*Journal of Harbin Engineering University*
*ISSN: 1006-7043*

Vol 46 No. 10
september 2025

# PhyQA: An Ontology-Based Question Answering System for Formula Identification in Physics Subject

**Catherine Hii Seng Jing[1], Bong Chih How[2]**

[1, 2] Faculty of Computer Science and Information Technology, University of Malaysia Sarawak, Kota Samarahan, Malaysia

**Abstract**

In today's fast-paced world, it's crucial to access accurate and relevant information, especially in physics, where complex calculations are standard. Finding answers to physics questions can be difficult and time-consuming, particularly for those without a deep understanding of the subject. This paper proposes a novel solution called PhyQA - a question-answering system that utilizes ontology in the physics domain. By leveraging the power of ontologies to provide structured knowledge about physics concepts and their relationships, PhyQA can efficiently and accurately identify calculational questions. The development of this system was motivated by the need for a tool that could help bridge the gap between the complex calculations of physics and first-time learners' understanding. PhyQA provides clear and concise answers to physics questions, making it available to a broad audience, including students, researchers, and anyone interested in the subject. Using ontologies in this system provides a new technique for organizing and structuring knowledge, allowing for a more sophisticated representation of complex concepts. Ultimately, this paper emphasizes the importance of creating a question-answering system with ontology in the physics domain. It demonstrates its potential to revolutionize how we access and understand knowledge in this subject. By providing accurate and accessible answers to calculational questions, PhyQA has the potential to impact education and research in this field significantly.

**Keywords**: Question-answering system, physics, ontology, information retrieval, calculation.

## 1. Introduction

Developing an intelligent dialog system for increasing interest in studies exceeds human capabilities. This system should be able to handle engaging conversations and questions on a wide range of topics [1]. One popular tool that has emerged is a question-answering system (QAS), which has gained popularity in online education [2],[3],[4]. QAS has the potential to help students learn and complete their coursework independently. A Question-answer system can swiftly analyze extensive legal resources and provide answers or clarification to queries within seconds [5]. Due to the COVID-19 pandemic, many students have to switch to online learning platforms due to movement restrictions. This shift to virtual classrooms has raised concerns about the effectiveness of remote learning. Nevertheless, online education offers a wide range of benefits, including the convenient availability of various multimedia materials, access to current and global information sources, improved navigation features, and the ability for students to interact with experts in meaningful discussions [6]. Additionally, online learning provides flexible and cost-effective communication channels.

Question-answering systems (QAS) are applications of natural language processing (NLP) that are becoming increasingly popular. QAS aims to interpret and respond to inquiries from users in their native tongue [7].

Open-domain QAS have access to a vast corpus of textual data and can employ information retrieval techniques to locate relevant pieces of text that may contain the answer to the question [8]. Open-domain QAS answers a wide range of questions on any subject, as they are not limited to a specific domain or set of concepts; when it comes to open-domain QAS, the efficiency of generative models increased by carefully integrating text retrieval for supporting evidence [9]. However, closed-domain QAS has more constrained information and can only respond to queries under a domain or set of ideas. As a result, they are usually more specialized than open-domain QAS. Generally, people use it to address questions within a specific field or industry, such as medicine, law, or finance [10, 11,12]. The system consists of three modules: query processing, document processing, and answer processing [13]. The query processing module is a module that pre-processes the natural language question, which includes identification, classification, and reformulation. In QAS, the document processing module includes information retrieval (IR), which retrieves the information from the knowledge base or documents related to the natural language question. The answer-processing module extracts the final element from the previous module. Ideally, QAS can find all possible answers to every question. In reality, QAS is precise and accurate only when solving subject-specific questions. This issue has led many researchers to seek a solution, but most studies yield irrelevant results as most of the current QAS depend on content matching rather than the semantics of the answer. QAS is an acronym for Question Answering System, which provides solutions to theory-based questions like definitions, factoids, and descriptions. QAS analyses the input questions and extracts the subject, relationship, and object triples. However, the relationships between the entities in the questions are usually not further processed and need clarification. Ontology can enhance QAS by comprehending the words and sentence structure by providing additional information to an answer [14]. In the current study, the researchers use a closed domain ontology-based focusing on Physics, PhyQA.

The structure of this paper presents in the following manner: Section 2 presents related works for ontology-based QAS. Section 3 explains the PhyQA architecture, while Section 4 discusses the results. The last section provides a concluding statement.

## 2. Research Methodology

This research aims to fulfil the demand for a reliable question-answering system in secondary-level physics education in Malaysia. The main objective is to create PhyQA, a question-answering system that utilizes ontology-based methods, to enhance the recognition and utilization of appropriate physics formulas for calculation-oriented questions. Concurrently, this project aims to develop PhyQA, a formalized physics ontology designed to capture the semantics of Malaysia's Sijil Pelajaran Malaysia physics curriculum.

The importance of this research lies in its ability to address the challenges students encounter when attempting to comprehend physics principles and solve mathematical exercises. Furthermore, it seeks to assess and contrast the effectiveness of PhyQA with Fong and Bong's QAS (2017), closed domain question answering system in the physics field, in terms of their implications for promoting physics education. Ultimately, the primary goal of this study is to significantly improve physics instruction within the secondary educational framework in Malaysia. Thus, the ontology, PhyOnto, covers several physics learning units, including Measurement, Force, Motion I, Gravitation, Heat, Waves, Light and Optics, Electricity, Electromagnetism, Electronics, and Pressure [15].

