International Journal of
Geo-Information

MDPI

Article

Geographical Entity Management Model Based on Multi-Classification

Lin Shi [1], Xiaoji Lan [1,*], Ming Xiao [2] and Ning Liu [2]

[1] School of Civil and Surveying & Mapping Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China; chgfeng100@nnu.edu.cn
[2] Territorial Space Survey and Planning Research Center of Ganzhou, Ganzhou 341000, China; xiaoming1993@outlook.com (M.X.)
* Correspondence: lanxiaoji@jxust.edu.cn

Abstract: Scientific and logical classification is crucial for efficient information storage, management, and sharing. However, there are numerous existing classification systems for geographical entities, and the categories to which the same geographical entity belongs are often different in the business databases constructed according to different classification systems, which brings great obstacles to the management and sharing of geographical information. This study analyzes the complexities of multiple classifications of geographical entities and proposes a multi-classification model for geographical entities based on directed hypergraph theory. This model integrates and transforms different classification systems for the same geographical entity, creating a unified method for expressing multiple classifications. We also designed a data structure to support this unified expression. By implementing this model, the study enables the effective management of geographical entity data, facilitating improved sharing and the exchange of geographical information across different industries and applications. In practical, the multi-classification model proposed in this paper allows geographical entities from different classification systems to be stored and managed within a single geographical database. Data views are then used to provide tailored services to various industry sectors and business applications. This approach effectively reduces data duplication and enhances the efficiency of managing and sharing geographical information. Using land use classification as an example, this study constructs a unified expression of three different land use classification systems based on the multi-classification model. An experiment managing land use data for a specific city was conducted using this model in PostgreSQL. The results indicate that the proposed method not only reduces data redundancy but also improves the query efficiency by over 10% on average compared to the mainstream relational database management mode. This confirms the effectiveness and practical value of the proposed method.

Keywords: geographical entity; classification system; multiple classifications; directed hypergraph; entity management

Citation: Shi, L.; Lan, X.; Xiao, M.; Liu, N. Geographical Entity Management Model Based on Multi-Classification. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 334. https://doi.org/10.3390/ijgi13090334

Academic Editors: Wolfgang Kainz and Eliseo Clementini

Received: 28 June 2024
Revised: 6 September 2024
Accepted: 13 September 2024
Published: 19 September 2024

1. Introduction

In the second half of the 19th century, Fukuzawa Yukichi, a prominent enlightenment thinker during Japan's Meiji Era, was the first to translate the Western concept of "science" into the Japanese term "science". He believed that science represented knowledge divided into various categories or disciplines. In this context, "science" refers to the classification and grading of different fields of study [1,2]. Classification is a fundamental research method that plays a crucial role in both social management and scientific research. In governance, the state organizes the government into various functional departments based on management needs. Similarly, in scientific research, fields are divided into different disciplines according to their specific areas of focus. Scientific and systematic classification is not only a common method in social management but also the foundation for data management in the digital age. Organizations such as the National Aeronautics and Space Administration (NASA) and the International Organization for Standardization (ISO) have developed their own data classification systems. Additionally, the Open Geospatial Consortium (OGC) has established numerous classification standards specifically for geosciences. In practice, various business application systems organize data hierarchically according to a specific classification system and classify and manage entity information through different data tables in databases [3,4]. Scientific and systematic classification is crucial for effective data management. The classification system significantly influences the design of information systems, as well as the processes of data storage, retrieval, analysis, application, and sharing [5–7].

Geographical entities, which are uniquely identifiable natural or artificial objects in the real world, serve as the primary carriers of geographical information. They are physical abstractions that humans use to describe and represent geographical phenomena and are the core components of geographical scenes [8,9]. The classification system for geographical entities is vital for organizing, managing, analyzing, and applying geographical information; furthermore, the classification system is essential for the unified management of geographical data and serves as a critical foundation for developing related application systems [10,11]. However, varying disciplinary perspectives, industry applications, and management needs can lead to different classifications of geographical entities. Historically, in the management and analysis of mapping and geographical information, government departments and various industries have conducted extensive research to develop classification systems tailored to their specific business areas and application requirements. For example, some standards are focused on basic surveying and mapping management, such as the "Classify and Codes for the National Land Information" [12], "Specifications for Feature Classification and Codes of Fundamental Geographic Information" [13], and "Specification for Geographical Entity Spatial Data" [14]. Others are tailored to specific sectors, such as the public security sector's "Classification and Codes for General Geographical Entities of the Police" [15] and the transportation sector's "Coding rules for Entity Identification Code of Traffic Management Geographical Information—City Road" [16]. Moreover, the categorization of geographical entities is not static and tends to change with time and management needs. For example, consider China's land classification standards. The first national land detailed survey, conducted in 1984, classified land into eight primary categories and 46 secondary categories. Later, the " Current Land Use Classification" [17], revised by the Ministry of Land and Resources, expanded this to twelve primary categories and 73 secondary categories, enhancing the definitions and refining the secondary categories while also adjusting some category names. In the third national land survey, land was classified into 13 primary categories and 73 secondary categories. Over time, as concepts and methods for managing geographical entities have evolved, land use classification has become increasingly detailed and has undergone numerous revisions and updates.

Different disciplines and departments have long maintained their own classifications of geographical entities, leading to a wide variety of classification standards. As a result, the classification systems for the same geographical entities are often incompatible and lack universality. The information management systems developed by different industries or for specific business applications vary significantly in how they manage geographical data. This diversity in classification standards, driven by different perspectives, causes overlap and inconsistency in entity categories. These discrepancies create significant challenges for the cross-sector transmission, application, sharing, and exchange of geographical information [11,18]. For instance, in land-use classification, various countries and international organizations use their own classification systems tailored to their specific application objectives. This practice makes it challenging to exchange and share geographical data across different systems. Scholars have introduced the concept of semantic similarity for geographical entities to address these issues. They have analyzed different classification systems at a semantic level, developed semantic reference trees, and investigated methods for integrating and making different classification systems interoperable [19–22].

Recently, many scholars proposed developing a unified classification system for natural resources, suggesting exploratory principles and ideas for this approach [23,24]. However, this task is highly challenging. Creating a geographical entity classification system that is universally applicable and shareable across different sectors and applications is extremely difficult [25]. Classification is often closely tied to specific application needs and management perspectives, which are influenced by human subjectivity. As a result, multiple classification systems have long coexisted both between industries and within sectors, continually evolving to meet changing application needs [26].

Considering the fact that multiple classification systems coexist for the same geographical entity in different industries, sectors, and applications, and in order to improve the cross-sector transmission, management, sharing, and application of geographical entity information, this study explored developing a geographical entity management method that accommodates multiple classification systems based on directed hypergraph theory. The method develops a unified expression model by integrating and transforming multiple classification systems. This model allows for the unified management of geographical entities from different classifications within a single data table, facilitating the transmission and sharing of geographical information across different systems and departments. The rest of the article is organized as follows. Section 2 examines the characteristics of various classification systems and develops a logical expression to represent these multiple classifications. Section 3 introduces an excellent data structure for logically representing multi-classification and explains how to perform conversion operations between different classification systems within the unified expression model. Section 4 applies this model and data structure to unify three different land use classification systems, demonstrating how to manage diverse land use data within PostgreSQL. This section evaluates the method's feasibility and its benefits for sharing and exchanging geographical information. Section 5 presents the discussion and conclusions.

2. Logical Expression of Multi-Classification

2.1. Multiple Classifications and Their Characteristics

Classification has an obvious hierarchical nature, and the classification system is usually an ordinary tree structure that can be logically represented by a hierarchical model. For example, Figure 1 illustrates two land classification systems. In the current land use classification (Figure 1A), the woodland is categorized into seven secondary categories. In contrast, the third national land survey classifies woodland into four secondary categories, with mangrove land, forest swamp, and scrub swamp being sub-categories of wetland (Figure 1B). Thus, the same seven geographical entities are classified into different primary categories across these two classification systems. If different departments or applications build their data management systems based on these two classification systems, it will lead to varying ways of organizing and managing the same entities. Additionally, the unique coding of these entities in different systems will differ. This creates significant obstacles to sharing and exchanging geographical information between departments and applications.

