



Review

# Question Classification for Intelligent Question Answering: A Comprehensive Survey

Hao Sun <sup>1</sup>, Shu Wang <sup>2,\*</sup> , Yunqiang Zhu <sup>2,3</sup>, Wen Yuan <sup>2</sup> and Zhiqiang Zou <sup>1,4</sup> 

<sup>1</sup> College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 1221045412@njupt.edu.cn (H.S.); zouzq@njupt.edu.cn (Z.Z.)

<sup>2</sup> Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; zhuyq@reis.ac.cn (Y.Z.); yuanw@reis.ac.cn (W.Y.)

<sup>3</sup> Jiangsu Center for Collaborative Innovation on Geographic Information, Nanjing 210023, China

<sup>4</sup> Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

\* Correspondence: wangshu@igsnrr.ac.cn

**Abstract:** In the era of GeoAI, Geospatial Intelligent Question Answering (GeoIQA) represents the ultimate pursuit for everyone. Even generative AI systems like ChatGPT-4 struggle to handle complex GeoIQA. GeoIQA is domain complex IQA, which aims at understanding and answering questions accurately. The core of IQA is the Question Classification (QC), which mainly contains four types: content-based, template-based, calculation-based and method-based classification. These IQA\_QC frameworks, however, struggle to be compatible and integrate with each other, which may be the bottleneck restricting the substantial improvement of IQA performance. To address this problem, this paper reviewed recent advances on IQA with the focus on solving question classification and proposed a comprehensive IQA\_QC framework for understanding user query intention more accurately. By introducing the basic idea of the IQA mechanism, a three-level question classification framework consisting of essence, form and implementation is put forward which could cover the complexity and diversity of geographical questions. In addition, the proposed IQA\_QC framework revealed that there are still significant deficiencies in the IQA evaluation metrics in the aspect of broader dimensions, which led to low answer performance, functional performance and systematic performance. Through the comparisons, we find that the proposed IQA\_QC framework can fully integrate and surpass the existing classification. Although our proposed classification can be further expanded and improved, we firmly believe that this comprehensive IQA\_QC framework can effectively help researchers in both semantic parsing and question querying processes. Furthermore, the IQA\_QC framework can also provide a systematic question-and-answer pair/library categorization system for AIGCs, such as GPT-4. In conclusion, whether it is explicit GeoAI or implicit GeoAI, the IQA\_QC can play a pioneering role in providing question-and-answer types in the future.

**Keywords:** Intelligent Question Answering (IQA); GeoAI; Question Classification (QC); IQA\_QC framework; evaluation metrics; literature review



**Citation:** Sun, H.; Wang, S.; Zhu, Y.; Yuan, W.; Zou, Z. Question Classification for Intelligent Question Answering: A Comprehensive Survey. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 415. <https://doi.org/10.3390/ijgi12100415>

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 25 July 2023

Revised: 6 October 2023

Accepted: 9 October 2023

Published: 10 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intelligent Question Answering (IQA), an important part of artificial intelligence, is a technology that enables computers to understand natural language and determine answers to questions intelligently [1–5]. Especially in geoscience, the high demand for an IQA system is due to the complexity and diversity of geographical questions and types. Question classification is the key to the IQA system, which is the core mission of both explicit and implicit GeoAI. The IQA system has the ability to accurately understand the user's query intention and query knowledge, so that the computer can completely replace the general query system (requires manual interaction for many times), directly interact with human beings and feedback knowledge to answer to the user. The IQA

system contains two core steps: a semantic parsing step for identifying the user questions as corresponding fixed queries and an answer query step for obtaining accurate answers from the database with the fixed queries. The IQA\_QC framework has a great influence on accuracy and performance of the IQA system, whether it is to obtain a fixed type of question in the semantic parsing step or an accurate answer in answer query step. Therefore, the research of the IQA\_QC framework makes sense in both aspects of IQA and information science development [6–8].

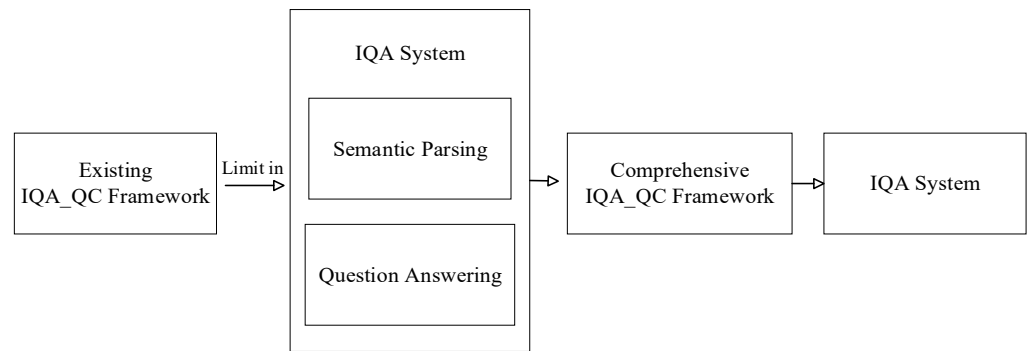
At present, the question classification mainly consists of four IQA\_QC frameworks from different perspectives: content-based [3,9], template-based [10–12], calculation-based [13,14] and method-based [15,16] classification. From the perspective of content, questions can be divided into querying different facts like when, where and who. From the perspective of template, questions can be divided into multi-hop questions and single-hop ones. From the perspective of calculation, questions can be divided into different conditions. From the perspective of method, questions can be divided into different processes of IQA including semantic parsing and question querying (specific descriptions and discussions are shown in Section 2). The above four classifications can guide semantic parsing and question querying by choosing appropriate approaches or models. But the above four classifications are not ideal, as there are still some defects and limits. Specifically, it is necessary to set the content of the classification to query the object and to set up different templates and calculations and to investigate deeper questions. That is to say, the first three are all based on the operational level. But the operational level involves content and depth. Therefore, a more comprehensive classification framework is needed. Although the last one, method-based classification, is not based on the operational level, it cannot provide a systematic and detailed category due to its broad classification. Overall, the four classifications are classified from different aspects, which have certain reference value for other researchers [1,15,17–19], but there is not yet a comprehensive system to systematically support IQA semantic parsing and question answering assistance. Therefore, a hierarchy-based classification, which consists of three levels, is proposed to address the above problems.

Next, we will describe two aspects including the motivations and organization of the paper. Motivations mainly describe the reason for constructing a comprehensive IQA\_QC framework. Organization mainly describes the structure of this paper.

### 1.1. Motivations

In the IQA system, especially in the geographical field, the question handling is rather crude, and in most cases only the general semantic analysis to related questions occurs [20–22]. This paper reviewed the existing IQA\_QC frameworks and found that these classifications were not sufficient to guide the IQA system in the aspects of semantic parsing and question querying. In view of these problems, a comprehensive and systematic IQA\_QC framework is required to guide the IQA system both on the semantic parsing step and answer query step. In summary, compared with the existing frameworks, we proposed a hierarchy-based classification, covering not only the types of questions in common classifications but also the types of questions that are missing. Moreover, compared with common metrics, evaluation metrics of broader dimensions are proposed, which would help measure geographical domain questions. Figure 1 presents a block diagram of the IQA\_QC framework. In this paper, we construct the IQA\_QC framework starting from the geographical questions. Due to the diversity of division angles in spatial and temporal scales in the geographical field, if our framework could cover the questions in the geographical field, it could help answer geographic-related questions for geography users.