In order to achieve the specified goals, this study is guided by a set of research inquiries:

1. How can the development of PhyQA, which is a question-answering system based on ontology, effectively enhance its ability to identify appropriate physics formulas for calculation-based questions among secondary students in Malaysia?
2. How does integrating PhyOnto, a contextual physics ontology, enhance the precision and accuracy of responses generated by PhyQA, especially when addressing calculation-oriented physics questions?
3. PhyQA and Fong and Bong's QAS have notable differences in their approach, encompassing distinct features and characteristics. Hence, these differences affect their effectiveness in meeting the specific needs of secondary students studying Physics in Malaysia.

The research questions presented here form the basis of our inquiry into enhancing physics education in Malaysia. We aim to comprehensively understand the contributions of PhyQA and PhyOnto to improved learning outcomes and assess their effectiveness relative to existing models, like Fong and Bong's QAS (2017). In essence, this study aims to develop PhyQA and PhyOnto tools to enhance the value of physics teaching by providing accurate solutions for calculation-based questions. The study will also evaluate their performance against Fong and Bong's QAS, with the ultimate objective of helping Malaysian secondary students.

## 3. Related Works

In the following, we highlight three existing QAS that are relevant and motivate us to the study, which are QAPD (ontology-based question-answering system in the physics domain), QAS (a hybrid QAS to support physics learning), and QUASE (ontology-based domain-specific natural language QAS [16,17,18].

QAPD and Fong and Bong's (2017) QAS aim to answer physics questions. Still, QAPD leverages the representational power of ontology better to understand the relationships between concepts in the domain, while Fong and Bong's (2017) QAS uses a more straightforward knowledge representation. QAPD, armed with its robust ontology, represents knowledge in the physics domain as a network of interconnected concepts, allowing it to make logical inferences and provide more accurate answers to complex questions. On the other hand, a more straightforward approach uses a simpler knowledge representation to answer questions [17]. Although it may not be as complex as QAPD, it is still a formidable opponent. It quickly and reliably answers various questions using a hybrid information retrieval approach.

On the other hand, QUASE is the wildcard in this trio, able to adapt to new domains and knowledge. QUASE makes it ideal for answering questions in various fields, including biology, but it may need help solving specific and complex questions in physics [18].

In terms of performance, QAPD has shown superior accuracy and efficiency compared to both QAS and QUASE, especially when addressing complex physics questions. However, QUASE is a more flexible architecture better suited for adapting to new domains and knowledge.

Current ontology-based QASs primarily address theoretical questions, such as wh-questions, Boolean questions, and others, as demonstrated by their evaluated work. Our goal is to design and develop PhyQA, a QAS capable of answering calculation questions in physics subject. Upon conducting a thorough analysis, we discovered that the existing physics ontology needs to be revised to meet the demands of solving calculation questions. Therefore, we aspire to create an ontology that effectively captures the semantic nuances of physics, incorporating its units and measurement tools.

## 4. PhyQA Architecture

The implementation of PhyQA involves the utilization of an ontology to facilitate automatic question-answering. Hence, it consists of four main modules: (1) Ontology Construction, (2) Question Pre-processing, (3) Candidate Answers Retrieval, and (4) Answer Formulation. The ontology serves as a knowledge base for PhyQA to recognize and understand physics concepts and their relationships. Figure 1 illustrates the architecture of PhyQA and its modules.

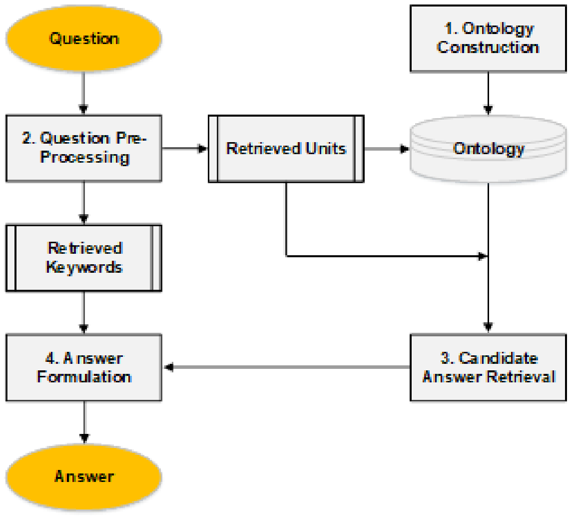


**Figure 1. The PhyQA architecture.**

Figure 1 depicts the process for solving calculation physics questions. Below, we outline each module.

**a. Ontology Construction**. This module created an ontology, PhyOnto, a knowledge base for physics concepts and their relationships. It identifies commonly used units in the field and is produced by analyzing physics textbooks. Classes, hierarchies, properties, and instances are defined.

**b. Question Pre-Processing**. This module extracts keywords and units from user inquiries using natural language processing. Employing techniques such as tokenization, stopword removal, lemmatization, stemming, and entity recognition, we identify standard units mentioned in the question and use them to find relevant formulas for solving the inquiry.

**c. Candidate Answers Retrieval**. This module uses the PhyOnto knowledge base to source the candidate's responses. We formulate a SPARQL query to identify physical quantities that can be measured using the retrieved units; this suggests relevant formulas or equations to solve the given question.

**d. Answer Formulation**. Our system selects the best answer using a formula matching the units. If multiple physical quantities match the units, we find the most relevant answer by identifying the physical quantities. We present the final formulas or equations to the user as potential solutions.

In the upcoming sections, we will thoroughly examine particular modules and provide a comprehensive review.

