An enhanced domain ontology model of database course in computing curricula

Nur W. Rahayu1,2, Ridi Ferdiana2, Sri S. Kusumawardani2
1Department of Informatics, Faculty of Technology Industrial, Universitas Islam Indonesia, Yogyakarta, Indonesia
2Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia

ABSTRACT
The ACM/IEEE Computing Curricula 2020 includes the study of relational databases in four of its six disciplines. However, a domain ontology model of multidisciplinary database course does not exist. Therefore, the current study aims to build a domain ontology model for the multidisciplinary database course. The research process comprises three phases: a review of database course contents based on the ACM/IEEE Computing Curricula 2020, a literature review of relevant domain ontology models, and a design research phase using the NeOn methodology framework. The ontology building involves the ontology reuse and reengineering of existing models, along with the construction of some classes from a non-ontological resource. The approach to ontology reuse and reengineering demonstrates ontology reusability. The final domain ontology model is then evaluated using two ontology syntactic metrics: Relationship Richness and Information Richness. These metrics reflect the diversity of relationships and the breadth of knowledge in the model, respectively. In conclusion, the current research contributes to the Computing Curricula by providing an ontology model for a multidisciplinary database course. The model, developed through ontology reuse and reengineering and the integration of non-ontological resources, exhibits more diverse relationships and represents a broader range of knowledge.

Keywords: Multidisciplinary database course, Non-ontological resource, Ontology reuse and reengineering, Ontology syntactic metrics, Relational database structured query language

Corresponding Author:
Ridi Ferdiana
Department of Electrical and Information Engineering, Universitas Gadjah Mada
Yogyakarta, Indonesia
Email: ridi@ugm.ac.id

1. INTRODUCTION
Being a valuable resource, data requires effective management encompassing tasks like data cleaning, database storage, and data-based decision-making. Nowadays, the role of the database system as data storage is becoming increasingly important. It is also supported by more varied database models, e.g., relational data models, key-value stores, time series databases, graph databases, and spatial databases. Of these various models, the relational model is the most widely used with 71.9% popularity [1]. Therefore, it is not surprising that relational database learning is included in the curriculum for students in computing disciplines.

The relational database is a core course in computing disciplines [2]. Computing disciplines have evolved from five disciplines in 2005 to six disciplines in 2020. Previously, the joint task force ACM/AIS/IEEE released Computing Curricula 2005 which covered 5 disciplines, i.e., Computer Engineering, Computer Science, Information Systems, Information Technology, and Software Engineering [3]. Later, the ACM/IEEE Computing Curricula 2020 has 6 disciplines, i.e., the five existing disciplines plus Cybersecurity [4].

This is an open access article under the CC BY-SA license.
Along with the growth of database system technology, the increasing needs of database user organizations, and the development of international curricula, database learning methods and tools are also increasing. Moreover, learning databases can be challenging. The difficulty arises from database design and its programming language, which is abstract [5], [6]. Therefore, teaching and learning database is generally carried out in theory and practice sessions [7], [8] and enriched with various learning tools [6], [9].

The learning tools for databases vary, ranging from visualization [5], and web-based applications, to serious games [2]. The coverage of materials across these tools is also diverse, although no single tool encompasses all materials [6], [10]. With the growing demand for advanced and personalized learning tools, the range of techniques and methods being utilized is expanding, including the use of ontology. Previous studies have shown that ontologies are popularly used in e-Learning and are often collaborated with other artificial intelligence techniques or fields [11], [12]. Despite the potential to enhance tool flexibility, there is currently no ontology available for Database courses across various computing disciplines. To address this gap, we propose the following research questions:

RQ1: What are the Database course contents in multiple disciplines?
RQ2: How to develop a domain ontology model for a multidisciplinary database course?

Ontologies are frequently employed as data models in e-learning recommender systems. The findings of this study are expected to provide a valuable resource for developers of intelligent learning technologies by offering an established ontology model, thereby minimizing the necessity for constructing one from the ground up. With reusability capability, the new ontology can be used across various application platforms, including the learning management system.

The structure of this paper is as follows: Section 1 presents the research background and continues with section 2 which contains literature on Computing Curricula 2020 and ontology. Next, section 3 presents the methodology, and the results are covered in section 4. Discussion of research results is presented in section 5 and this paper closes with conclusions in section 6.

2. BACKGROUND
2.1. ACM/IEEE Computing Curricula 2020
The association for computing machinery (ACM) and IEEE computer society started efforts to develop curricula for computing disciplines in 2017. It was stated that after 2005, "the number and type of computing degree programs available to students has dramatically increased" [4]. As computing includes "a family of study areas", the Computing Curricula 2020 document contains curricula for computer engineering (CE), computer science (CS), cybersecurity (CSEC), information systems (IS), information technology (IT), software engineering (SE) and data science (DS).