In addition, the core contribution of this paper, a hierarchy-based IQA\_QC framework, on the one hand can provide the guidance for the IQA system especially for GeoIQA; on the other hand, it can provide a training corpus for generative AI systems like ChatGPT in certain fields like geoscience.



**Figure 1.** Block diagram of IQA\_QC framework.

### 1.2. Organization

The remainder of this paper is organized as follows: Section 2 reviews the existing IQA\_QC framework proposed by different researchers. Section 3 describes our hierarchy-based IQA\_QC framework in detail. Section 4 proposes the IQA evaluation metrics of broader dimensions by using the IQA\_QC framework. The relevant analysis and limitations are discussed in Section 5. Finally, the conclusion and future direction are presented in Section 6.

## 2. Existing IQA\_QC Frameworks

Question classification is a significant issue for the IQA system, since it can guide IQA to choose which approach and model to use to answer the question in a more efficient way. The existing IQA\_QC framework contains four directions, content-based, template-based, calculation-based and method-based, as shown in Table 1. In this section, the four kinds of classification will be summarized and illustrated in detail.

**Table 1.** Four main classifications.

Question Classification of the IQA System	Explanation
Content-based	This classification is based on the different content of the questions such as factoid, confirmation and so on.
Template-based	This classification is based on the different templates of the questions such as question-to-fact, fact-to-answer and question-to-answer (fact).
Calculation-based	This classification is based on the different calculation processes of the questions such as numerical comparison and numerical condition.
Method-based	This classification is based on the different methods of the questions such as information retrieval-based methods and semantic parsing-based methods.

### 2.1. The Content-Based Classification

According to different content, questions can be further divided into different types which includes factoid, confirmation, definition and so on [3,9,23]. The specific details are mentioned in Table 2.

Content-based classification is mainly based on the results of semantic parsing to query the object. Thus, the main advantage of content-based classification is strong pertinence. That is to say, if the content is determined according to the query, the corresponding answer can be directly acquired according to the entity and relationship of the content. Furthermore, complex natural language processing to extract answers is not needed which guarantees the time complexity of the system.

**Table 2.** Classification based on the content.

Sub-Type	Explanation
Factoid (when/who/where)	Questions that essentially require a single fact or a small piece of text to be determined as an answer.
Confirmation (yes/no)	Questions that require a yes/no answer.
Definition	Questions that require answers that are definitions of terms.
Causal	Questions that require that the answer should be one or more consequences of a fact.
Procedural	Questions that require a set of actions needed for accomplishing something.
Comparative	Questions that require, as an answer, a set of differences between two or more subjects.
Examples	Questions that aim to find examples that best describe the reference point of the question.
Opinionated	Questions that require to find the opinion of someone about a subject or a fact.

However, too much pertinence will lead to countless kinds of questions and poor accuracy of semantic parsing types. Also, the types of questions that can be directly answered are limited, and complex questions (such as those involving reasoning and calculation) cannot be answered [15].

In general, we can refer this principle to the fine level, because this detailed classification allows for precise judgment at the end of semantic parsing. However, this classification only focuses on the operation level, which is too fragmentary and leads to great trouble in semantic parsing. The limitation determines that the classification is not comprehensive. Further improvement is needed to supplement more question types through variety (e.g., Sections 2.2 and 2.3).

## 2.2. The Template-Based Classification

According to the query object, questions can be divided into different templates, including “question-to-fact”, “fact-to-answer” and “question-to-answer” [10–12,24]. For example: through the following templates <Jackie Chan, star, New Police Story> <New Police Story, Director, Mu-sung Chan> [25], we reasoned that Mu-sung Chan is a correct answer if the question is “Who directed New Police Story starring Jackie Chan?”. The details are shown in Table 3.

**Table 3.** Classification based on the template.

Sub Type	Explanation
Question-to-Fact	Some questions need to make up the gap between concepts in the question and the core fact.
Fact-to-Answer	Some questions need to make up the gap between concepts in the core fact and the answer choices.
Question-to-Answer (Fact)	Some questions need additional knowledge to connect concepts in the question to the answer, based on the core fact.

The application of these different templates is mainly to solve the question of reasoning in a Knowledge Graph (KG) which requires more than one template. The superiority of the template-based type is the chain approach, which uses transitivity to solve implicit questions [26]. This type of classification makes good use of the derivability of KG and utilizes templates (triplets) to deduce the answer to the deep questions.

The characteristic of a KG makes this classification depend on the richness of graph, while the content and branches of the knowledge graph are not necessarily enough to reason in many cases. For example, the question “Can you recommend some films like The Shawshank Redemption?” cannot be reasoned via one or more templates in the KG if there is lack of a rich-enough KG.

Thus, this type of question could be guided to improve the former classification and supplement the operational level to some extent. However, this classification is limited in that it only focuses on the reasoning of the KG, and other kinds of questions which need semantic parsing or text understanding are ignored. So, this classification should take account of more sophisticated questions.

### 2.3. The Calculation-Based Classification

Considering the weaknesses of content-based and template-based classification, calculation-based classification was proposed. The calculation-based classification divides questions into numerical comparison and numerical condition [13,14]. For example, for numerical comparison question “What is the second longest river in the world?”, the calculation process is to compare the lengths of rivers. And for the numerical condition question “How many age groups make up more than 7% of the population?”, the calculation process needs to know which age groups make up more than 7% of the population to count the group number to acquire the answers. The specific details are shown in Table 4.

**Table 4.** Classification based on the calculation.

Sub-Type	Explanation
Numerical Comparison	The answers to the questions can be directly obtained via performing numerical comparison in the documents, such as sorting and comparison.
Numerical Condition	The answers to the questions cannot be directly obtained through simple numerical comparison in the documents.

On the basis of the content-based and template-based classification, the calculation-based classification is further improved. The pro of the calculation-based type is goal specificity, that is, the classification covers the questions that require calculation to obtain the answer such as counting the quantity or finding the maximum value. However, the con of this classification is that the range of questions included is limited. That is to say, it is precisely because of the clarity of the goal that this classification still starts from the operational level of the IQA which contains only a small fraction of the semantic parsing results.

Hence, this idea can be referred to operational level, and calculation-based classification may be an auxiliary type for the first two classifications. Nevertheless, this type can only carry out simple reasoning and calculations (such as comparing the number of two entities); how to incorporate more sophisticated symbolic reasoning abilities into the IQA system is also a challenge [13].

### 2.4. The Method-Based Classification

To deal with more sophisticated situations like symbolic reasoning, method-based classification was proposed. The method-based classification divides questions into two types, information retrieval-based and semantic parsing-based, according to whether they focus on semantic parsing or answer query [15,16,18], as shown in Table 5.

Tracing back to the solutions for IQA tasks, a number of studies on IR-based and SP-based methods have been proposed [18]. The advantage of the classification lies in high integration. The classification starts from a higher level than the operational level of the IQA system, which covers a more integrated classification than content-based, template-based and calculation-based classification. While the advantage brings an improved classification,

the high integration of this classification may lead to a lack of expansibility and lack of a more detailed class. That is to say, the method-based classification involves question understanding as a primary step which is short of discussion of some complex issues and deep thought of the sub-level (i.e., operational level).

**Table 5.** Classification based on the methods.

Sub-Type	Explanation
Information Retrieval-Based (IR-Based) Methods	IR-based approach constructs a question-specific graph delivering the comprehensive information related to the question and ranks all entities in the extracted graph based on their relevance to the question.
Semantic Parsing-Based (SP-Based) Methods	SP-based approach represents a question by a symbolic logic form, and then executes it against the Knowledge Base (KB) to obtain the final answers.