**4.1. Ontology Construction**

In physics knowledge representation, various ontology-based systems enhance understanding and information retrieval. One such system is QAPD, which utilizes the EAEONT ontology as its foundation for constructing a comprehensive knowledge repository. The EAEONT ontology focuses on domain theories, specifically electricity and electromagnetism within physics. Furthermore, the Physics Concept Ontology (PCO) captures fundamental physics ideas and their interactions [17]. Both EAEONT and PCO are notable for accommodating instances of independent entities, making them valuable tools for facilitating knowledge retrieval.

To effectively solve physics problems that involve calculations, particularly those that require the recognition of specific units and measurement tools, it is essential to develop a specialized ontology that aligns with the physics curriculum of Malaysia's SPM examination. In this case, PhyOnto becomes crucial as it is specifically designed to accommodate these requirements and contributes towards the objectives of PhyQA.

To establish a strong base, we analyzed multiple methodologies for ontology [19] and ultimately decided on implementing the "101 methods" [20]. This approach guides the creation of PhyOnto, an ontology designed to support PhyQA by capturing information about physics units and measurement instruments.

The initial and essential phase of ontology development involves establishing the scope of the ontology. In our study, we specifically examine the standards of the Malaysian SPM examination, comparable to GCSE or GCE Ordinary Level (O Level). This specification serves multiple purposes. Firstly, the Malaysia SPM examination plays a vital role in assessing students; understanding of physics concepts and ability to solve problems within the Malaysian educational system. Consequently, aligning our ontology with these examination standards ensures that our knowledge representation directly caters to the educational needs of secondary-level physics students in Malaysia.

Moreover, aligning with secondary school physics curriculum textbooks is backed by prior research. Textbooks serve as primary written sources of knowledge, and they frequently contribute significantly to educational research by describing differences in the understanding of learners across various levels of the education system [21]. Therefore, our ontology development process involves using these textbooks as the source of data [22, 23]. We aim to incorporate the basic concepts, theories, and principles emphasized in the official curriculum into our ontology, named PhyOnto. This approach aligns PhyOnto with structured learning pathways that students follow in their classrooms. In short, we have chosen to focus on Malaysia's SPM examination standard and rely primarily on textbooks because this examination holds a central position within the Malaysian education system. Textbooks are recognized as foundational educational references, making them an appropriate resource for our comprehensive and directly relevant PhyOnto catering to secondary students studying physics in Malaysia.

The development of ontology required a thorough examination of the content in the textbook to identify and extract relevant physics concepts along with their interconnections. These identified concepts and relationships were then methodically arranged and represented within the structure of the ontology.

We determined terms in PhyOnto by extracting and enumerating them from textbooks. We considered the extracted nouns as potential classes within the ontology and assigned verbs as properties. This approach aimed to ensure a systematic and well-organized representation of knowledge. Nouns typically represent entities or concepts that can be translated into classes in the ontology, while verbs describe relationships between these entities or concepts and fall under the category of properties. For example, the noun "units" represents a concept that aligns with the class structure of the ontology, while the verb "has unit" effectively captures the relationship between a physical quantity and its corresponding unit, functioning as a property.

**Table 1. The essential terms of PhyOnto.**

| Items | Important terms | Description with example |
|---|---|---|
| 1 | Unit | A unit is a standard quantity used to express measurements. Examples of units include meter per second (m/s), kilogram (kg), and second (s). |
| 2 | SI Unit | SI unit stands for International System of Units, which is the modern form of the metric system. Scientists and engineers commonly use SI units in their applications. Examples of SI units include meter (m), kilogram (kg), and ampere (A). It is usually a subclass of the Unit. |
| 3 | Physical quantity | A physical quantity is a measurable property of matter or energy. Physical quantities include length, mass, time, speed, and pressure. |
| 4 | Base quantity | A base quantity is an independent physical quantity that cannot be effectively present as other physical quantities. The SI system has seven base quantities: length (meter), mass (kilogram), time (second), electric current (ampere), temperature (kelvin), amount of substance (mole), and luminous intensity (candela). |
| 5 | Derived quantity | A derived quantity is a physical quantity expressed in one or more base quantities. Examples consist of speed (meter per second), acceleration (meter per second squared), and force (kilogram meter per second squared). |
| 6 | Measurement tool | A measurement tool is an instrument used to measure physical quantities. Examples of measurement tools include rule (length), balance (mass), stopwatch (time), thermometer (temperature), ammeter (electric current), and spectrophotometer (luminous intensity). |
| 7 | Application | An application is a specific use or purpose for a physical quantity. Examples of applications include: <br>● measuring the length of a piece of paper with a ruler <br>● measuring the mass of an object with a balance <br>● measuring the time it takes for an object to fall with a stopwatch. |

In PhyOnto, there is a total of four object properties which are "hasUnit," "hasSIUnit," "hasMeasurementTool," and "hasApplication." To describe the physics formula, it needs "hasUnit" to define the units used. For example, the force has a formula, "F=ma," where "F" stands for force, "m" stands for mass, and "a" stands for acceleration. The SI unit force is Newton (N), mass is the kilogram (kg), and acceleration is the meter per second square ($m/s^2$). To calculate the force, it needs the value of kg and $m/s^2$. We thus defined "force hasUnit kg" and "force hasUnit $m/s^2$". In other words, mass and acceleration are necessary to understand force.