Addressing the challenge of multiple classifications for the same geographical entity by developing methods to merge and unify different classification systems is a valuable approach. This effort aims to enhance the organization and management of geographical entities and support the sharing and exchange. As early as 1959, Guttenberg defined the concept of land use with multiple dimensions and attempted to create a distinct land use classification system. His idea of multi-dimensional classification introduced a new perspective for categorizing land use [27]. In this paper, the "multi-classification" refers to different classification systems used for the same geographical entity. The integration and unified expression of these multiple systems are achieved through analytical reconstruction. Unlike a single hierarchical classification, which organizes data in a straightforward tree structure, multi-classification must simultaneously represent various systems within a unified framework. This creates a more complex logical structure. Simple hierarchical models are effective for one-to-many relationships (1:m) but cannot clearly depict how the same entity relates to multiple classification systems. The network model utilizes a graph structure to describe the connections between things and can more intuitively express the many-to-many associations (m:n) that prevail among entities in the real world. In the real world, multi-dimensional higher-order associations are prevalent among entities, and such complex associations often exhibit network-like characteristics (e.g., social relationships, biological networks, and literature citations), and the network model is capable of expressing both one-to-many associations and many-to-many associations [28].

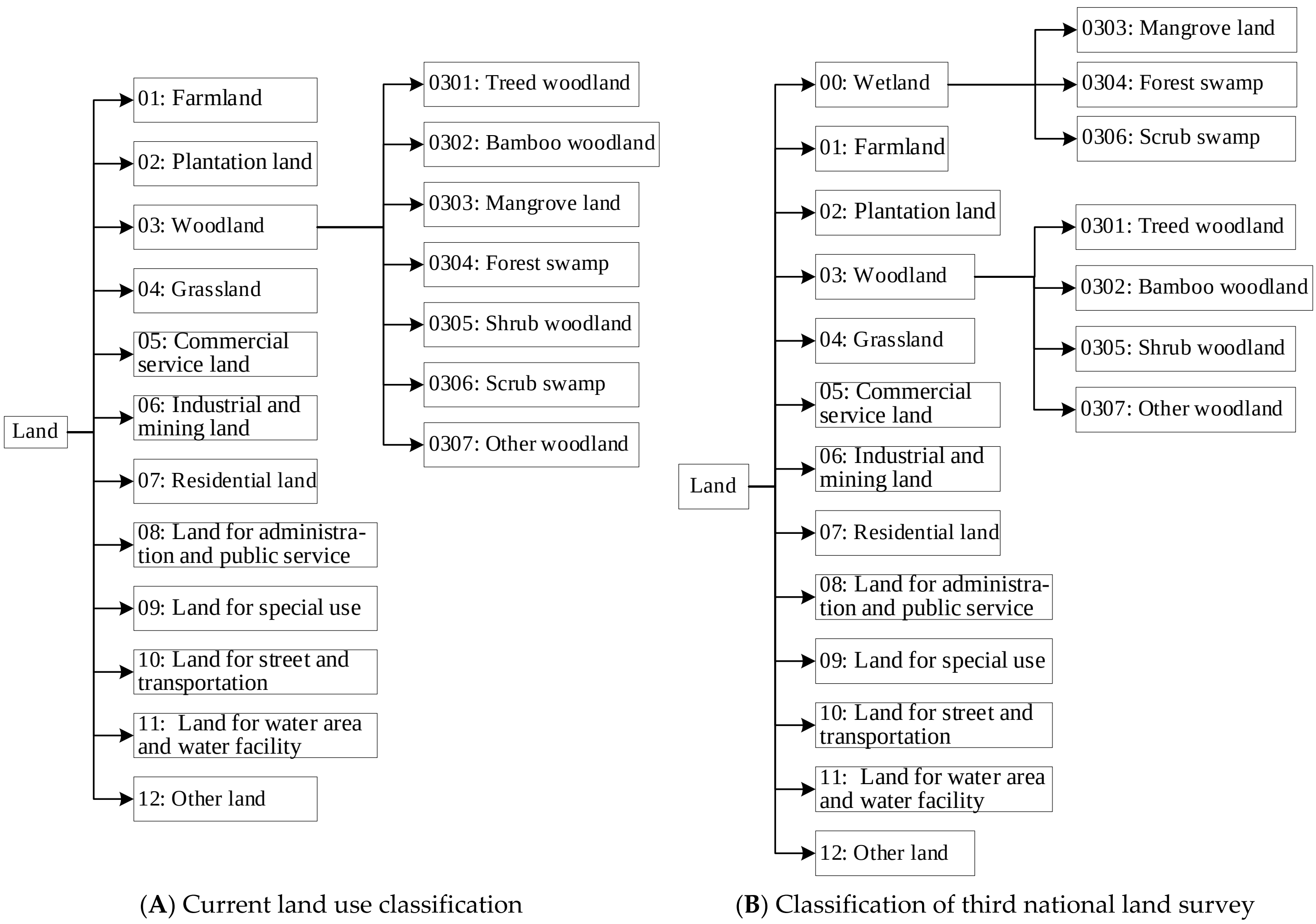


Figure 1. Different classifications of woodland.

Therefore, the network model, which uses a graph structure, is better suited for intuitively describing complex, multivariate relationships among objects in the real world. However, in a standard graph, an edge can only connect two nodes (entities), representing binary associations. As a result, the typical network model cannot effectively capture or represent the complex, multi-dimensional relationships between multiple entities and classification systems using just a single arc edge. A hypergraph, an extension of the traditional graph structure, allows a single edge to connect multiple nodes simultaneously. This capability makes hypergraphs well-suited for representing complex relationships involving multiple entities at once. They are an ideal mathematical model for expressing multi-element and multi-dimensional associations [28,29]. In order to better describe the multiple and complex connections of the same geographical entity in multiple classification systems, this study constructed a logical representation of multi-classification of geographical entities based on the hypergraph theory (Figure 2).

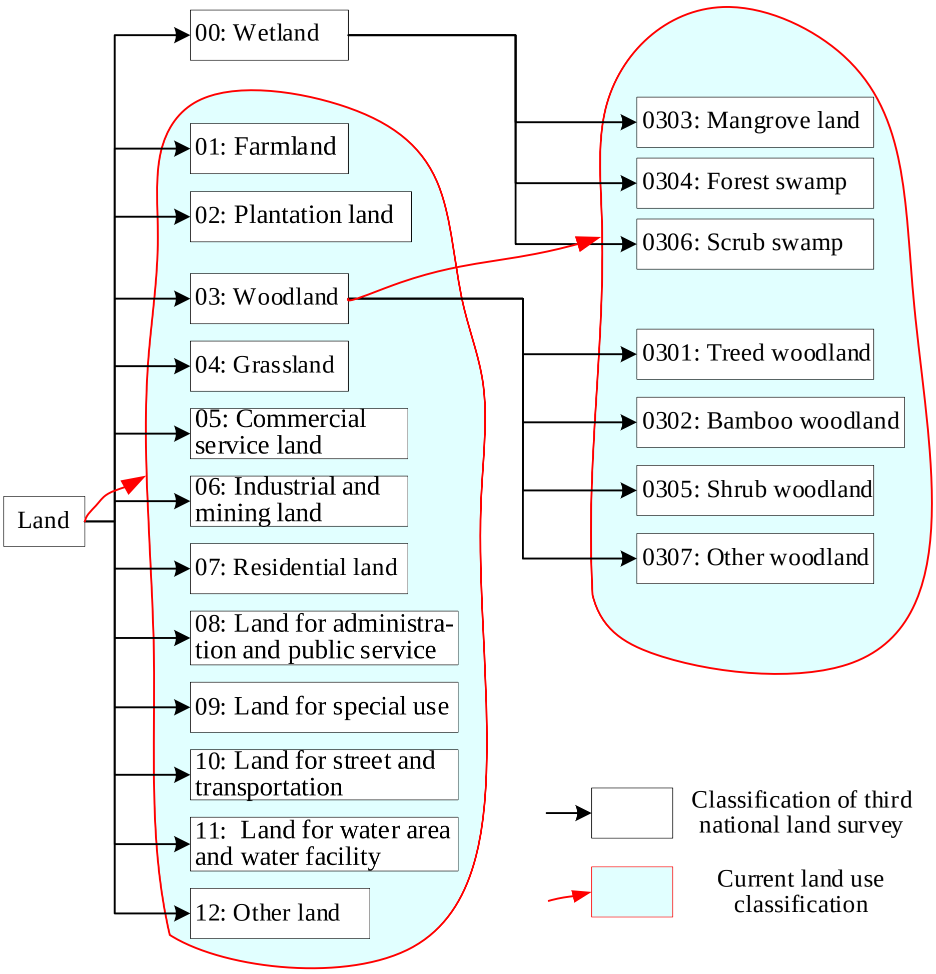
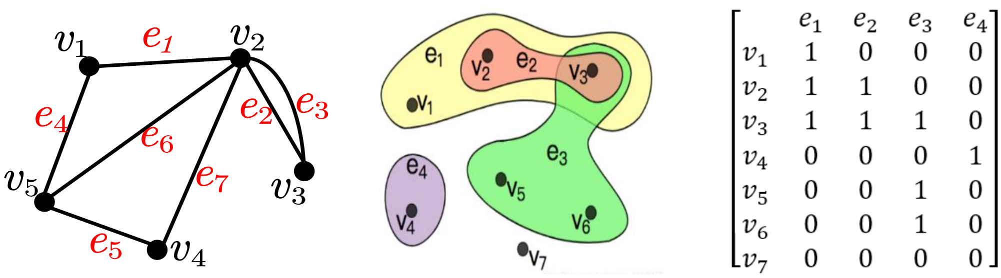