The ACM/IEEE Computing Curricula 2020 (CC2020) emphasizes that learning in the current computer field leads to competence. It uses competence definition as "the quality or state of having sufficient knowledge, judgment, skill, or strength". The CC2020 competency model consists of 4 components, i.e., knowledge, skills, and disposition in carrying out tasks. Knowledge focuses on factual information, skills refer to the capability to utilize this knowledge, and attitude steers individuals toward the proper application of these competencies [4].

2.2. Ontology
An ontology is defined as "a logical structure of terms used to describe a domain of knowledge, including both the definitions of applicable terms and their relationships" [13], [14]. Technically, an ontology contains classes, individuals, and properties [15]. As part of the knowledge base, ontologies are popularly used in e-Learning environments. For example, previous research [16] states that ontologies are commonly used for curriculum modeling, describing learning domains, describing learner data, and describing e-Learning services.

In relation to learning contents, ontology is classified into two categories: subject domain ontology, which stores database learning materials, and learning task ontology, which archives activities and assessment items.

Ontologies can be constructed using various methodologies, one of which is the NeOn methodology. This approach offers nine different scenarios to build an ontology model. The scenarios are: (1) From specification to implementation, (2) Reusing and reengineering non-ontological resources, (3) Reusing ontological resources, (4) Reusing and reengineering ontological resources, (5) Reusing and merging ontological resources, (6) Reusing, merging, and reengineering ontological resources, (7) Reusing ontology design patterns, (8) Restructuring ontological resources, and (9) Localizing ontological resources [17].

Furthermore, the quality of the ontology model can be evaluated at three distinct levels, i.e., (i) syntactical (considering formal structure and language), (ii) semantic (determining meanings); and (iii) pragmatic (regarding intentions and usefulness) [18], [19]. At the syntax level, there are several metrics to assess the ontology model, such as Relationship Richness and Information Richness.
The Relationship Richness (RR) is useful for assessing the diversity of relationships among classes. RR is determined by dividing the count of non-inheritance relationships (P) by the total count of both inheritance (H) and non-inheritance relationships (P), as shown in (1). To elaborate, subclasses represent inheritance relationships, whereas object properties, equivalent classes, and disjoint classes fall under the category of non-inheritance relationships [20].

$$RR = \frac{|P|}{|H| + |P|}$$  \hspace{1cm} (1)

The Information Richness (IR) metric evaluates the depth or breadth of an ontology model, calculated as the average number of subclasses per class. This value distinguishes between two types of ontologies: (1) horizontal ontologies, where classes have numerous direct subclasses, and (2) vertical ontologies, characterized by classes with a limited number of direct subclasses [20].

3. METHOD

To address the first research question concerning course content, we conducted a review of database course materials across various disciplines. The second research question, pertaining to ontology development, was addressed through a combined approach of literature review and design research. These three phases are outlined below,

Phase 1: Review the database course content across multiple disciplines [21]. This phase is performed by referring to the CC2020 document and the curriculum of each computing discipline. Analysis was conducted on knowledge areas, knowledge units, and learning outcomes. Synthesis was subsequently performed by searching for common learning outcomes across various disciplines.

Phase 2: Conduct a literature review related to domain ontology models in the Database course. Literature was searched on Google Scholar using the keyword “ontology design database course” and documented using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (“the PRISMA 2020 statement”) [22]. Afterward, we analyzed the ontology classes of the existing models.

Phase 3: Design research to build a new domain ontology model. Refers to the NeOn methodology framework, this study combines Scenario 2 and Scenario 4 [17]. This phase includes four steps:

− Create a mapping from the necessary database contents (Phase 1) to the existing ontologies (Phase 2). The purpose of this step is to identify the need for adding, reusing, or reengineering classes.
− Add new classes and properties (Scenario 2) based on a Database textbook. The “Database System Concepts 7th edition” book [23] was selected as a design reference because it is the latest and widely recognized textbook [2].
− Modify classes and properties according to Scenario 4 based on the required course contents.
− Evaluate the new ontology model using RR and IR measurements. The results are then compared with the previous ontology model.

4. RESULTS

4.1. Phase 1: Review the database course content

Among the six disciplines accommodated by CC2020, four study databases, i.e., Computer Engineering (ACM/IEEE-CS Computer Engineering Curricula 2016) [24], Computer Science (ACM/IEEE-CS Computer Science Curricula 2013) [25], Information Systems (ACM/AIS Information System Curricula 2020) [26] and Information Technology (ACM/IEEE-CS Information Technology Curricula 2017) [27]. The competency analysis, including detailed knowledge areas, knowledge units, and learning outcomes across the four disciplines, is presented in Table 1. The analysis process is further advanced by assigning specific topic keywords to each learning outcome. This step produces 6 topics from 36 learning outcomes, i.e., data models (n=9), database systems (n=9), normalization (n=2), procedural language (n=2), structured query language (SQL) of data definition (n=6), and SQL query (n=8). The keyword distribution is depicted in Figure 1. The figure illustrates that out of the six most discussed topics, four are prominent: data modeling, database systems, SQL data definition, and SQL queries. The detail for each topic is below.