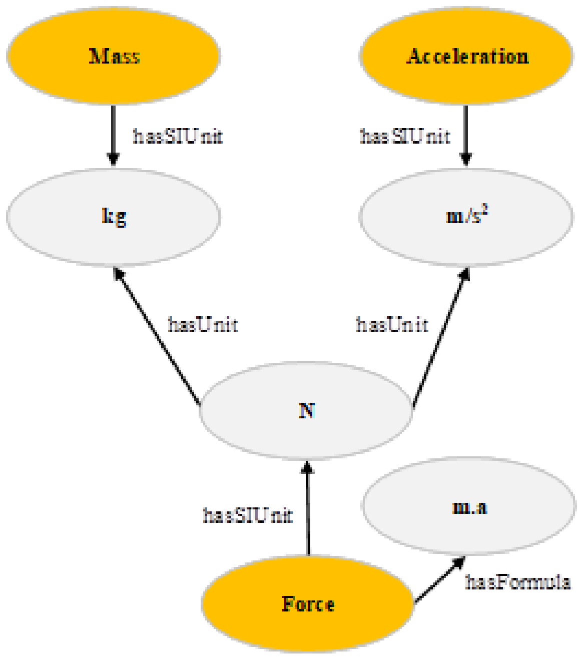


**Figure 2. The relationship between N, kg, and $m/s^2$.**

Additionally, PhyOnto uses OWL for illustration. PhyOnto consists of 9 classes, four object properties, 153 individuals, and 704 axioms. The PhyOnto can be found on GitHub[1]. The information in PhyOnto can be retrieved using SPARQL [23], the standard language for querying RDFS data.

### 4.2. Question Pre-Processing

We developed the question pre-processing module with a specific goal in mind to effectively handle different inquiries related to the field of physics. Its primary aim is to accurately respond to calculation-based questions that commonly emerge within this discipline. Queries involving numerical values and units require specific formulas for accurate solutions. By extracting pertinent information from such inquiries, this module facilitates the identification of relevant physics theories and enables the retrieval of precise answers.

The NLP techniques employed in the question pre-processing module of PhyQA include tokenization, stopword removal, lemmatization, stemming and entity recognition [25, 26].

This module begins by tokenizing the user's question, breaking it down into individual words or tokens. Following this, common non-essential words get removed through stopword removal. Subsequently, lemmatization is applied to reduce words to their base forms and stemming further truncates them to their stems. These procedures contribute towards standardizing and streamlining the text for analysis purposes consistently with an academic tone.

Furthermore, the module conducts entity recognition to ascertain customary units referred to in the inquiry. For instance, if the question pertains to "the velocity measured in meters per second", entity recognition would identify "meter per second" as a unit of measurement. These extracted keywords and units fulfil a requirement in identifying pertinent physics formulas and potential solutions for the query. Table 2 lists the examples of questions, extracted keywords and units.

**Table 2: Retrieved keywords and units.**

| | |
|---|---|
| **Question** | A car travels from a stationary position and reaches a velocity of 36 m/s in 8 seconds. |
| **Keywords** | 'car', 'travel', 'stationary', 'position', 'reach', 'velocity', '36', 'ms', '8', 'second', |
| **Standard Units** | 'm/s,' 'seconds' |

This module employs various natural language processing techniques to standardize and streamline the user's question for subsequent modules.

### 4.3. Candidate Answers Retrieval

The candidate answer retrieval module is a necessary component of the PhyQA as it extracts pertinent information from the ontology to solve physics questions that involve calculations. Algorithm 1 guides this module to identify units in user questions while connecting them with relevant physical quantities within the ontology [27].

The entire process begins within the question pre-processing module, where we get keywords and units from the user's question. These extracted elements are then smoothly communicated to the candidate answer retrieval module, where they undergo further processing and analysis. For example, when presented with a user's question such as "A car travels from a stationary position and reaches a velocity of 36 m/s in 8 seconds. What is the acceleration of the car?" this module effectively utilizes advanced natural language processing techniques to identify and accurately extract units like "m/s" and "seconds".

The following Algorithm 1, acts as the underlying mechanism for recognizing units based on ontology. It works in conjunction with a comprehensive physics ontology, smoothly navigating through its collection of knowledge to establish links between the retrieved units and their respective physical quantities. By employing a dynamic approach, even units expressed in various formats like "m/s" or "$ms^{-1}$" can be effectively connected to their corresponding physical quantities within the ontology.

[1] https://github.com/CatherineHsj/PhyOnto.git

**Algorithm 1. Units Recognition**

Input: Retrieved units, Ontology.

Output: Label retrieved units as physical quantities.

1. Let U be the retrieved units from the question pre-processing module.
2. Let Q be the physical quantity with the SI unit of U.
3. Let N be the number of the retrieved units.
4. Set l=0
5. For each unit U in the retrieved units N:
   i. Increment l by 1.
   ii. Search for the physical quantity Q in the ontology using the first **SPARQL [1]** query.
   iii. If the Q is present in the ontology, assign the corresponding label to U and jump to step 6.
   iv. If Q is not in the ontology, jump to step 6 and iterate until l < N.
6. Output the labelled retrieved units.

By Algorithm 1, the query defined as "SPARQL[1]" can be defined as follows:

$$SELECT\ ?\,x\ WHERE\ \{?\,x\ : hasSIUnit\ ?\,y.\,?\,y\ rdfs{:}seeAlso\ ?\,z.FILTER\ (?\,z = \ '" + item + "')\}ORDER\ BY\ ?\,x$$

**SPARQL[1]**

In the query above, the variable '?x' is selected for physical quantities with the SI unit 'y', and '?y' is linked to another resource '?z' through the 'rdfs:seeAlso' property. The value of '?z' is dynamically inserted as an 'item' in the query to retrieve physical quantities with an SI unit equivalent to the retrieved units. The query matches the value of '?z' with the 'rdfs:seeAlso' property of the corresponding unit resource to account for different unit representations. This query enables the algorithm to retrieve physical quantities with an SI unit equivalent to the retrieved units, even if the representation varies. For example, one can write the meter per second as 'm/s' or 'ms-1'. The results are then sorted by physical quantity using the 'ORDER BY' clause.