Figure 2. Unified representation of different classification systems.

2.2. Hypergraph Theory

Berge first introduced the concept of hypergraphs in 1973 [30]. A hypergraph is a theoretical approach combining graph theory and set theory and is a generalized variant of traditional graphs (Figure 3) [31]. The arc edges of hypergraphs are called hyperedges, and a hyperedge can connect multiple vertices, which can help to more intuitively and naturally describe the prevailing multi-element higher-order associations between entities. Hypergraphs have been successfully applied in various computer vision tasks, including classification and retrieval [32–34]. Hypergraphs offer a broad, complex, and detailed framework that enhances the ability to describe relationships between entities in the real world. This makes them particularly well-suited for addressing complex network problems.



(**A**) Graph (**B**) Hypergraph (**C**) Association matrix

Figure 3. Graph and hypergraph.

Let the vertices $V = \{v_1, v_2, \cdots, v_n\}$ be a finite set. Then, the hypergraph $H = \{e_1, e_2, \cdots, e_m\}$ on V can be defined as a finite subset of clusters of V, such that $e_i \neq \varnothing$ and $U_{i=1}^{m} e_i = V$. Here, $v_1, v_2, \cdots v_n$ are the vertices of the hypergraph, and $e_1, e_2, \cdots, e_m$ are the hyperedges of the hypergraph. Since the hyperedge e_i is a non-empty subset of the vertices, a hyperedge can contain multiple vertices. The hypergraph shown in Figure 3B can be represented as $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}, E = \{e_1, e_2, e_3, e_4\} = \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$.

From an application perspective, a hypergraph is composed of a set of vertices V and a set of hyperedges E. In this model, real-world entities are represented as vertices, while hyperedges are a finite set of vertices that illustrate the connections between these entities. Mathematically, a hypergraph can be represented using an association matrix, which captures these relationships [35]. As shown in Figure 3C, the vertices of a hypergraph are the rows of the matrix, the hyperedges are the columns of the matrix, and if a vertex belongs to a hyperedge, the intersection of the vertex and the hyperedge in the matrix is one; otherwise, it is zero. The structure of an association matrix can be represented as a two-dimensional array in computers, which is very convenient for the implementation of computer programs and arithmetic processing.

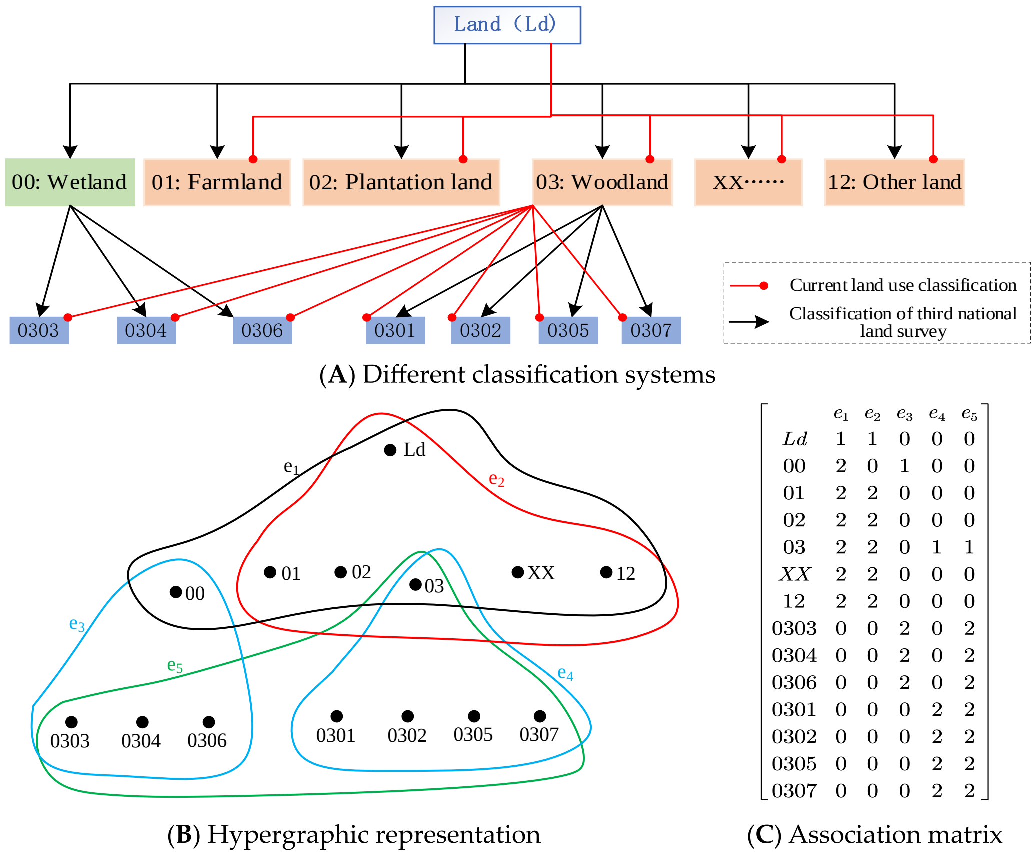
2.3. Hypergraph-Based Representation of Multi-Classification

Although the nodes in a hypergraph represent the entities themselves, and the hyperedges can connect multiple nodes, this structure is particularly effective for representing multiple complex relationships between entities. However, when representing multi-classification, it is important to not only integrate and reconcile different classification systems but also preserve the hierarchical structure of each individual system. This ensures that hierarchical relationships within each classification are maintained, allowing for the efficient extraction of a specific classification system based on application or management needs. To clarify the hierarchical relationships between parent and subclass entities in a classification system, it is essential to use directed hypergraphs. In a directed hypergraph, each hyperedge has a direction that indicates a starting point and an endpoint. This direction helps differentiate between parent classes and subclasses, making it possible to clearly represent the hierarchical structure. In this study, we used directed hypergraphs to create a unified expression model that integrates multiple classification systems for geographical entities.