Therefore, semantic parsing and induction level of method-based classification are relatively high due to high abstraction ability. Nevertheless, method-based classification still aims to query the answers from the database. Thus, we believe that if we take a higher-level view than that of method-based classification, the degree of induction for the IQA\_QC framework will be better.

In view of the current research status, the above four IQA\_QC frameworks all have advantages and disadvantages. The first three classifications, content-based, template-based and calculation-based classification, are too trivial to be a comprehensive IQA\_QC framework. The method-based classification is improved at the operational level, while it does not take into account that the answer may not be queried from the database. And we can find that most scholars focus on the operational level of IQA. However, such IQA\_QC frameworks are not sufficient to support the IQA system both in semantic parsing and question querying. For example, content-based classification relies too much on the results of semantic parsing which imposes a heavy strain on the precision of semantic parsing; method-based classification can only query questions whose answers should be stored in the database. Therefore, we speculate on current and expected IQA\_QC framework trends that arise in response to new challenges in IQA field. Furthermore, there is a lack of a bridge between theory and practice regarding how the IQA\_QC framework guides the IQA system in the aspects of semantic parsing and question querying. In order to well unify theory and practice, we apply the hierarchy-based IQA\_QC framework to geosciences [21,27,28] issues, which covers variety of complex questions (detailed in Section 3). From three levels, we first consider whether the answer to the question exists in the database. Then, we should consider which method to use to answer the question. Finally, we have to think about the implementation of the specific questions. In short, a more comprehensive and systematic IQA\_QC framework which would effectively guide semantic parsing and question querying of the IQA system is proposed.

### 3. Question Classification for Intelligent Question Answering

In this section, we will describe our IQA\_QC framework in detail. Firstly, we will introduce the basic idea of our hierarchy-based classification, which indicates the process of summarizing and generalizing. Then, we will explain the IQA\_QC framework concretely, which is divided into Retrieval Question Answering (RQA) and Generative Question Answering (GQA) according to the popular classification based on whether the answer exists in an existing database. Finally, more specific categories will be detailed.

#### 3.1. Basic Idea

Compared with the four frameworks mentioned in Section 2, we proposed a hierarchy-based classification, which consists of three levels. We follow these principles: the first level, regarding whether the answers can be retrieved or whether they need to be generated in the

IQA system, that is, IQA existence and essence; the second level, regarding which method to acquire the answer in the IQA system, that is, IQA form; the third level, regarding which special implementation means to answer the questions of the IQA system, that is, IQA implementation. Figure 2 presents the above ideas.

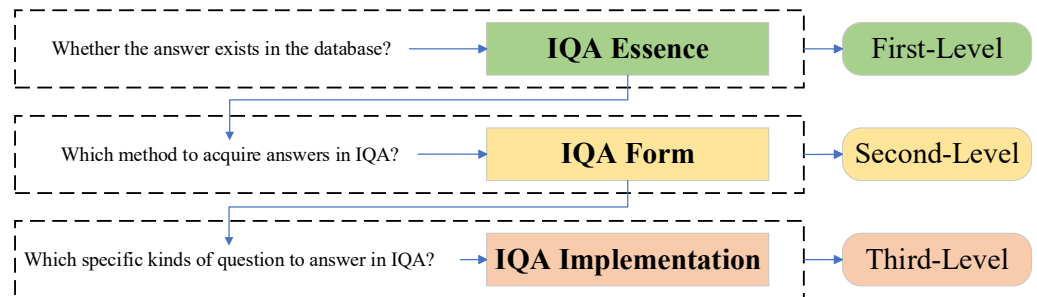


Figure 2. The principles of the IQA\_QC framework.

Based on the above principles, we create a specific and detailed IQA\_QC framework. From the first level, in which the answer to the question may be in or not be in the database, IQA generally falls into two types: Retrieval Question Answering (RQA) and Generative Question Answering (GQA). In the following sub-sections, their definitions and more detailed classifications are presented in turn.

The hierarchy-based IQA\_QC framework is shown in Figure 3. As can be seen from Figure 3, the IQA\_QC framework is divided into three levels according to three principles, i.e., essence, form and implementation of the IQA system, as presented in Figure 2. Each type surrounded by dotted lines corresponds to the principles. Obviously, our IQA\_QC framework covers more types of questions than existing IQA\_QC frameworks, which will be discussed in Section 5.1. This IQA\_QC framework will be described in the following sub-sections.

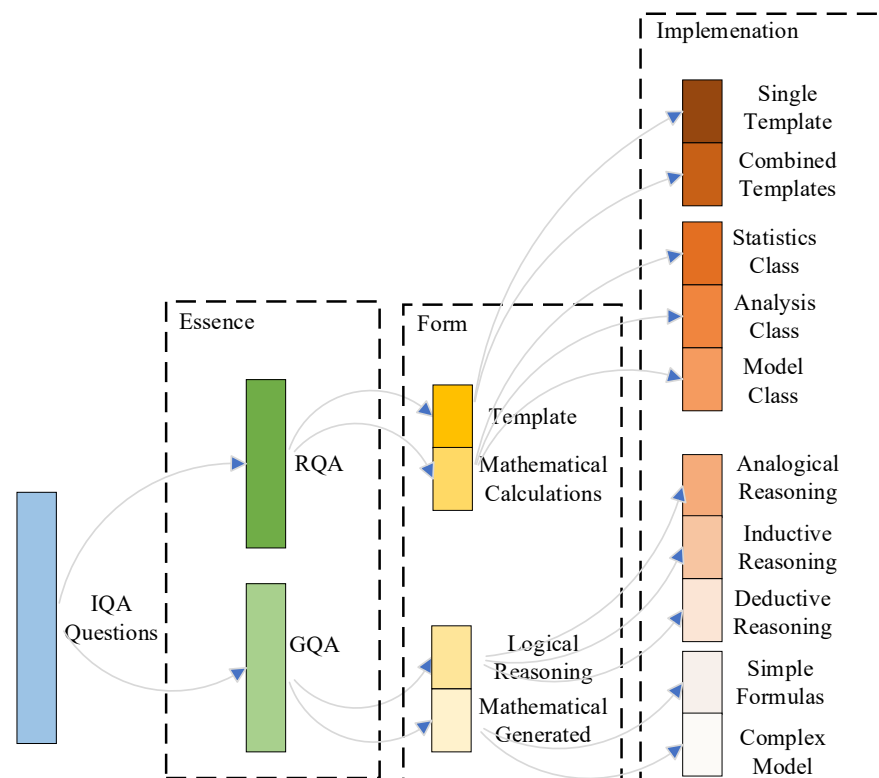


Figure 3. The components of the IQA\_QC framework.

To better understand the abstract description of the principle and specific components of the hierarchy-based IQA\_QC framework proposed in this paper, the following explanation is combined with geosciences-related questions in implementation (third level) which also well unifies theory and practice. Specifically, for each question type, the Beautiful China Ecological Civilization Pattern in the geosciences field is taken as an example to illustrate. Also, “Pattern” in the following example refers to the development pattern, such as “Beautiful Country Pattern”, “Characteristic Town Pattern” [21], etc.

### 3.2. Retrieval Question Answering

RQA determines the most appropriate answers as replies according to the user’s input questions from a large amount of corpus data. In other words, RQA is mainly based on the dataset stored in the database. When a user asks a question, the database can query directly through the template and simple calculation and determine the answer. Thus, the sub-types can be categorized into queries based on template and mathematical calculations.