Utilizing the question pre-processing module, we obtained two units: "m/s" and "seconds" from the user's query. The estimated research into an ontology of the system proclaimed which physical quantities are related to these specific SI units. Following the successful retrieval of this relevant information, labels corresponding to these physical quantities enhance our access to valuable insights from the ontology's extensive knowledge repository.

In this particular situation, the labels obtained encompassed "Speed/Velocity", denoted by the units "m/s"; "Time" represented by the unit "seconds". With this vital information, the system effortlessly progressed to the next module, seamlessly combining these recognized units with their corresponding physical quantities. This integration creates a fundamental basis for generating accurate and contextually appropriate responses to complex physics queries, enhancing PhyQA's overall effectiveness.

Besides, in the following, this second SPARQL retrieved the physical quantity, measured using the retrieved units. The main intention is to find formulas to solve the question problem. The SPARQL query appears as shown below:

$$SELECT\ ?\,x\ WHERE\ \{?\,x\ : hasUnit\ ?\,y.\,?\,y\ rdfs{:}seeAlso\ ?\,z.FILTER\ (?\,z = \ '" + item + "')\}ORDER\ BY\ ?\,x$$

**SPARQL[2]**

The SPARQL query selects the variable '?x' for physical quantities with the unit '?y' and unit '?y' is linked to another resource '?z' through the 'rdfs:seeAlso' property. The value of '?z' is dynamically inserted as an 'item' in the query to retrieve physical quantities that have a unit equivalent to the retrieved units. The results are then sorted by physical quantity using the 'ORDER BY' clause.

The objective of this query is to retrieve the physical quantities that can be measured using the retrieved units, which can be applied to identify relevant formulas. The query can recognize the physical quantities compatible with the retrieved units by linking the units to their corresponding physical quantities using the 'hasUnit' property. These physical quantities can then suggest relevant formulas or equations.

Referring to the example in Table 2, the retrieved units such as 'm/s' can be used to measure several physical quantities, such as acceleration, centripetal acceleration, centripetal force, and first linear motion, among others. The results of the second SPARQL query are passed to the answer formulation module to determine the most relevant answer to the user's question. This module analyses the user's query and the retrieved physical quantities and formulas to identify the most suitable solution for the user.

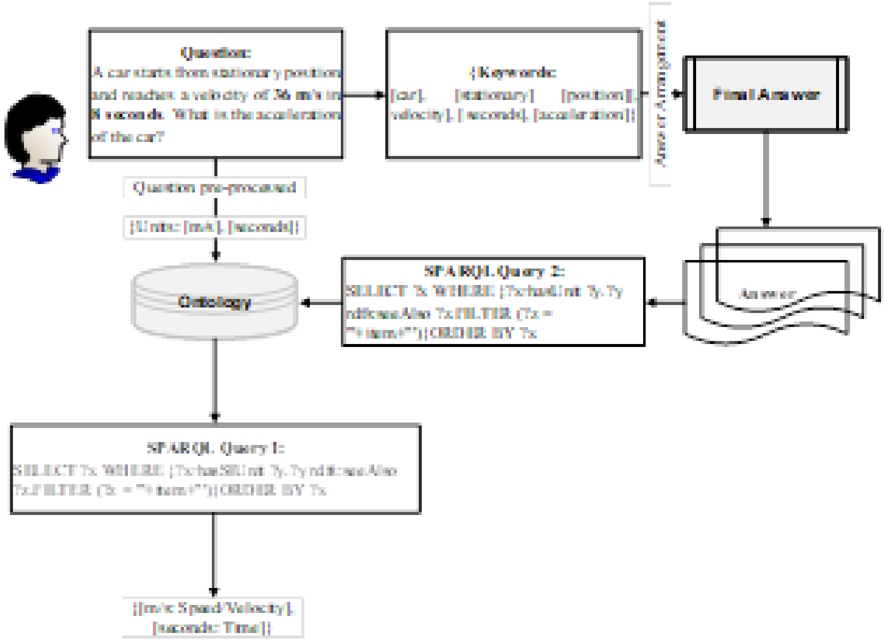


**Figure 3. The process of PhyQA on retrieving the answers from ontology using SPARQL queries.**

The utilization of SPARQL queries in PhyQA, as depicted in Figure 3, showcases the navigation of our ontology, known as PhyOnto, to retrieve potential answers. While this diagram provides a complete outline of the process, a clear understanding of the underlying algorithms and methodologies is essential to retrieve answers accurately and in an appropriate global manner. In essence, PhyQA's module for retrieving candidate answers plays a critical role in delivering accurate physics responses that are relevant within the aid; having a comprehensive grasp of the system's mechanisms is essential for optimal utilization.

### 4.4. Answer Formulation

This section provides a detailed explanation of the process through which PhyQA offers accurate responses to inquiries about physics.

From a set of physical quantities and their corresponding formulas obtained from the SPARQL[2], an intersection operation assists in identifying the optimal answer. However, in certain circumstances, if only one physical quantity matches the retrieved units, the intersection process is unnecessary, and the corresponding formula is the answer.

**a. Single answer**

If only one physical quantity matched the retrieved units, the corresponding formula can be directly considered as the answer. In this case, PhyQA can immediately provide the final formula to the user for solving the questions without further intersection or comparison with other formulas.