− Data modeling: the relationship among data models, examples, the context of use, and Entity Relationship Diagram design from a case study.
− Database systems: history, components, and types.
− SQL data definition: syntax to create and modify schemas or tables, rows, columns, keys, and views.
− SQL query to learn how to:
  − translate user stories into SQL statements,
− use SELECT statements using filtering and sorting
− make JOINs,
− utilize calculated fields and aggregate functions, and
− optimize the query.

Table 1. Comparison of database course content among four computing disciplines

<table>
<thead>
<tr>
<th>No</th>
<th>Discipline</th>
<th>Knowledge areas*</th>
<th>Knowledge unit**</th>
<th>No. of learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Computer Engineering (CE)</td>
<td>CE-SWD (Software design)</td>
<td>CE-SWD-9 (Data modelling)</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Computer Science (CS)</td>
<td>Information Management (IM)</td>
<td>IM/Database Systems</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Information Systems (IS)</td>
<td>Data / Information Management</td>
<td>Query the relational model</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Information Technology (IT)</td>
<td>ITE-IMA (Information Management)</td>
<td>ITE-IMA-03 (Data modelling)</td>
<td>5</td>
</tr>
</tbody>
</table>

Total 36

* This column header accommodates the terms “Knowledge Area” used in CE and CS, “Competency Area” in IS, and “IT Domain” in IT.

** The current study limits the maximum of two knowledge units that contain more database-related learning outcomes.

Figure 1. The intersection of learning outcomes in the multidisciplinary database course

4.2. Phase 2: A literature review of ontology models

Initially, a Google Scholar search using the keyword "ontology design database course" yielded a result of 230,000 articles. Considering time efficiency, the current research is limited to the first 50 articles. Subsequently, the process of article selection is carried out through abstract scanning. The process resulted in seven articles related to domain ontologies for database courses.

After full-text reading, we found 2 articles used general class names (i.e., chapter or topic), 1 article referred to the same ontology, and 1 article did not detail the ontology. Therefore, at the end of the selection
An enhanced domain ontology model of database course in computing curricula (Nur W. Rahayu)

4.3. Phase 3: Design research of a new ontology model

As mentioned in the Method section, ontology design research consists of four sequential steps. The steps are mapping the necessary contents, using Scenarios 2 and 4 for ontology building, and assessing the new model. The outcomes are presented below.

Step 1: Mapping the required database contents to the existing ontologies.

Phase 1 results indicate that popular database contents include data models, database systems, SQL data definitions, and SQL queries. However, the content of existing ontology models varies. The mapping result is summarized in Table 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Data model</td>
<td>Data model types and database design stages</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2.</td>
<td>Database system</td>
<td>Data storage</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3.</td>
<td>SQL data definition</td>
<td>—</td>
<td>Yes</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>4.</td>
<td>SQL query</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Step 2: Reuse and reengineer the non-ontology resources.

The step is performed as there is no complete model exists. Scenario 2 (“Reusing and reengineering non-ontological resources”) was conducted for the first two required contents (“Data model” and “Database system”). We developed a new ontology model utilizing a top-down approach, transitioning from broad to specific concepts. To address this, the textbook ‘Database System Concepts 7th Edition’ is our primary reference:

- Chapter 1, ‘Introduction’, constructs the ‘DatabaseComponent’ and ‘DataModel’ classes.
- Chapter 6, ‘Database Design Using the ER Model’ is for the “EntityRelationshipModel” class (a subclass of “DataModel”).
- Chapter 20, ‘Database System Architectures’, constructs the “DatabaseArchitecture” class.

The "SimpleAttribute" class exemplifies the relationships among classes. It is a subclass of the "AttributeType" class and is disjoint with the "CompositeAttribute" class. An excerpt of the ontology model for Scenario 2, generated using the OntoGraph plugin, is presented in Figure 3(a).

Step 3: Reuse and reengineer the ontology model

Scenario 4, as applied to studies [29] and [30] focuses on the last two materials: ‘SQL data definition’ and ‘SQL query’. In the new model, the ‘SQL data definition’ is represented by two top-level classes, ‘DDLStatement’ and ‘Data’. Similarly, the ‘SQL query’ is mapped to the ‘Clause’ and ‘SelectStatement’ classes. These four newly introduced classes are elaborated further through subclass hierarchical relationships, as depicted in Figure 3(b).