In this study, the geographical entity multi-classification model based on directed hypergraphs involves three types of relationships between vertices and hyperedges. In the association matrix of the hypergraph, A_{ij} is the value of node v_i in the hyperedge e_j in the association matrix. Then, the value of each element is as in Equation (1). If the weight of A_{ij} is zero, it means that node v_i does not belong to hyperedge e_j; if the weight A_{ij} is one, it means that node v_i is the starting point (head) of hyperedge e_j, i.e., it is a parent class in the classification system; if the weight A_{ij} is two, it means that node v_i is an endpoint (tail) of hyperedge e_j, i.e., it is a subclass of the classification system. In this way, the multi-dimensional association between multiple nodes can be well-expressed by a directed hypergraph, and the hierarchical relationship between nodes can also be expressed.

$$A_{ij} = \begin{cases} 1, v_i \in head_j \\ 0, v_i \notin e_j \\ 2, v_i \in tail_j \end{cases} \quad (1)$$

As shown in Figure 4A, using land classification as an example, the current land use classification [17] divides land into twelve primary categories (in order to simplify the graphical representation, the category "XX......" represents eight land categories coded 04-11). According to the classification of third national land survey, land is divided into thirteen primary categories, including an additional category called "wetland", which is not present in the current land use classification. Furthermore, the differences between the two systems are more pronounced at the secondary category level. For example, some land categories classified as woodland in the current land use classification are categorized as subtypes of wetland in the classification of third national land survey. Clearly, the multiple classifications for land form a complex, high-dimensional network. This complexity can be effectively represented using the directed hypergraph structure (Figure 4). Figure 4C illustrates the association matrix of the directed hypergraph, and the element value of 1 in the matrix indicates that the node is in the hyperedge and is the starting point of the hyperedge, i.e., the parent class of the current level in the classification system.



	e_1	e_2	e_3	e_4	e_5
Ld	1	1	0	0	0
00	2	0	1	0	0
01	2	2	0	0	0
02	2	2	0	0	0
03	2	2	0	1	1
XX	2	2	0	0	0
12	2	2	0	0	0
0303	0	0	2	0	2
0304	0	0	2	0	2
0306	0	0	2	0	2
0301	0	0	0	2	2
0302	0	0	0	2	2
0305	0	0	0	2	2
0307	0	0	0	2	2

Figure 4. Multi-classification of land use.

3. Data Structure for Multi-Classification

The data structure is a key factor influencing the performance of storing and querying geographical information. Designing an appropriate data structure based on the characteristics of the logical model is crucial for implementing the multi-classification model for geographical entities. Hypergraphs, being a more advanced and complex extension of graph structures, involve more intricate data operations and physical implementations compared to traditional graphs. To support the representation of multi-classification and manage geographical entity data, this study designed a data structure that includes category nodes (ClsNode), hyperedges (HyperEdge), and node references (NodeRef, which link category nodes across different hyperedges). This structure, shown in Figure 5, enables the implementation of a hypergraph model for multi-classification of geographical entities. In this data structure, the category node ClsNode corresponds to the vertices of the hypergraph, representing different categories of real-world entities, in which the relevant information of entity categories is recorded. The HyperEdge corresponds to the hyperedge of the hypergraph and represents the specific categorization of the entities in a specific viewpoint. The NodeRef records whether the category node is referenced by a specific hyperedge. This reference system uses a bidirectional linked list, which effectively captures the multi-dimensional associations between each category node and various classification systems, as well as the relationships among category nodes and other nodes within a hyperedge.

```
// Definition of class node
typedef struct ClassNode
{
    long     cls_id;       // class node ID
    string   cls_name;     // class node name
    string   cls_note;     // description of the class node
    NodeRef* ref_id;       // first reference of this class node
};

// Definition of node reference
typedef struct NodeRef
{
    long       ref_id;     // node reference ID
    ClassNode* cls_id;     // corresponding class node
    HyperEdge* edge_id;    // hyperedge ID containing the noderef
    long       view_id;    // data view ID
    NodeRef*   prev_id;    // previous node reference
    NodeRef*   next_id;    // next node reference
    NodeRef*   ref_link;   // neighboring node reference
};

// Definition of hyperedge
typedef struct HyperEdge
{
    long      edge_id;     // hyperedge ID
    string    edge_name;   // hyperedge name
    string    edge_note;   // description of the hyperedge
    NodeRef*  first_ref;   // first node reference of the hyperedge
};
```

Figure 5. Data structure for multi-classification.

Based on this data structure, the multi-classification of the land shown in Figure 4 can be represented by ClsNode, HyperEdge, and NodeRef as in Figure 6. The rectangles in the first column on the left side of the figure are all the category nodes constituting the hypergraph, the rectangles in the first row on the top side represent each of the hyperedges of the hypergraph, and the white rectangles at the intersections of the rows and columns represent a NodeRef of a category node in the different hyperedges. It can be seen that if a category is included in a hyperedge, the intersection of the row where the category is located and the column where the hyperedge is located is the node reference of the category in the hyperedge; if the category is not included in the hyperedge, the intersection of the row where the category is located and the column where the hyperedge is located is null, and all the node references in the columns where the hyperedge is located correspond to the category nodes that constitute all the categories of the hyperedge. Since the hyperedge and node references are bi-directionally associated with each other, it is very easy to traverse a hyperedge to obtain all the categories constituting the hyperedge, and it is also very easy to traverse from a node reference (category) to other node references (categories) associated with it, i.e., to obtain the same category associated with the same category in more than one classification system.

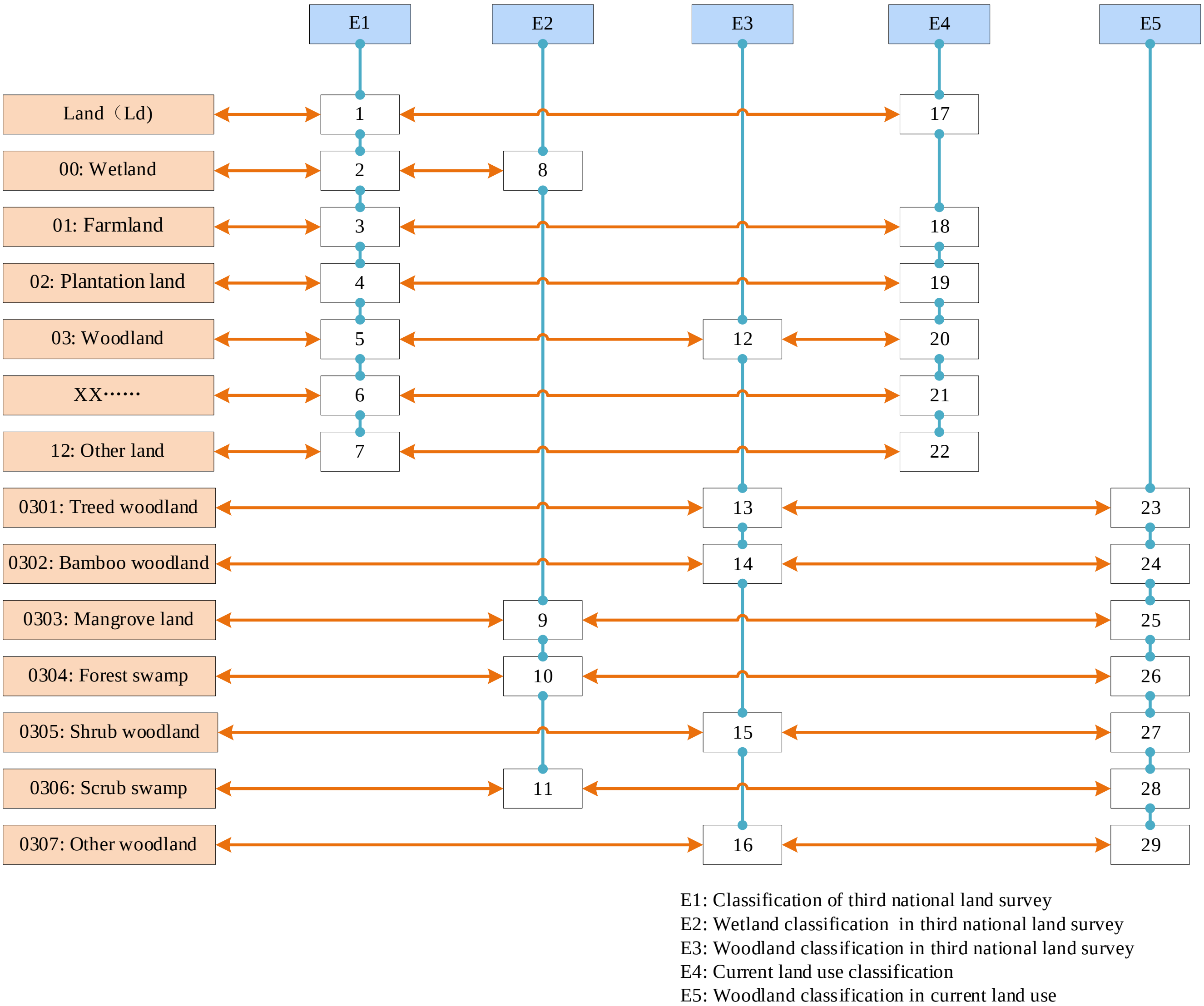


Figure 6. Multi-classification representation of land use based on the data structure of this study.