**b. Multiple possible answers**

If multiple physical quantities match the retrieved units, using the intersection process to identify the most relevant solution, the intersection between the results for different units is defined as the set of physical quantities with both units. For example, an intersection between the results for "m/s" and "seconds" can be defined as the set of physical quantities that have the unit "m/s" and also the unit "seconds." It appears in the query as **m/s** ∩**seconds**={physics:physicalQuantity(hasUnit)∈ **m/s** and physicalQuantity(hasUnit)∈**seconds**}. There are two possible results after the intersection: one final answer or many final answers.

**i. One Final Answer**

In some cases, the intersection process may result in a single physical quantity that satisfies the criteria, and the corresponding formula will be the final answer. PhyQA can then present this formula to the user as the solution to the question without further comparison.

**ii. Many Final Answers**

In other cases, the intersection process may result in multiple physical quantities that satisfy the criteria. After completing the intersection process and identifying potential physical quantities, PhyQA formulates answers depending on the keywords extracted from the user's query (see also 3.2). PhyQA then proposes these answers to the user as possible solutions, which may include the final formula for solving the question. This step further enhances the system's utility as an educational tool, as it provides answers and promotes understanding and learning by guiding users toward the appropriate formulas and equations for problem-solving.

PhyQA's approach of leveraging an ontology-based knowledge base, utilizing an intersection process, and providing final problem-solving formulas showcases its potential as a promising question-answering system for physics education. The feature of presenting the final formula to users enhances the system's utility as an educational tool by promoting understanding, learning, and active engagement in problem-solving. It empowers users to apply their knowledge and skills in physics, enhancing their critical thinking and problem-solving abilities while reinforcing their understanding of physics concepts. Moreover, by fostering self-directed learning and enabling users to develop a stronger foundation in physics, PhyQA adds value as an educational aid. The following section will further discuss the results of PhyQA.

## 5. Results and Discussion

### 5.1. Gold Standard

The proposed QAS is benchmarked and evaluated against a gold standard comprising 50 calculation questions and their corresponding answers. We selected these questions manually from the textbooks, which we used as the primary source. The 50 questions cover a scope of topics in Physics, and you can discover many related questions in the Appendixes. Although we have attempted to obtain as many questions as possible to cover the topic spectrum, it is crucial to note that not all topics delivered the same number of calculation questions, and some provided only factoid questions. Therefore, we only include questions that demand calculation.

### 5.2. Evaluation Metrics

The evaluation metrics are typically calculated based on the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) generated by the system. TP refers to the number of questions for which the system generated a correct answer, TN signifies the number of remaining queries that the system accurately identified as unanswerable, FP refers to the number of questions for which the system generated an incorrect answer, FN refers to the number of questions for which the system failed to create any solution. Table 3 shows the results of PhyQA in the confusion matrix.

**Table 3. Table of Confusion of PhyQA.**

| | True | False |
|---|---|---|
| Positive | 38 | 12 |
| Negative | 0 | 10 |

The absence of true negatives in Table 3 is due to the nature of the task evaluated, specifically, question-answering. PhyQA attempts to generate a single answer for each question, which is the formula for the question, and the evaluation depends on how well the generated answer corresponds to the correct answer. Since there is only one answer for each question, any different response is considered a false positive, and any solution the system fails to generate is supposed to be a false negative. Therefore, there are no true negatives in question answering evaluation.

From Table 3, out of the total questions evaluated, there were ten false negatives because some require diagrams or graphs to be solved. However, we still included these questions in the evaluation since they required formula and mathematical skills. Furthermore, our evaluation did not exclude difficult questions, including those that need higher-order thinking (HOT) skills, and some of these questions did not directly provide the units or required the use of more than two formulas to solve. While this approach allowed us to test the PhyQA's ability to handle more complex questions, it also contributed to false negatives.

We report accuracy, precision, recall, and F1 score to evaluate the system's ability to render correct answers. Accuracy measures the overall correctness of the system's answers, calculated as the ratio of correctly answered questions to the total number of questions in the test set. Precision measures the proportion of the system's correct answers, calculated as the ratio of correctly answered questions to the total number of questions the system answered. Recall measures the proportion of the right solutions retrieved by the system, calculated as the ratio of correctly answered questions to the total number of correct answers in the test set. The F1 score is a harmonic mean of precision and recall. It provides a balanced measure of the system's performance. Table 4 shows the outcome of the evaluation.

Comparing PhyQA with Fong and Bong's QAS is relevant because it benchmarks against an existing Physics QAS. Fong and Bong's QAS is a QAS that was developed for physics education and aimed to answer five types of questions. Table 4 displays the results.

**Table 4. Accuracy, Precision, Recall, and FI score of PhyQA**

| Evaluation Metric | PhyQA | Fong and Bong |
|---|---|---|
| Accuracy | 0.450 | 0.012 |
| Precision | 0.514 | 0.014 |
| Recall | 0.782 | 0.083 |
| F1 Score | 0.620 | 0.024 |

**5.3. Advantages**

From Table 4, we can see that PhyQA achieved an accuracy of 0.45, precision of 0.51, recall of 0.78, and F1 score of 0.62. When comparing these metrics against Fong and Bong, we can see that PhyQA outperforms Fong and Bong in accuracy, precision, and recall. However, it's worth noting that PhyQA's F1 score of 0.62 is higher than Fong and Bong's QAS F1 score of 0.02, indicating better overall performance regarding a balance between precision and recall.

For example, when posed with a question such as "A car travels from a stationary position and reaches a velocity of 36 m/s in 8 seconds. What is the acceleration of the car?", PhyQA can provide the appropriate formula for acceleration based on the units given in the question which are "m/s" and "seconds." This capability to identify and utilize relevant formulas enhances its ability to solve Physics calculation questions. In contrast, Fong and Bong's QAS returns a list of theories related to car crashes, indicating a limitation in its ability to provide accurate and relevant answers for physics calculations. Table 5 compares the solutions, highlighting the differences in their performance in answering physics calculation questions.

