layer, enabling applications in predictive analyses.

Research plans include developing recommendation modules and integrating them with machine learning tools, enabling automated inference based on graph structure. Future work will consist of recommendation modules and integration with ML to automate inference.

The planned directions of further research include:

- Expansion of the database to include quantitative data (frequency of occurrence, effectiveness of treatment).
- Implementation of a recommendation module using logical rules and semantic links.
- Integration with a web interface allowing for the exploration of the graph by non-technical users.
- Transferring the model to other domains, such as horse breeding, cat breeding or genetic analysis in medicine.

References

1. Angles, R. (2018). *An introduction to graph data management*.
2. Brachman, R.J. (2004). *Knowledge representation and reasoning*. Elsevier.
3. Franconi, E. (2002). Knowledge bases and conceptual modeling. In: *Conceptual Modeling: Foundations and Applications*.
4. Fundation, A. (2021). *Advanced Graph Databases and Their Applications.* Retrieved from: https://opensourcejournal.com
5. Groth, P. (2020). Knowledge graphs: Introduction, history, and perspectives. *AI Magazine*, *no. 41(4),* 23-36.
6. Hogan, A.B. (2021). Knowledge graphs. *ACM Computing surveys (CSUR*), *no. 54(4),* 1-37.
7. Mansfeld, T. (2020). *Scaling Graph Databases in Cloud environments*, https://www.graphdataconference.com.
8. Memgraph Ltd. (2021). *Memgraph.* Retrieved from: https://memgraph.com
9. Microsoft (2021). *NoSQL Database - What is NoSQL?* Retrieved from: https://azure.microsoft.com/pl-pl/overview/nosql-database/
10. Nowakowska, K. (2022). Reprezentacja wiedzy o hodowli psów wybranych ras z wykorzystaniem grafowej bazy danych [*Representation of knowledge about breeding dogs of selected breeds using a graph database*]. Kielce.
11. Peng, B.P. (2018). NoSQL databases: Critical analysis and comparison. *International Journal of Computer Applications*, *no. 182(43),* 1-9.

12. Robinson, I.W. (2015). *Graph databases: New opportunities for connected data (2nd ed.).* O'Reilly Media.
13. Sadalage, P.J. (2012). *NoSQL distilled: A Brief Guide to the Emerging World of Polyglot Persistence.* M.F. Pramod, J. Sadalage (eds.). Addison-Wesley Professional.
14. Schwitter, R. (2010). *Controlled natural languages for knowledge representation*. Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics.
15. Sowa, J.F. (2000). *Knowledge representation: Logical, philosophical, and computational foundations*.
16. The Graph Fundation (2021). *OnGDB*. Retrieved from: https://www.graphfoundation.org/ongdb/