In this study's data structure, the bidirectional linked list design of the node references (NodeRef) enables detailed analysis of the hypergraph structure from both horizontal and vertical dimensions. The horizontal dimension refers to the node references of each category node across different hyperedges, which represent various classification perspectives of geographical entities. The frequency with which each category node is referenced in this dimension indicates the number of different classification systems that include that entity category. For example, the category node "Land (Ld)" is referenced twice (node references 1 and 17) in the hypergraph, indicating that "Land" can be categorized from two perspectives, namely the third national land survey (hyperedge E1) and the current land use classification (hyperedge E4). Similarly, "woodland" is cited in hyperedge E1, E3, and E4, i.e., "woodland" is categorized from three perspectives in this multi-classification system. The vertical dimension refers to the references of all nodes within each hyperedge. This dimension represents the classification results from the current perspective, showing the parent category along with all its subcategories under that specific classification system. For example, the column where hyperedge E3 is located contains five node references. This indicates that node reference 12 corresponds to the category "woodland", which is divided into four subcategories in the third national land survey. Specifically, node references 13 through 16 represent "treed woodland", "bamboo woodland", "shrub woodland", and "other woodland".

Moreover, the combination of vertical and horizontal dimensions can be analyzed to quickly extract a complete individual classification system from multi-classification. Because each column in the vertical represents a hyperedge, the first node reference in the hyperedge represents the parent node of the classification, and the rest of the node references represent all the subclasses of the parent node. Therefore, starting from a single hyperedge in the vertical dimension and combining it with correlations between that hyperedge and the other columns in the horizontal dimension, a complete construction of the classification system represented by that hyperedge is possible. For example, starting with the vertical dimension hyperedge E1, we can interpret it as follows: according to the third national land survey, land can be categorized into types such as wetland, farmland, plantation land, woodland, "xx. . . . . .", and other land. Combining this with horizontal dimension, node references 2 and 5 in hyperedge E1 are related to node reference 8 in hyperedge E2 and node reference 12 in hyperedge E3. This means that "wetland" (node reference 2/8) can be further divided into the subcategories of mangrove land, forest swamp, and scrub swamp (corresponding to node references 9–11). Similarly, the "woodland" category at node reference 5/12 can be further divided into treed woodland, bamboo woodland, shrub woodland, and other woodland (corresponding to node references 13–16). By using this method, the complete classification system for land in the third national land survey, as shown in Figure 6, can be fully reconstructed (Figure 7). By following this process, each classification system within the multi-classification can be individually and quickly reduced to a complete classification tree for geographical entities. The data structure designed in this study facilitates both the fusion and unified expression of multiple classification systems while preserving the hierarchical structure of each individual system. This approach enables an effective and coherent integration of multiple classifications.

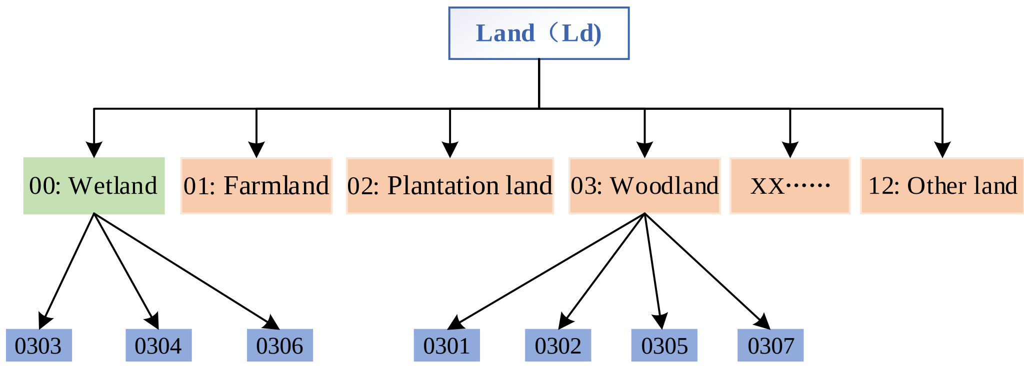


Figure 7. Land use classification system (perspective of the third national land survey in Figure 6).

4. Geographical Entity Management and Case Study

4.1. Multi-Classification of Land Use

Today, society faces significant challenges related to resources, population, food, and the environment. Among these, the use and protection of land resources are crucial because they directly affect food supply, population survival, and environmental quality. The effective management of land resources is essential for sustainable economic and social development. The current land use serves as a fundamental reference for countries when developing major strategies and important policies for economic and social development. As social management practices have become more detailed and sophisticated, the understanding of land use status has also evolved. China has many land use classification standards. This study analyzed the evolution of land classification systems by examining three examples: the current land use classification [17], the Classification of third national land survey, and the "Standard for Classification of Urban Green Space" [36]. The analysis focuses on how these classifications have developed in the context of natural resource management and urban construction management. The differences between various classification systems were analyzed, and the three land classification systems were fused and unified using the multi-classification model of geographical entities proposed in this paper. This approach explored the possibilities for unified management and cross-sectoral sharing and exchange of land use data across different industries and applications.

To address the demands of economic and social development and more detailed management, the classification of third national land survey was developed based on the current land use classification [17]. This updated classification has refined and merged some land categories. For example, "wetland" was introduced as a new primary category, and certain subclasses of "woodland" from the previous classification have been reclassified as subclasses under "wetland". The Standard for classification of urban green space [36] issued by the Ministry of Housing and Urban-Rural Development divides urban greenland into five primary categories, which are a further refinement of the classification of "park and greenland" in the third national and survey. These three standards all pertain to the classification of land as a geographical entity. However, the business management needs of various applications and departments differ, leading to both overlaps and significant differences among the different classification systems. Using the directed hypergraph model for multi-classification proposed in this study, the three classification systems were fused and expressed (Figure 8). Given the large number of secondary categories, the figure focuses on the primary categories of each classification system and highlights some representative secondary categories for clarity. Other secondary categories have been omitted to maintain a clear and readable graph.