**Table 5. Answers of PhyQA and Fong and Bong's QAS solve the question: "A car travels from a stationary position and reaches a velocity of 36 m/s in 8 seconds. What is the acceleration of the car?"**

| Question | A car travels from a stationary position and reaches a velocity of 36 m/s in 8 seconds. What is the acceleration of the vehicle? |
|---|---|
| PhyQA | Given unit(s):<br>seconds: Time<br>m/s: Speed/Velocity<br>Expected Formula:<br>Acceleration<br>Acceleration (m $s^{-2}$) = (Final velocity (m/s) - Initial velocity (m/s) / Time for the velocity change (s) |
| Fong and Bong | Car manufacturers must perform safety tests on each newly invented car before marketing. Below is a crash test executed by a car manufacturer :<br>a) crash<br>the impact of the crash starts chemical reactions which produce nitrogen gas that quickly inflates the airbag.<br>(b) inflate<br>the fully inflated airbag absorbs the initial impact of the dummy driver by pushing forward due to its inertia.<br>(c) safe<br>the dummy gets protection from sharp objects by cushioning. When gas escapes from the holes behind the airbag, the bag collapses. |

Table 5 compares the answers generated by both QAS in response to the question. PhyQA was able to accurately identify the relevant units of measure and provide the correct formula to solve the problem, while Fong and Bong provided an utterly unrelated answer. This comparison highlights the advantages of using PhyOnto for solving physics calculation questions, as it demonstrated a much higher accuracy and relevance than Fong and Bong.

One advantage of PhyQA is that it provides a structured and standardized approach to answering physics calculation questions, which can be practical for educational purposes. The high recall value of PhyQA suggests that it effectively identifies relevant answers. Still, the relatively low precision value may also include some irrelevant answers. We can enhance this through additional advancement of the system. Nevertheless, PhyQA's performance metrics demonstrate its potential as a reliable and effective QAS for physics education, surpassing the performance of Fong and Bong, as shown in the comparison of the evaluation metrics.

**5.4. Limitation**

However, a limitation of PhyQA is that it relies on the availability and accuracy of the underlying knowledge base and may need to perform better on questions that require more complex reasoning or interpretation. The system has the potential to yield false positive answers when dealing with situations where two or more physical quantities in a formulation share identical units. For example, in the question: "A ball of plasticine weighing 0.058 kg is thrown and strikes a wall at a speed of 10 m/s, sticking to the wall, what is the impulse of the plasticine? Thus, the system answered "{'momentum', 'centripetal force', 'principle of conversion of momentum', 'wavelength', 'impulsive force'}". Meanwhile, momentum, centripetal force, the principle of conversion of momentum, wavelength, and impulsive force have mass and velocity in the formula. These five results are the results of the intersection between units "kg" and "m/s."

Besides, some calculation questions involved diagrams, graphs, or charts. PhyQA cannot solve this type of calculation question. It has no image processing capacity to extract the parameters. For instance, PhyQA could not solve the question "The graph above shows the force acting on a trolley of 5 kg mass over a distance of 10 m. Find the work done by the force to move the trolley." because some of the required information was in the graph.

In addition, the system cannot solve questions requiring HOT skills. One of the questions is, "Ranjit runs up a staircase of 35 steps. Each step is 15 cm in height. Given that Ranjit's mass is 45 kg, find the work done by Ranjit to reach the top of the staircase.". The question provided 35 steps, 15 cm, and 45 kg to calculate the work done. The work done formula is W = fd (W is work done with SI unit J; f is force with SI unit N; and d is the displacement with SI unit m). However, the system could not calculate the question using this formula because the question did not provide the value of N. The calculation for the quantity of N is necessary. The formula for force is F=ma. However, the question did not give the acceleration in this formula. Simultaneously, another formula to determine the quantity of N uses the weight formula, W=mg (W is the weight with SI unit N; m is mass with SI unit kg; g is the acceleration of gravity with SI unit m/s2). The acceleration of gravity is a constant value of 9.8 m/s2. The weight formula and the value of gravity acceleration are the techniques to solve the question.

PhyQA is one of many question-answering systems designed for physics education. For example, Fong and Bong (2017) developed a hybrid QAS based on ontology and topic modelling. However, while both systems aim to assist students in solving problems, their approaches are different. Unlike PhyQA, which identifies relevant physical quantities and units in a question, Fong and Bong retrieve keywords only from calculation questions and do not identify units. As a result, their system may struggle with questions involving multiple quantities with the same units, which PhyQA can address through its intersection process.

**6. Conclusion and Future Works**

PhyQA represents a pioneering and innovative approach to addressing physics as it utilizes the capability of an ontology-based knowledge base, PhyOnto, specifically designed in this research to fulfil the requirements of solving calculation questions. The PhyQA system comprises four modules: ontology construction, question pre-processing, candidate answer retrieval, and answer formulation. The candidate answer retrieval module works with PhyOnto to retrieve the most relevant answers to the user's question. Based on the evaluation metrics, PhyQA had relatively low accuracy, correctly answering only 45% of the questions. However, it has a moderate precision of 51%, which suggests that when it delivered an answer, it was accurate more than half of the time. The recall was relatively high at 78%, indicating that the system could identify many correct answers but created ten false negatives. The F1 score, which combined measures of precision and recollection, was 0.61, indicating that the system performed moderately well overall. In conclusion, it is clear that, at its current stage, Fong and Bong's QAS is not sufficiently advanced to serve as a substitute for a dedicated physics tutor.