Step 4: Evaluation of the newly developed ontology model

The newly developed ontology model has 141 classes. The top-level classes are ‘Clause’, ‘Data’, ‘DatabaseArchitecture’, ‘DatabaseComponent’, ‘DataModel’, ‘DDLStatement’, and ‘SelectStatement’. The model is openly accessible on Figshare [31], with detailed metrics provided in Table 3.
Table 3. New ontology model metrics

<table>
<thead>
<tr>
<th>No</th>
<th>Metrics</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Class count</td>
<td>141</td>
</tr>
<tr>
<td>2.</td>
<td>subClassOf count</td>
<td>134</td>
</tr>
<tr>
<td>3.</td>
<td>Object property count</td>
<td>8</td>
</tr>
<tr>
<td>4.</td>
<td>Equivalent classes count</td>
<td>0</td>
</tr>
<tr>
<td>5.</td>
<td>Disjoint classes count</td>
<td>13</td>
</tr>
</tbody>
</table>

The existing model is then assessed using Relationship Richness (RR) and Information Richness (IR) metrics. These values are subsequently compared with the most comprehensive existing ontology model, as detailed in Table 4. The table reveals that the new ontology model exhibits a greater diversity of relationships, as indicated by its higher RR value. Furthermore, the new model is more horizontally oriented compared to the previous one [29]. This is attributed to the broader knowledge representation in the new model, while the existing model encapsulates more detailed knowledge (vertical ontology).

Table 4. The comparison of ontology evaluation metrics

<table>
<thead>
<tr>
<th>Evaluation metrics</th>
<th>The existing ontology model (Study [29])</th>
<th>The new ontology model</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>( RR = \frac{2}{196 + 2} = 0.01 )</td>
<td>( RR = \frac{21}{134 + 21} = 0.16 )</td>
</tr>
<tr>
<td>IR</td>
<td>( IR = \frac{17}{7} = 15.08 )</td>
<td>( IR = \frac{21}{7} = 19.14 )</td>
</tr>
</tbody>
</table>

5. DISCUSSION

The field of databases is rapidly evolving, necessitating the advancement of database education to keep pace with technological progress. The current research has discovered that the most popular topics in databases across various disciplines are data modeling, database systems, SQL data definition, and SQL queries. These findings support previous research stating that data modeling and its implementation are the core of database learning in many higher education institutions [2]. Furthermore, the four subjects are also present in the master’s level curricula to some extent [32]. Hence, this ontology model is expected to be utilized and expanded at the undergraduate and master’s levels.

Materials on database design and programming are often abstract, prompting the development of learning tools. These tools vary not only in the platforms they utilize but also in the scope of their content. However, no single tool provides comprehensive coverage of all materials [6], [10]. This fact underscores the need for these tools to adapt to diverse learning requirements. Therefore, the newly developed ontology model with high information richness is more suitable for reuse across multiple disciplines due to its broader scope.

6. CONCLUSION

The current research was conducted to answer the two research questions mentioned in the Introduction section. We performed the document analysis of the ACM/IEEE Computing Curricula 2020, a literature review of existing ontology models, and design research of a new domain ontology model. We found that the database course is taught in four disciplines: Computer Engineering, Computer Science, Information Systems, and Information Technology. In addressing our first research question (RQ1: “What are the database course contents in multiple disciplines?”), we discovered that despite differences in materials and learning outcomes, there are four popular topics: data modeling, database systems, SQL data definition, and SQL queries.

Existing ontology models differ in content, with none encompassing all four topics. To address the second research question (RQ2: “How to develop a domain ontology model for a multidisciplinary database course?”), we used the NeOn methodology framework in our design research. We utilized Scenario 2 (“Reusing and reengineering non-ontological resources”) and Scenario 4 (“Reusing and reengineering ontological resources”). Therefore, the ontology model construction involved two scenarios: (1) reusing textbook material, and (2) reusing and reengineering existing ontology models. As a result, our newly developed ontology model comprises 141 classes. Compared to its predecessor, this model exhibits more diverse relationships (evidenced by a 0.15-point increase in RR value) and represents a broader range of knowledge (with a higher IR score of 4.06 points).

Ontology models, renowned for their reusability and interoperability, offer a robust framework for data modeling. Future research could integrate the model into applications such as Learning Management Systems, Computer-Aided Instruction, or Intelligent Tutoring Systems. Furthermore, exploring database...
models beyond relational databases could diversify the learning contents and provide learners with a broader understanding of various industry-relevant database models.

ACKNOWLEDGEMENTS

This work was funded by RTA Program Universitas Gadjah Mada with Grant Number 5075/UN.I.P.II/Dit-Lit/PT.01.01/2023.

REFERENCES


An enhanced domain ontology model of database course in computing curricula (Nur W. Rahayu)