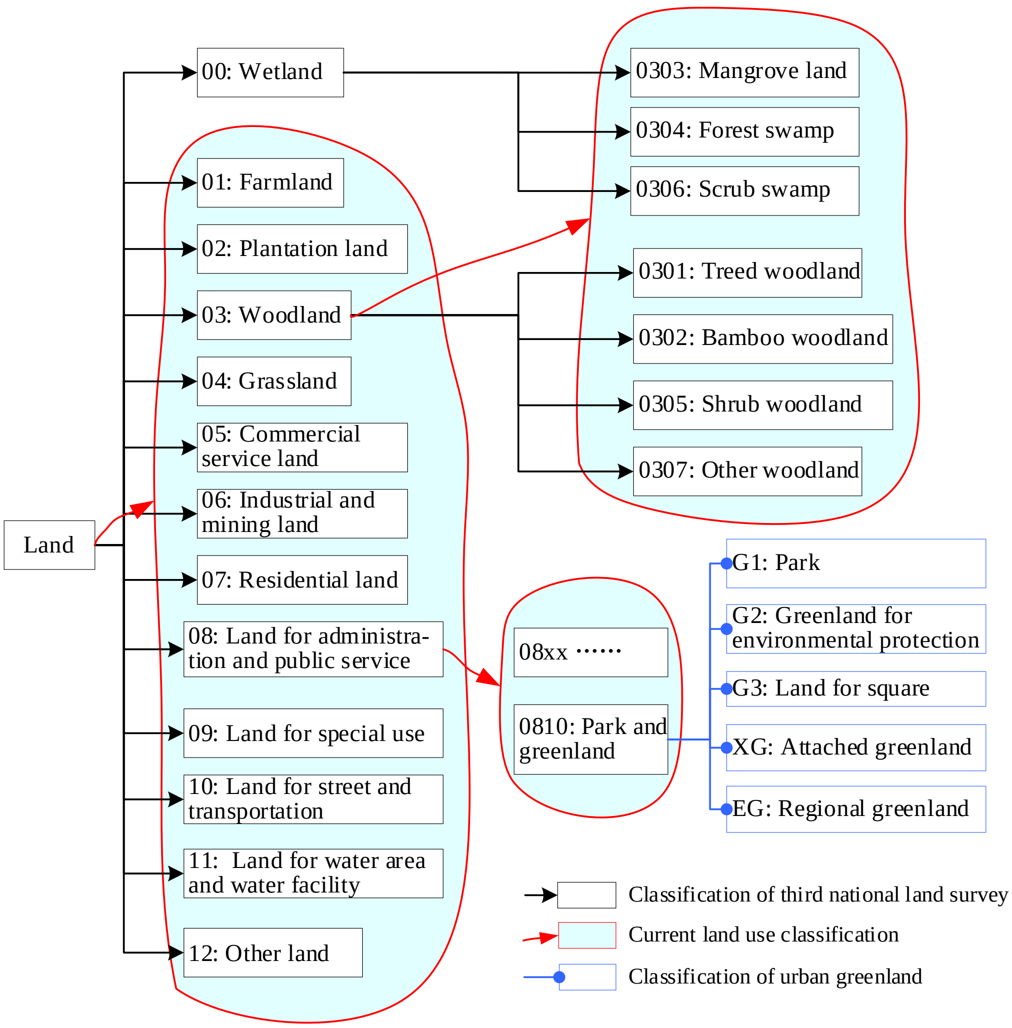


Figure 8. Multi-classification expression for land use.

4.2. Geographical Entity Management Based on Multi-Classification

Present-day geographical data management is primarily handled using relational databases. In this approach, geographical information is often organized based on geometric types like points, lines, and surfaces, or according to specific classification systems. This data is managed hierarchically [37], represented as various data layers in GIS softwares, and mapped to different relational tables in a database. Managing geographical entity information using relational data tables presents several challenges, especially in the era of big data. This approach often complicates data sharing and exchange. Recently, scholars have suggested focusing on geographical entities as the core for managing and analyzing geographical information to address these issues. Geographical entities are natural or artificial features that exist independently and can be uniquely identified [9,38]. They are also the cohesive nucleus of geographical information [10,39]. Their inherent properties remain unchanged regardless of shifts in management or application needs. Different departments may interpret and describe these entities from their specific business perspectives. Building on this concept, this study proposes a model for managing geographical entities across various industries and departments, leveraging a multi-classification system to support diverse business applications.

In this study, the multi-classification model for geographical entities integrates various classification systems, allowing for the addition of new systems as application requirements evolve. This model is inherently flexible and adaptable, making it less suited for structured relational databases. Instead, it requires an unstructured approach to effectively manage and describe geographical entities in practical applications. PostgreSQL, a widely used open-source relational database, excels in spatial data processing owing to its PostGIS extension, which fully supports the OpenGIS specification. Additionally, PostgreSQL handles unstructured data effectively through its JSONB type. JSONB supports the nesting of document-type objects and allows for index creation, resulting in improved storage efficiency and faster query and retrieval speeds. Given the schema-less and unstructured nature of geographical entity data under multiple classification systems, along with the need for spatial operations and analysis, this study implemented geographical entity management for multi-classification using PostgreSQL.

Because the data structure design of this study is oriented to multi-classification, the PostgreSQL database requires three relational tables to manage the multi-classification model. Among them, the ClassNode table stores the category nodes, the HyperEdge table stores the hyperedge structure, and the NodeRef table stores the node references. The structure definitions of these three relational tables are shown in Table 1, Table 2, and Table 3, respectively. Geographical entities, as independently existing features, are not affected by their classification systems, and therefore, a separate GeoEntity table was used to store geographical entity information (Table 4). This table mainly includes the entity unique identifier "ent_id", the geometry field "ent_shp", and the JSON type field "ent_attrs". The "ent_attrs" field stores various descriptive information about the entities provided by different departments and business applications. Geographical entities are associated with specific classification systems through data views. A view is a virtual table derived from one or more tables in the database, and only the definition of the view is stored in the database, not the data related to the view. Therefore, the view (Table 5) consists of the unique identifier "view_id", description field "view_note", and field "sql_stat". The text type "sql_state" records the view's SQL statement, and the relationship between the five tables is shown in Figure 9. This structure supports a geographical data management model based on multi-classification, as shown in Figure 10. Each geographical entity is uniquely stored in the GeoEntity table, while different classification systems are linked to these entities through distinct data views. Various application systems then use these views to manage and manipulate entity information. This approach emphasizes entity-centered management, allowing different applications and departments to describe entities according to various classification perspectives (views).

Table 1. ClassNode table structure.

Field Name	Value Type	Meanings	Description
cls_id	Serial	Class Unique ID	Automatic system generation and maintenance.
cls_name	Text	Class name	Class Node Name.
cls_note	Text	Class description	Other descriptive information relevant to the category.
ref_id	NodeRef	First node reference ID	The first reference of this class node in the hypergraph structure.

Table 2. NodeRef table structure.

Field Name	Value Type	Meanings	Description
ref_id	Serial	node reference ID	Automatic system generation and maintenance.
cls_id	ClassNode	Corresponding class node	Points to the actual class node.
edge_id	HyperEdge	The ID of the hyperedge	Which hyperedge the node reference belongs to.
view_id	Serial	View name	Corresponding data view.
prev_id	NodeRef	Previous node reference	The ID of the last node reference in this hyperedge.
next_id	NodeRef	Next node reference	The ID of the next node reference in this hyperedge.
ref_link	NodeRef	Neighboring node reference	Reference ID of this reference node in another hyperedge.

Table 3. HyperEdge table structure.

Field Name	Value Type	Meanings	Description
edge_id	Serial	Hyperedge unique ID	Automatic system generation and maintenance.
edge_name	Text	Hyperedge name	Description of hyperedge name.
edge_note	Text	Hyperedge description	Additional descriptive information related to this hyperedge.
first_ref	NodeRef	First node reference ID	Points to the first node reference associated with this hyperedge, i.e., the parent node of the classification.

Table 4. GeoEntity table structure.

Field Name	Value Type	Meanings	Description
ent_id	Serial	Entity unique ID	Automatic system generation and maintenance.
ent_shp	Geometry	Geometry of geographical entity	Geometric shape description.
ent_attrs	JSONB	Geographical entity attributes	Various attribute information related to the entity.

Table 5. View table structure.

Field Name	Value Type	Meanings	Description
view_id	Serial	View unique ID	Automatic system generation and maintenance.
view_note	Text	View description	Description of this view.
sql_stat	Text	SQL statement	SQL statement to create this view.