In this research, we identify, describe, and analyze limitations in the PhyQA system despite its promising results. The researchers predict that PhyQA will enhance with advanced image processing techniques and merge with other QASs to expand its capabilities, and this means it can answer a vast range of physics questions, not just calculations. This step will contribute to advancing PhyQA towards fulfilling its potential as a comprehensive physics question-answering tool, offering users a more streamlined and effective means of accessing the necessary knowledge.

**Authors Contributions:**

Both authors contributed equally

**Funding:**

This research received no external funding.

**Conflicts of Interest:**

The authors declare no conflict of interest.

**References**

[1] Mahmood1, M. Z. ·. W. E. Z. ·. Q. Z. S. ·. A., 6 September 2022. Conversational question answering: a survey.

[2] Agarwal, A., Sachdeva, N., Yadav, R. K., Udandarao, V., Mittal, V., Gupta, A., & Mathur, A. (2019, May). EDUQA: Educational domain question answering system using conceptual network mapping. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8137-8141). IEEE.

[3] Lende, S. P., & Raghuwanshi, M. M. (2016, February). Question answering system on education acts using NLP techniques. In 2016 world conference on futuristic trends in research and innovation for social welfare (Startup Conclave) (pp. 1-6). IEEE.

[4] Liang C. (2012) Improved Intelligent Answering System Research and Design. In: Wu Y. (eds) Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009). Advances in Intelligent and Soft Computing, vol 116. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-11276-8_7

[5] Martinez-Gil, J., 2023. A survey on legal question–answering systems. Computer Science Review, Volume 48.

[6] Yuhanna, I., Alexander, A., & Kachik, A. (2020). Advantages and disadvantages of Online Learning. Journal Educational Verkenning, 1(2), 13-19.

[7] Pundge, A. M., Khillare, S. A., & Mahender, C. N. (2016). Question answering system, approaches and techniques: a review. International Journal of Computer Applications, 141(3), 0975-8887.

[8] Zhu, F., Lei, W., Wang, C., Zheng, J., Poria, S., & Chua, T. S. (2021). Retrieving and reading: A comprehensive survey on open-domain question answering. arXiv preprint arXiv:2101.00774.

[9] Izacard, G., & Grave, E. (2020). Leveraging passage retrieval with generative models for open domain question answering. arXiv preprint arXiv:2007. 01282.

[10] Haendel, M. A., Chute, C. G., & Robinson, P. N. (2018). Classification, ontology, and precision medicine. New England Journal of Medicine, 379(15), 1452-1462.

[11] Obeid, C., Lahoud, I., El Khoury, H., & Champin, P. A. (2018, April). Ontology-based recommender system in higher education. In Companion Proceedings of the The Web Conference 2018 (pp. 1031-1034).

[12] Vegesna, A., Jain, P., & Porwal, D. (2018). Ontology based chatbot (for e-commerce website). International Journal of Computer Applications, 179(14), 51-55.

[13] Allam, A. M. N., & Haggag, M. H. (2012). The question answering systems: A survey. International Journal of Research and Reviews in Information Sciences (IJRRIS), 2(3).

[14] Bouziane, A., Bouchiha, D., Doumi, N., & Malki, M. (2015). Question answering systems: survey and trends. Procedia Computer Science, 73, 366-375.

[15] Abdi, A., Idris, N., & Ahmad, Z. (2018). QAPD: an ontology-based question answering system in the physics domain. Soft Computing, 22(1), 213-230.

[16] Abdi, A., Idris, N., & Ahmad, Z. (2018). QAPD: an ontology-based question answering system in the physics domain. Soft Computing, 22(1), 213-230.

[17] Fong K.S. and Bong C.H. (2017). A hybrid question answering system based on ontology and topic modeling. Journal of Telecommunication, Electronic and Computer Engineering, 9(2-10), 151-158

[18] Mishra, V., & Khilwani, N. QUASE: AN Ontology-Based Domain-Specific Natural Language Question Answering System.

[19] Iqbal, R., Murad, M. A. A., Mustapha, A., & Sharef, N. M. (2013). An analysis of ontology engineering methodologies: A literature review. Research journal of applied sciences, engineering and technology, 6(16), 2993-3000.

[20] Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.

[21] Oakes, J., & Saunders, M. (2004). Education's most basic tools: Access to textbooks and instructional materials in California's public schools. Teachers College Record, 106(10), 1967-1988.

[22] Choy, C. S., Chuan, K. K., Beng, O. H., Md Mustafa, M. K. A., & Ragavan, R. (2019). Physics Form 4. Sasbadi Sdn Bhd.

[23] Chuan, K. K., Choy, C. S., Bongkek, N. R., Kasron, J., Md Mustafa, M. K. A., & Chakrabarty, P. K. (2020). Physics Form 5. Penerbit Bestari Sdn Bhd.

[24] Prud'hommeaux, E., & Seaborne, A. (2008). SPARQL query language for RDF. World Wide Web Consortium. Retrieved from https://www.w3.org/TR/rdf-sparql-query/

[25] Elnozahy, W. A., El Khayat, G. A., Cheniti-Belcadhi, L., & Said, B. (2019). Question Answering System to Support University Students' Orientation, Recruitment and Retention. Procedia Computer Science, 164, 56-63.

[26] Jurafsky, D., & Martin, J. H. (2000). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall.

[27] Zamani, H., Dumais, S., Craswell, N., Bennett, P. and Lueck, G., 2020, April. Generating clarifying questions for information retrieval. In Proceedings of the web conference 2020 (pp. 418-428).