#### 3.2.1. Queries Based on Template

Questions raised by users can be answered based on template query, which is mainly divided into a single template and combined templates.

##### (a) Single template

The function of a single template is to query entities or attributes which means querying entity B (attribute) by entity A (attribute). A single template query is typically based on a triple template query entities or attributes such as <Subject, Relation, Object> or <Subject, Attribute, Attribute Value> to acquire the correct answers (raw data in the database). For the Beautiful China Ecological Civilization Pattern database, the entity attributes that can be queried are shown in Table 6.

**Table 6.** The examples of single template.

Sub-Type	Example	Single Template
Factoid	Where is Giant Panda National Park?	<Giant Panda National Park, locates, Sichuan Province>
Confirmation	Is Nanshan National Park applying National Park Pattern for development?	<Nanshan National Park, applies, National Park Pattern>
Definition	What is the meaning of Characteristic Town Pattern?	<Characteristic Town, means, Towns with different development characteristics>
Example	What are the specific villages of Beautiful Country Pattern?	<Beautiful Country Pattern, has, Jinlong Country and Dapo Country... >

As can be seen from Table 6, these questions belong to content-based classification as discussed in Section 2.1. The answer can be obtained effectively by identifying the corresponding entities, attributes and the relation. For example, for the question “Where is Giant Panda National Park?”, we use one template <Giant Panda National Park, locates, Sichuan Province> and find that the answer is Sichuan Province. This type of question is the most common question in IQA. These questions are content-based classification, extracting raw data from the database and returning it to the user.

##### (b) Combined templates

In fact, in IQA systems, the question types are diverse, and the answer is not just a single entity or attribute most of the time. Unlike the single template, combined templates use multiple triples to query entities and attributes at the same time. Combined templates are composites of multiple triples which combine <Subject, Relation, Object> and <Subject, Attribute, Attribute Value> to implement more complex queries than a single template.



Three notable sub-types are question-to-fact, fact-to-answer and question-to-answer. Table 7 shows the examples.

**Table 7.** The examples of combined template.

Sub-Type	Example	Single Template
Question-to-Fact	What patterns can Chaoyang District apply for development?	<Chaoyang District, Belong to, Beijing>, <Beijing, Apply, Agricultural Innovation Park Pattern>
Fact-to-Answer	What will cause floods in northwest of China?	<Rainfall, reach to, flood >, <Northwest, rainfall, flood>
Question-to-Answer	What intangible cultural heritage does the capital of China have?	<China, Capital, Beijing>, <Beijing, Intangible Cultural Heritage, Beijing Opera>

An alternative solution for KG reasoning is to infer missing facts by synthesizing information from multi-hop paths [29]. As can be seen from the above examples, users may want to know about multiple entities and attributes during the IQA process. According to the template-based classification, this is the multi-hop query in the classification which indicates that only one triple query cannot find the desired answer. For example, for the question “What intangible cultural heritage does the capital of China have?”, we use combined templates <China, Capital, Beijing>, <Beijing, Intangible Cultural Heritage, Beijing Opera> and find that Beijing Opera is the final answer after we know Beijing is the capital of China. In geoscience questions, multiple triple templates are required for the combined query. Furthermore, combined templates are also necessary for the IQA system in other fields.

### 3.2.2. Mathematical Calculation

In the process of acquiring the answers, the approach that relies on templates is not comprehensive. Sometimes, the data need to be calculated, sorted or reasoned to determine the final answer. Therefore, considering that those questions cannot be directly answered, mathematical calculation classification has been proposed, which consists of statistic class, analysis class and model class, as described in Figure 3.

#### (a) Statistics class

This class is mainly aimed at some statistical questions, such as finding the maximum, minimum and average values. For example, for the question “Which city is the largest in area in Jiangsu Province?”, we should first find all the cities in Jiangsu Province as candidates for the answer and then calculate and compare the area of each city to obtain the result with the largest area. The data need to be calculated or a statistical function used to determine the final answer. The notable sub-types are max, min and average. Examples are shown in Table 8.

**Table 8.** The examples of statistics class.

Sub-Type	Example
Max	Which city is the largest in area in Jiangsu Province?
Min	What pattern has been applied for development the least in Jiangsu Province?
Average	How many Ecological Civilization Patterns are there on average in Jiangxi Province?

This kind of question corresponds to the calculation-based classification mentioned in Section 2. We need to calculate data or use a statistical function to obtain the final answer. Statistical questions are mainly for quantitative questions and determine answers from the candidates.

## (b) Analysis class

Analysis class is mainly for those questions that focus on understanding certain complex natural language. For example, words like “in recent years” and “around” may appear abundantly in questions. How to make the computer understand the meaning of the above words is the key. In other words, semantic parsing of questions for determining the corresponding answer is important. The notable two are temporal type and spatial type [30]. Specific examples are shown in Table 9.

**Table 9.** The examples of analysis class.

Sub-Type	Example
Temporal Type	In recent years, what kind of Ecological Civilization Pattern has been applied in Xihu District for development?
Spatial Type	Around Yangtze River, where are Ecological Planting Patterns being applied for development?

Analysis class corresponds to semantic parsing-based (SP-based) methods mentioned in Section 2. For the question “In recent years, what kind of Ecological Civilization Pattern has been applied in Xihu District”, we have to make computer understand the specific time of “In recent years” such as in the last 5 years or 10 years. For the question “Around Yangtze River, where are Ecological Planting Patterns being applied for development?” we have to make the computer understand the specific space of “around” such as within 5 or 10 miles. These questions which often occur in the IQA system require semantic parsing and concrete analysis. How to make the computer understand the meaning of the question is the most important.

## (c) Model class

To answer the questions that cannot be addressed by the statistical class and analysis class, the model class is proposed to tackle this problem. The model class is mainly for the questions that use computer models to acquire answers. The most notable two types are the recommendation model and prediction model. Furthermore, different computer models may be used to solve questions in different fields in IQA systems. Examples are shown in Table 10.

**Table 10.** The examples of model class.

Sub-Type	Example
Recommendation Model	What areas are suitable for the Ecotourism Sustainable Development Pattern?
Prediction Model	When will mudslides appear in the northwest of China?

The model class uses the computer models to deal with the entity or attribute or attribute values that exist in the database and determine the results. For “What areas are suitable for the Ecotourism Sustainable Development Pattern?”, the solution is not to calculate the maximum value, nor to parse every word, which the statistical class and analysis class cannot solve. This is a complex reasoning process that must be based on computer models and some properties in the database to acquire the final answer.

## 3.3. Generative Question Answering

While a fact-based question usually has a single best answer, an opinion question often has many relevant answers, which may reflect a variety of different viewpoints [4,31]. The question types in Section 3.2 only consider the condition that the answer exists in the database, while the answer to the question needs to be generated and does not exist in the database in most cases. Generative Question Answering (GQA) is introduced which covers

logical reasoning and mathematical generation, since GQA covers more deep-seated and complex issues.

### 3.3.1. Logical Reasoning

Solving grounded language tasks often requires reasoning about relationships among objects in the context of a given task [32,33]. In generative questions, there are a lot of questions that need to be answered logically, which relies on the original database to deduce the answer. Thus, logical reasoning, which mainly refers to generated answers based on logical reasoning, is proposed and covers analogical reasoning, inductive reasoning and deductive reasoning.