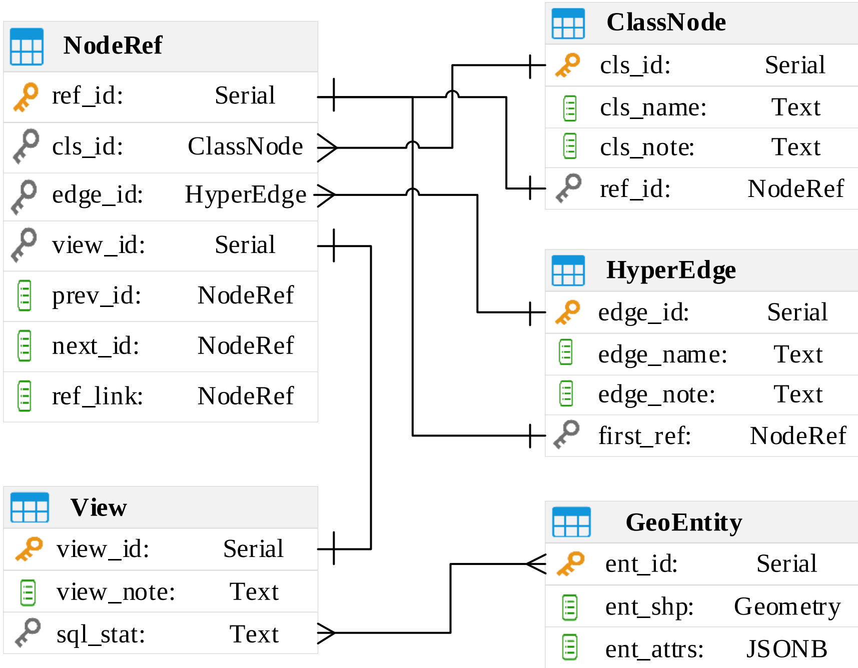


Figure 9. Data tables and their relationships in the multi-classification model.

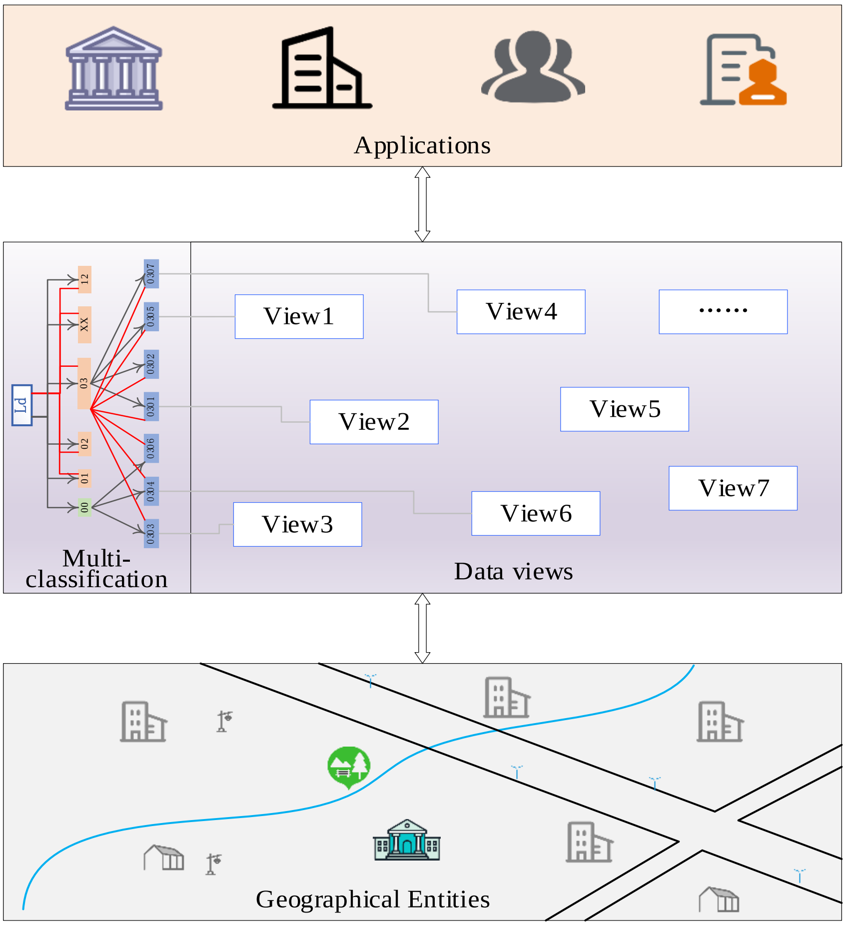


Figure 10. The geographical data management model based on multi-classification.

4.3. Prototype Systems and Case Study

Based on these principles, this study developed a geographical entity management system designed for multi-classification. This system integrates and reconstructs various classification systems (Figure 11) and provides a unified expression (Figure 12). It conducts unified management and application experiments for land use data across multiple classifications. The system includes case studies for managing geographical entities from different perspectives, such as the land use status, the third national land survey, and urban greenland management (Figure 13). This approach enhances the unified management of geographical information and supports the sharing and exchange of geographical entity information across different departments and applications. For example, in urban planning management, the unified geographical entity management system developed in this study allows the housing and construction department to swiftly identify the relationship between Greenland classifications and the classifications of third national land survey within the multi-classification system of land use. This system facilitates easy access to urban greenland status information from the third national land survey data, enabling seamless sharing and rapid conversion between national land survey data and urban greenland data. Similarly, when the natural resources department needs information on farmland classifications, it can directly retrieve the relevant data view from the node references associated with the corresponding hyperedge. This data view then allows the department to access the farmland classification data from the geographical entity data table.

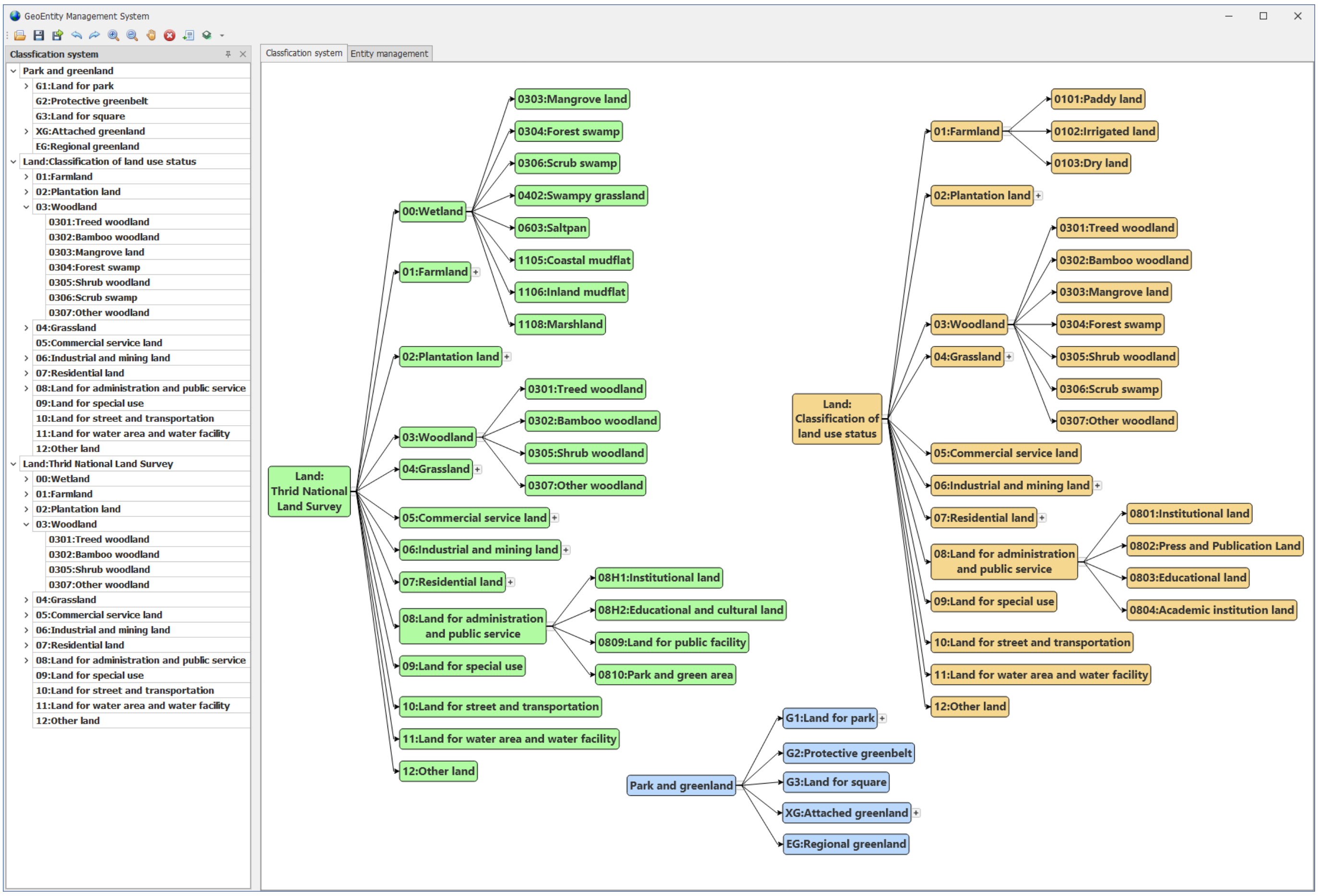


Figure 11. Different classification systems of land use.