#### (a) Analogical reasoning

Analogical reasoning is the process of considering that two objects are the same or similar in some attributes by comparing them and inferring that they are the same in other attributes. For instance, for the question “Is Qixia District suitable for applying Characteristic Town Ecological Civilization Pattern in Gong’an County?”, it is necessary to reason whether certain attributes of Qixia District are similar to that of Gong’an County. By analogical reasoning, Qixia District can apply the same pattern for development as Gong’an County because the two areas have similar attributes. We wish that the computer could deal with the concept of “similar” here. “Similar” may mean the similarity of either natural or social conditions. If the similarity of certain characteristics is satisfied, it can be judged that similar areas can apply the same Ecological Civilization Pattern for development.

#### (b) Inductive reasoning

Inductive reasoning refers to the transition from a certain degree of view on individual things to a wider range of views and the derivation of general principles from specific cases. For example, Qixia District, Gulou District and Qinhuai District all apply an Ecological Planting Pattern with rainfall of more than 300 mm. According to the above condition, we can reason the rule that the areas with rainfall of more than 300 mm are suitable for applying the Ecological Planting Pattern for development. For the question “Can Lishui District with rainfall of more than 300 mm apply the Ecological Planting Pattern for development?”, we answer “Yes” according to the general principle. We can summarize certain patterns from the large amount of data in the database by using inductive reasoning. Through inductive reasoning, a general principle is deduced, that is, what kind of Ecological Civilization Pattern can be applied for development if it meets certain conditions.

#### (c) Deductive reasoning

Deductive reasoning starts from a general premise and deductively leads to a specific statement or individual conclusion. For example, we firstly give a general premise that Hangzhou is suitable for applying the Ecotourism Ecological Civilization Pattern for development. Then, through the specific statement that Xihu District belongs to Hangzhou we can reason the rule that Xihu District is also suitable for applying the Ecotourism Ecological Civilization Pattern for development. In this case, Hangzhou is the big concept, and Xihu District is the small concept. This kind of reasoning from the general to the special is the core of deductive reasoning. Thus, we can answer “Yes” for the question “Is Xihu District suitable for the Ecotourism Ecological Civilization Pattern?” by using deductive reasoning. We can move from a general premise to a specific statement to solve the corresponding questions which need deductive reasoning.

### 3.3.2. Mathematical Generation

Mathematical generation involves calculations, including simple mathematical formulas and complicated models, to acquire the answers. Unlike the mathematical calculation class, the answer to such questions does not exist in the database. We use simple mathematical formulas and complicated models to generate the final answers.

## (a) Simple formulas

Simple formulas refer to the use of some common computations (addition, subtraction, multiplication and division) to come up with an answer instead of using off-the-shelf statistical modules (max, min and average). Simple mathematical formulas need to be customized and designed. For the class in GQA, the answer does not exist in the database, but a few simple mathematical formulas can be designed to acquire the user's desired results. Table 11 shows the examples.

**Table 11.** The examples of simple formulas.

Sub-Type	Example
Addition	What is the total area of Nanjing (area of each district of Nanjing)?
Subtraction	What is the population of Qinhuai District (total population)?
Multiplication	What is the forest area in Qixia District (forest coverage and total area)?
Division	What is the capita income of Gong'an County (per population and total income)?

As can be seen from the examples in the table, the database does not store attributes such as forest area or population. But a few simple calculations using forest land coverage and GDP and GDP per capita should give users the answer they want. In IQA systems, this type of question is also common but often ignored.

## (b) Complex model

A complex model refers to the use of computer models to acquire the answers which do not exist in the database. When questions become complicated from both semantic and syntactic aspects, models are required to have strong capabilities of natural language understanding and generalization [34,35]. Table 12 presents the examples which include the prediction model and decision model (the underline here means to distinguish them from the model class in Section 3.2.2). Unlike the model class in Section 3.2.2, the complicated model here is mainly used in the case of a lack of data in the face of complex questions.

**Table 12.** The examples of complex model.

Sub-Type	Example
Prediction Model	What patterns can be applied in Jiangsu Province for development in the next five years?
Decision Model	Is Xihu District suitable for an Ecological Planting Pattern for development?

In the case of a lack of data, many complex questions cannot be answered by simple calculation and reasoning [36]. For the question "What patterns can be applied in Jiangsu Province for development in the next five years?", the data on how Jiangsu Province will develop in the next five years cannot be stored in the corpus. Thus, prediction models are needed to deal with the above problem as well as to tackle the lack of demand data. And we may get the answer "Ecological Planting Pattern" by using the prediction model. Also, for the question "Is Xihu District suitable for an Ecological Planting Pattern for development?", the Ecological Planting Pattern is not stored in the dataset of Xihu District. But this does not mean that Xihu District cannot apply this pattern for development. We can obtain the answer "Yes" by using the decision model. In addition to the above examples, more computer models will be used to answer complex questions which are indispensable in the IQA\_QC framework. And this is also a relatively open problem in the field for the IQA system.

#### 4. IQA Evaluation Metrics

The evaluation metric measures the performance of the IQA system guided by the IQA\_QC framework. Different frameworks may have different effects of guidance. Thus, it is necessary to introduce some acknowledged evaluation metrics to make a judgment on the IQA system. At present, there are four common kinds of IQA evaluation metrics, which refer to accuracy, precision, recall and F1 [5,15,37–39]. We describe the four metrics in detail from the following three aspects, including the explanation, calculation methods and formulas. The formulas are based on true positive (TP), true negative (TN), false positive (FP) and false negative (FN) examples.

*Accuracy*: This metric shows how many questions are answered correctly (considering retrieving all the answers of a given question).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

*Precision*: *Precision* (also called positive predictive value) is the fraction of correct answers among the total retrieved answers.

$$Recall = TP / (TP + FN) \quad (2)$$

*Recall*: *Recall* is the fraction of correct answers that were retrieved among the total actual answers.

$$Precision = TP / (TP + FP) \quad (3)$$

*F1*: *F1* is a function of precision and recall that includes both metrics effects.

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

The above four evaluation metrics can measure the performance of the IQA system, but these metrics are only for the condition that the answers to the questions are stored in the database. For example, for “Where is Giant Panda National Park?”, the answer is “Chengdu of Sichuan Province” which is stored in the database. However, the condition that the answers to the questions are not stored in the database or the question itself has no exact answer lacks consideration. For example, for “What patterns can be applied in Jiangsu Province for development in the next five years”, it is obvious that data in the database for the next five years cannot be stored. We could not determine the exact answer so it is impossible to evaluate the answering performance. Also, for the question “When will mudslides appear in the northwest of China?”, there are absolutely no exact answers to the question, and we could not evaluate how well the question is answered. Therefore, the common four evaluation metrics cannot evaluate the performance of IQA systems well. A comprehensive evaluation system is summarized in Figure 4. Although there are four evaluation metrics, which have a certain range of application, there is a lack of corresponding evaluation metrics for some complex questions in the IQA\_QC framework.

As shown in Figure 4, an excellent evaluation system should contain three dimensions, answer performance, functional performance and systematic performance. The first dimension, answer performance, can assess the performance regarding the answer ability, such as *accuracy*, *precision*, *recall* and *F1*. Unfortunately, it is impossible to find the answer directly when the question is very complex. Thus, other dimensions are needed to evaluate the IQA system well [24,40,41]. The second dimension, functional performance, provides the overall effective metrics which include *accuracy*, *precision*, *recall* and *F1*, to evaluate IQA systems. However, it is impossible to evaluate a specific function of IQA systems. For example, chatting ability [17,42,43] in functional performance is one of the most important evaluation metrics which meets the users’ daily chat and dialogue needs. The third dimension, systematic performance, can provide the metrics of conversational ability which include overall effective metrics, chatting ability and so on to evaluate IQA systems. However, it cannot evaluate other systematic performance. For example, complexity of implementation [41,44] in systematic performance can evaluate the time and space complexity of the IQA system

to provide a reference for saving resources [5,12,34,38]. From that point of view, we have made a comprehensive induction and integration of the evaluation metrics, which can better help researchers have a more comprehensive understanding of the question types in IQA systems.

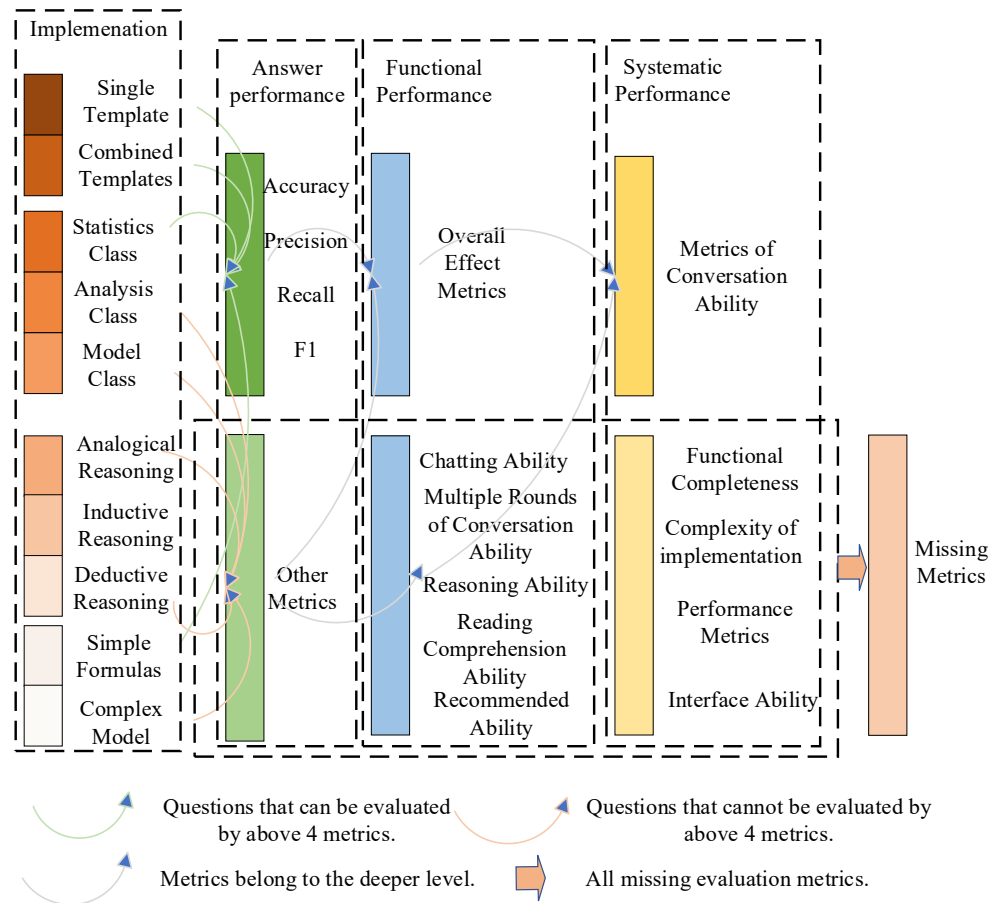


Figure 4. Performance that can be measured.

### 5. Discussion

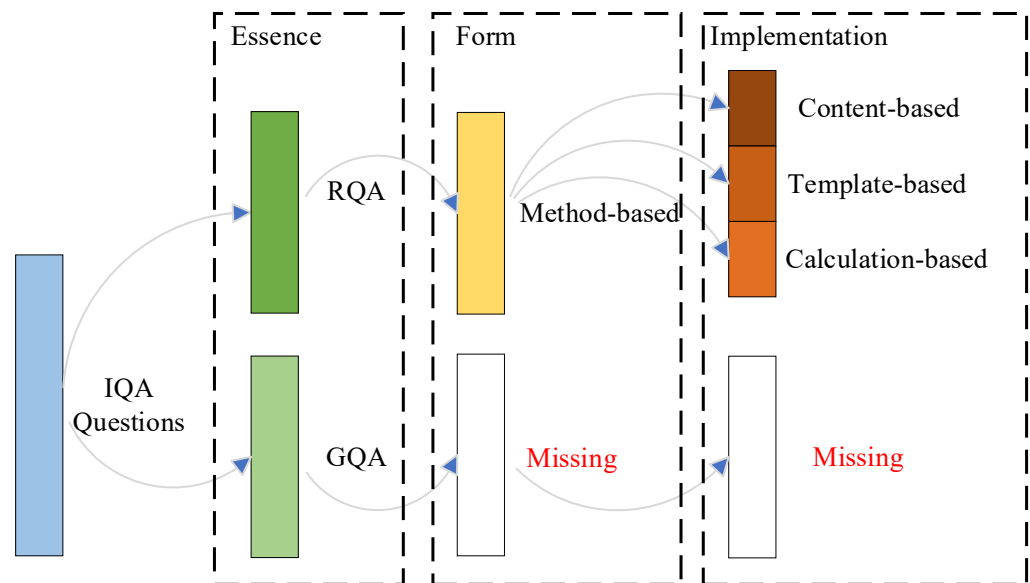
In this section, the main aim is to describe analysis and limitations of the IQA\_QC framework proposed in Section 3.

#### 5.1. Analysis of the IQA\_QC Framework

The comprehensiveness of the hierarchy-based IQA\_QC framework proposed in this paper is mainly reflected in the following three aspects.

Firstly, the hierarchy-based IQA\_QC framework covers more types of questions compared with the classification in Section 2. Figures 3 and 5 show the comparison of types in Sections 2 and 3. It can be seen directly from Figure 5 that 50% of the questions that are not currently paid attention to are found. On the one hand, model classes in RQA are added for those questions which need a computer model to carry out computation which the existing IQA\_QC classifications do not cover. But it is especially important for the IQA system currently for recommendations and predictions. On the other hand, logical reasoning and mathematical generation in GQA are supplementary to make conclusions and inferences due to semantic parsing and make calculations using formulas and complex models. For instance, when users want the system to recommend a certain kind of movie, a recommendation model is needed to acquire the result. Or users may want to know how much rain will fall when a mudslide occurs based on previous statistics. These kinds of question are common of geoscience in the IQA system. But it was ignored by all the existing

IQA\_QC frameworks. The existing IQA\_QC frameworks focus only on addressing the questions in general terms and then expect to be able to query the answer in the database. With the development of the IQA system, the questions will become more flexible, complex and changeable. That is, the existing IQA\_QC frameworks are unable to guide the IQA system well. The hierarchy-based IQA\_QC framework proposed in this paper makes up for the lack of existing IQA\_QC frameworks in Section 2 and makes the system develop towards the direction of more and more intelligence. As can be seen from Figures 3 and 5, the hierarchy-based IQA\_QC framework proposed in this paper is richer and contains more content and is more comprehensive compared with the existing IQA\_QC frameworks.



**Figure 5.** The comparison of our IQA\_QC framework with the existing frameworks mentioned in Section 2.

Secondly, we have made a comprehensive induction and integration of the evaluation metrics. The existing evaluation metrics are not enough for the IQA system from Section 4. That is to say, it is necessary to supplement metrics in all three dimensions, answer performance, functional performance and systematic performance, mentioned in Section 4. The comprehensive evaluation system can better help researchers have a deeper understanding of question types in the IQA system.

Finally, specific examples are given for the hierarchy-based IQA\_QC framework proposed in this paper. The examples in this paper are based on the geosciences field which has complex and diverse types of questions. The questions in the geoscience field are included in this hierarchy-based IQA\_QC framework which can prove the improvement and supplement compared with the hierarchy-based IQA\_QC framework in Section 2. Moreover, this framework can be applied to guide complex IQA systems in other fields as well.

### 5.2. Imitation of the IQA\_QC Framework

The IQA\_QC framework proposed in this paper still needs extensibility. Although the IQA\_QC framework proposed in Section 3 is hierarchy-based, which contains essence, form and implementation, the classification stopped at the third level (implementation) which does not contain more details [45–47]. Further down the line, there may be more and more detailed types which were not mentioned in Section 3. This is also the limitation of the hierarchy-based IQA\_QC framework proposed in this paper. However, because there is no more detailed classification of the third level (implementation), it can provide more extensibility for the hierarchy-based IQA\_QC framework. Thus, the hierarchy-based IQA\_QC framework in this paper is very extensible, and researchers do not need to be

limited to the modules and can further expand the third level (implementation) according to their own practical applications.

## 6. Conclusions

In this paper, we focus on the geographical IQA system and propose an IQA\_QC framework to solve the problems the current classification has. We faced the problems which are mainly in the following two aspects. First, the current IQA\_QC frameworks could not well guide the IQA system both in semantic parsing and question querying and, second, there is lack of a comprehensive IQA\_QC framework as well as systematic evaluation. To tackle the problems encountered in the field of IQA, this paper proposes a new IQA\_QC framework which refers to the three principles (essence, form and implementation) of IQA. Moreover, a number of metrics are summarized to evaluate the IQA system according to the IQA\_QC framework. The hierarchy-based IQA\_QC framework proposed in this paper not only provides the basis for the IQA system, especially for GeoIQA, but also provides a training corpus for generative AI systems like ChatGPT in certain fields like geoscience.

With the proposal of the hierarchy-based IQA\_QC framework, the IQA system can be improved from two aspects: the classification of semantic parsing results (class) is more accurate; various computer models and other methods are used to query answers which makes the query result more accurate. For the IQA system, especially for GeoIQA, a comprehensive and systematic hierarchy-based IQA\_QC framework can proceed to the specific system design. Then, there are roughly two steps. The first is semantic parsing, identifying the user's intention, which is how to make a computer understand natural language. Then, the second stage is to determine the answer, which requires various methods of querying, calculating, deriving the answer and so on. How to determine accurate answers is also a crucial step for the IQA system, which is related to whether it can meet the needs of users. Therefore, it is challenging to design and implement an IQA system, especially when thinking about what kinds of questions to ask. In summary, the content and direction of IQA to be carried out in the future on the basis of this hierarchy-based IQA\_QC framework have the following three aspects. Firstly, question classification can be improved by incorporating approaches and models into the existing IQA system in geoscience, which could help categorize complex geographical questions so that each type of question can be handled more specifically. Secondly, the impact of the hierarchy-based IQA\_QC framework on IQA performance can be further compared by different classification approaches, which indicates that a clear hierarchy-based IQA\_QC framework can better guide the IQA system to deal with different questions in geoscience and determine the answers more accurately. Finally, the evaluation system can be optimized by supplementing the evaluation metrics of specific questions in the hierarchy-based IQA\_QC framework, which means that the performance of the IQA system would be evaluated from a multi-dimensional perspective rather than just being limited to general metrics. In future research, we need to refine the hierarchy-based IQA\_QC framework to guide the IQA system and utilize different approaches to deal with different types of questions to improve the performance of the IQA system. Also, we need to evaluate the IQA system from a broader dimension, so that we have a more comprehensive understanding of the performance of the IQA system.

**Author Contributions:** Hao Sun: Conceptualization, Investigation, Analysis, Original draft preparation. Shu Wang: Conceptualization, Methodology, Analysis, Draft editing, Funding acquisition. Yunqiang Zhu: Funding acquisition, Supervision. Wen Yuan: Funding acquisition, Supervision. Zhiqiang Zou: Conceptualization, Draft editing, Funding acquisition, Supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (grant number 42101467), the Strategic Priority Research Program of the Chinese Academy of Science (grant number XDA23100100), the Chinese Scholarship Council (grant number 202008320044) and the National Key R&D Program of China (2022YFF0711601, 2022YFB3904202).



**Data Availability Statement:** This work does not involve additional data, and the relevant literature has been listed in the reference section.

**Acknowledgments:** This work was supported by the International Big Science Program of Deep-time Digital Earth (DDE). We would like to extend our thanks to the principal investigators of DDE, Chengshan Wang, Chenghu Zhou and Qiuming Cheng for their guidance and valuable comments. Meanwhile, we thank the editors and reviewers for their valuable comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arbaaen, A.; Shah, A. Natural language processing based question answering techniques: A survey. In Proceedings of the 2020 IEEE 7th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 18–20 December 2020; pp. 1–8.
2. Dwivedi, S.K.; Singh, V. Research and reviews in question answering system. *Procedia Technol.* **2013**, *10*, 417–424. [\[CrossRef\]](#)
3. Mishra, A.; Jain, S.K. A survey on question answering systems with classification. *J. King Saud Univ. Comput. Inf. Sci.* **2016**, *28*, 345–361. [\[CrossRef\]](#)
4. Moghaddam, S.; Ester, M. AQA: Aspect-based opinion question answering. In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, Canada, 11 December 2011; pp. 89–96.
5. Ojokoh, B.; Adebisi, E. A review of question answering systems. *J. Web Eng.* **2018**, *17*, 717–758. [\[CrossRef\]](#)
6. Azad, H.K.; Deepak, A. Query expansion techniques for information retrieval: A survey. *Inf. Process. Manag.* **2019**, *56*, 1698–1735. [\[CrossRef\]](#)
7. Charef, N.; Mnaouer, A.B.; Aloqaily, M.; Bouachir, O.; Guizani, M. Artificial intelligence implication on energy sustainability in Internet of Things: A survey. *Inf. Process. Manag.* **2023**, *60*, 103212. [\[CrossRef\]](#)
8. Zangerle, E.; Bauer, C. Evaluating recommender systems: Survey and framework. *ACM Comput. Surv.* **2022**, *55*, 1–38. [\[CrossRef\]](#)
9. Allam, A.M.N.; Haggag, M.H. The question answering systems: A survey. *Int. J. Res. Rev. Inf. Sci.* **2012**, *2*, 13.
10. Abujabal, A.; Yahya, M.; Riedewald, M.; Weikum, G. Automated template generation for question answering over knowledge graphs. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 1191–1200.
11. Chen, Y.; Wu, L.; Zaki, M.J. Bidirectional attentive memory networks for question answering over knowledge bases. In Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics (ACL): Kerrville, TX, USA, 2019.
12. Khot, T.; Sabharwal, A.; Clark, P. What’s Missing: A Knowledge Gap Guided Approach for Multi-hop Question Answering. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 2814–2828.
13. Kwon, H.; Trivedi, H.; Jansen, P.; Surdeanu, M.; Balasubramanian, N. Controlling information aggregation for complex question answering. In *Advances in Information Retrieval: 40th European Conference on IR Research, ECIR 2018, Grenoble, France, 26–29 March 2018—Proceedings 40*; Springer: Cham, Switzerland, 2018; pp. 750–757.
14. Ran, Q.; Lin, Y.; Li, P.; Zhou, J.; Liu, Z. NumNet: Machine Reading Comprehension with Numerical Reasoning. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 2474–2484.
15. Etezadi, R.; Shamsfard, M. The state of the art in open domain complex question answering: A survey. *Appl. Intell.* **2023**, *53*, 4124–4144. [\[CrossRef\]](#)
16. Jia, Z.; Pramanik, S.; Saha Roy, R.; Weikum, G. Complex temporal question answering on knowledge graphs. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Queensland, Australia, 1–5 November 2021; pp. 792–802.
17. Benamara, F. Cooperative question answering in restricted domains: The WEBCOOP experiment. In Proceedings of the Conference on Question Answering in Restricted Domains, Barcelona, Spain, 25 July 2004; pp. 31–38.
18. Lan, Y.; He, G.; Jiang, J.; Jiang, J.; Zhao, W.X.; Wen, J.R. Complex knowledge base question answering: A survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 11196–11215. [\[CrossRef\]](#)
19. Luo, K.; Lin, F.; Luo, X.; Zhu, K. Knowledge base question answering via encoding of complex query graphs. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2185–2194.
20. Wang, S.; Zhu, Y.; Qian, L.; Song, J.; Yuan, W.; Sun, K.; Li, W.; Cheng, Q. A novel rapid web investigation method for ecological agriculture patterns in China. *Sci. Total Environ.* **2022**, *842*, 156653. [\[CrossRef\]](#)
21. Zeng, X.; Wang, S.; Zhu, Y.; Xu, M.; Zou, Z. A Knowledge Graph Convolutional Networks Method for Countryside Ecological Patterns Recommendation by Mining Geographical Features. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 625. [\[CrossRef\]](#)
22. Yu, Z.; Wang, S.; Zhu, Y.; Yuan, W.; Dai, X.; Zou, Z. Unveiling Optimal SDG Pathways: An Innovative Approach Leveraging Graph Pruning and Intent Graph for Effective Recommendations. *arXiv* **2023**, arXiv:2309.11741. [\[CrossRef\]](#)

23. Wang, M. A survey of answer extraction techniques in factoid question answering. *Comput. Linguist.* **2006**, *1*, 1–14.
24. Lin, X.V.; Socher, R.; Xiong, C. Multi-Hop Knowledge Graph Reasoning with Reward Shaping. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018.
25. Roy, S.; Roth, D. Solving General Arithmetic Word Problems. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; Association for Computational Linguistics: Kerrville, TX, USA, 2015.
26. Sharath, J.S.; Banafsheh, R. Conversational question answering over knowledge base using chat-bot framework. In Proceedings of the 2021 IEEE 15th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 27–29 January 2021; pp. 84–85.
27. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 1544–1554. [[CrossRef](#)]
28. Xu, M.; Wang, S.; Song, C.; Zhu, A.; Zhu, Y.; Zou, Z. The Recommendation of the Rural Ecological Civilization Pattern Based on Geographic Data Argumentation. *Appl. Sci.* **2022**, *12*, 8024. [[CrossRef](#)]
29. Salunkhe, A. Evolution of techniques for question answering over knowledge base: A survey. *Int. J. Comput. Appl.* **2020**, *177*, 9–14. [[CrossRef](#)]
30. Ku, L.W.; Liang, Y.T.; Chen, H.H. Question analysis and answer passage retrieval for opinion question answering systems. *Int. J. Comput. Linguist. Chin. Lang. Process.* **2008**, *13*, 307–326.
31. Yang, W.; Xie, Y.; Lin, A.; Li, X.; Tan, L.; Xiong, K.; Li, M.; Lin, J. End-to-End Open-Domain Question Answering with BERTserini. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), Minneapolis, MN, USA, 2–7 June 2019; pp. 72–77.
32. Hu, R.; Rohrbach, A.; Darrell, T.; Saenko, K. Language-conditioned graph networks for relational reasoning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 10294–10303.
33. Somasundaran, S. QA with Attitude: Exploiting Opinion Type Analysis for Improving Question Answering in Online Discussions and the News. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM), Boulder, CO, USA, 26–27 March 2007.
34. Breja, M.; Jain, S.K. Analysis of why-type questions for the question answering system. In *New Trends in Databases and Information Systems: ADBIS 2018 Short Papers and Workshops, AI\* QA, BIGPMED, CSACDB, M2U, BigDataMAPS, ISTREND, DC, Budapest, Hungary, 2–5 September 2018—Proceedings 22*; Springer: Cham, Switzerland, 2018; pp. 265–273.
35. Zhong, V.; Xiong, C.; Socher, R. Seq2SQL: Generating Structured Queries from Natural Language Using Reinforcement Learning. *arXiv* **2018**, arXiv:1709.00103.
36. Khvalchik, M.; Revenko, A.; Blaschke, C. Question Answering for Link Prediction and Verification. In Proceedings of the European Semantic Web Conference, Portorož, Slovenia, 2–6 June 2019; pp. 116–120.
37. Dimitrakis, E.; Sgontzos, K.; Tzitzikas, Y. A survey on question answering systems over linked data and documents. *J. Intell. Inf. Syst.* **2020**, *55*, 233–259. [[CrossRef](#)]
38. Kaur, J.; Gupta, V. Effective question answering techniques and their evaluation metrics. *Int. J. Comput. Appl.* **2013**, *65*, 30–37.
39. Trivedi, P.; Maheshwari, G.; Dubey, M.; Lehmann, J. Lc-quad: A corpus for complex question answering over knowledge graphs. In *The Semantic Web—ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, 21–25 October 2017—Proceedings, Part II 16*; Springer: Cham, Switzerland, 2017; pp. 210–218.
40. Dubey, M.; Banerjee, D.; Abdelkawi, A.; Lehmann, J. Lc-quad 2.0: A large dataset for complex question answering over wikidata and dbpedia. In *The Semantic Web—ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019—Proceedings, Part II 18*; Springer: Cham, Switzerland, 2019; pp. 69–78.
41. Rajpurkar, P.; Zhang, J.; Lopyrev, K.; Liang, P. SQuAD: 100,000+ Questions for Machine Comprehension of Text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 2383–2392.
42. Etezadi, R.; Shamsfard, M. pecoq: A dataset for persian complex question answering over knowledge graph. In Proceedings of the 2020 11th International Conference on Information and Knowledge Technology (IKT), Tehran, Iran, 22–23 December 2020; pp. 102–106.
43. Fader, A.; Zettlemoyer, L.; Etzioni, O. Paraphrase-driven learning for open question answering. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Sofia, Bulgaria, 4–9 August 2013; pp. 1608–1618.
44. Rodrigo, A.; Penas, A. A study about the future evaluation of Question-Answering systems. *Knowl.-Based Syst.* **2017**, *137*, 83–93. [[CrossRef](#)]
45. Altinel, B.; Ganiz, M.C. Semantic text classification: A survey of past and recent advances. *Inf. Process. Manag.* **2018**, *54*, 1129–1153. [[CrossRef](#)]

46. Sasikumar, U.; Sindhu, L. A survey of natural language question answering system. *Int. J. Comput. Appl.* **2014**, *108*, 42–46. [[CrossRef](#)]
47. Wang, J.; Man, C.; Zhao, Y.; Wang, F. An answer recommendation algorithm for medical community question answering systems. In Proceedings of the 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Beijing, China, 10–12 July 2016; IEEE: New York, NY, USA, 2016; pp. 139–144.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.