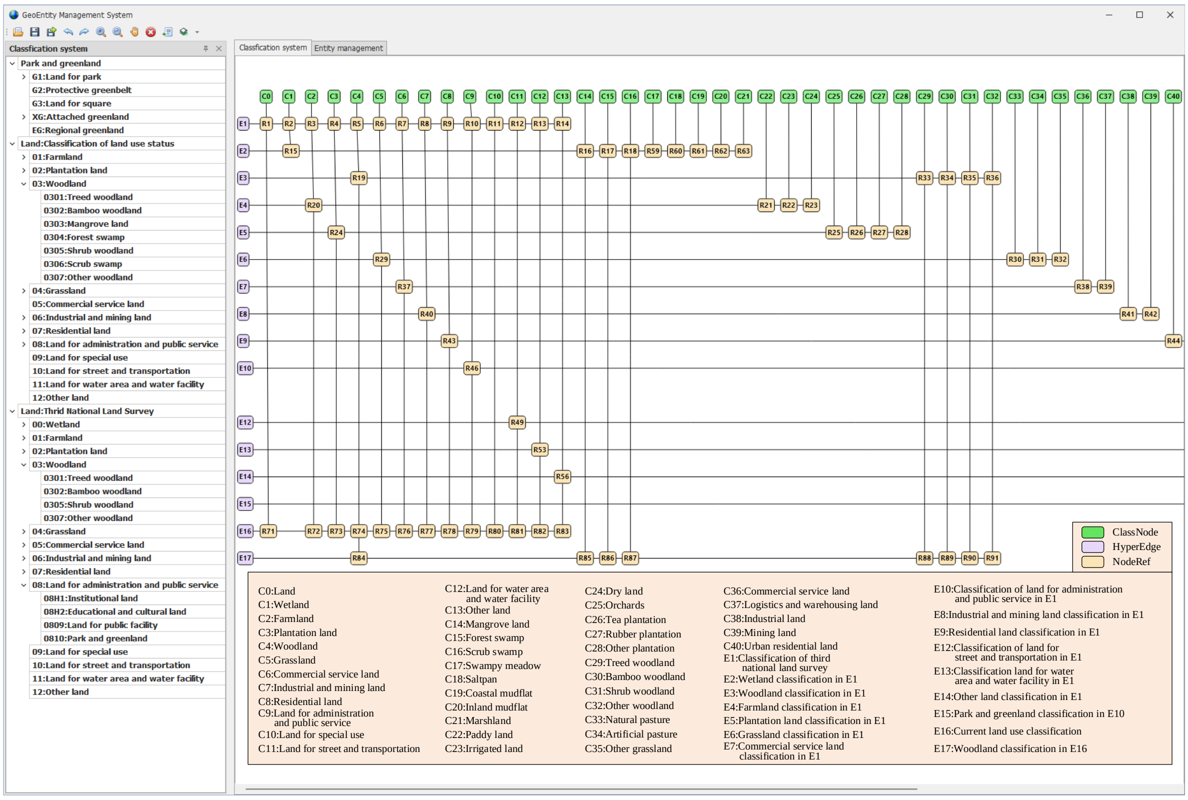


Figure 12. Unified expression of multiple classifications of land use.

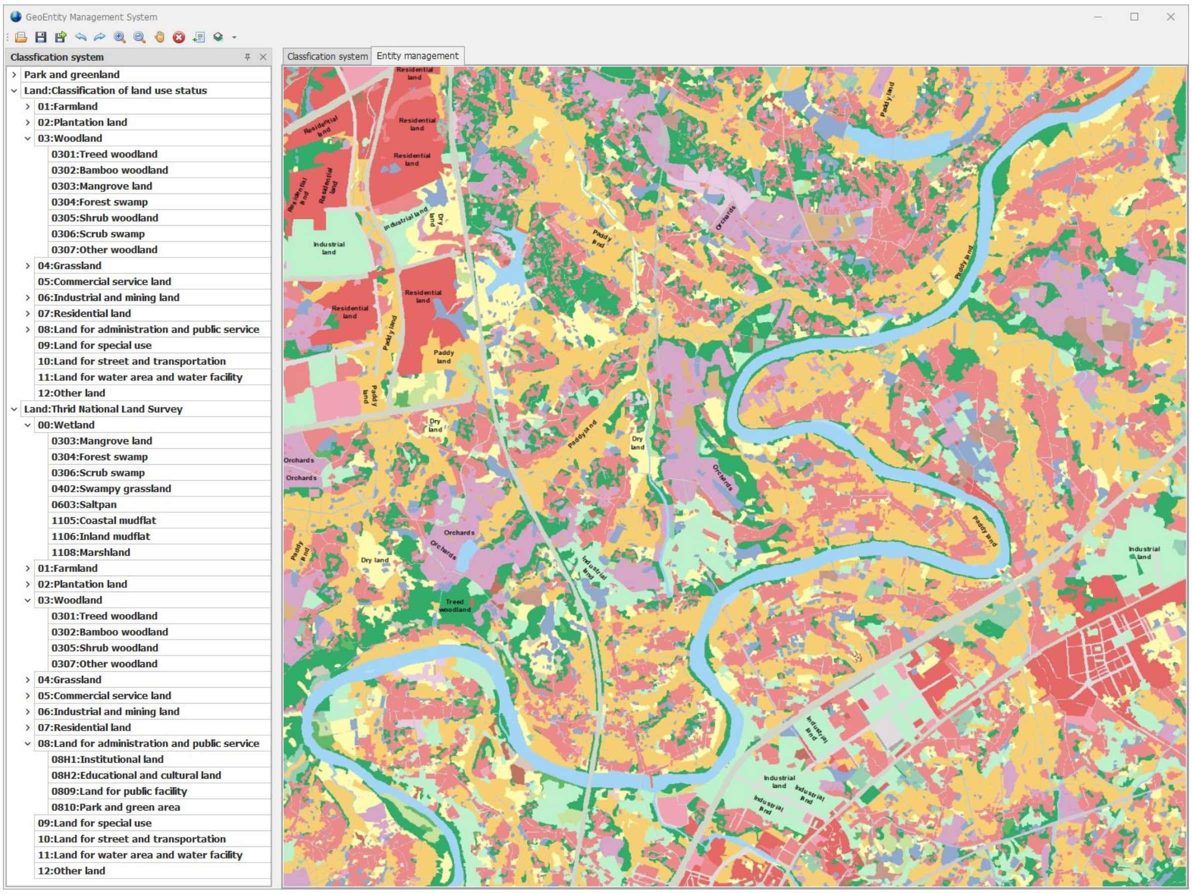


Figure 13. Land use data management based on multi-classification.

5. Discussion and Conclusions

In the following section, we discuss this study's geographical entity management model based on multi-classification from two aspects—geographical entity granularity and data query performance—and present the conclusions of this study.

5.1. Multiple Granularity of Geographical Entity

The real world is infinitely complex, and no information system can fully capture or describe it. To manage this complexity, we abstract the real world into various geomorphic entities based on specific spatial scales and application needs. These entities can be understood and analyzed at different scales; they can be broken down into smaller components or combined into larger entities [40]. For instance, in road traffic research, a road can be classified in various ways (Figure 14): it might be divided into two lanes going in opposite directions with a segregation facility in the middle, or it might be divided into three or four lanes going in the same direction. Additionally, the road can include various traffic facilities such as traffic lights. The different levels of granularity in road classifications affect the amount of information available about each entity. For example, the road entities depicted in Figure 14C, which include details like lane widths and traffic facilities, provide more granular information than those in Figure 14A,B. These differences in detail and granularity make it challenging to exchange and share geographical information across different systems and scales.

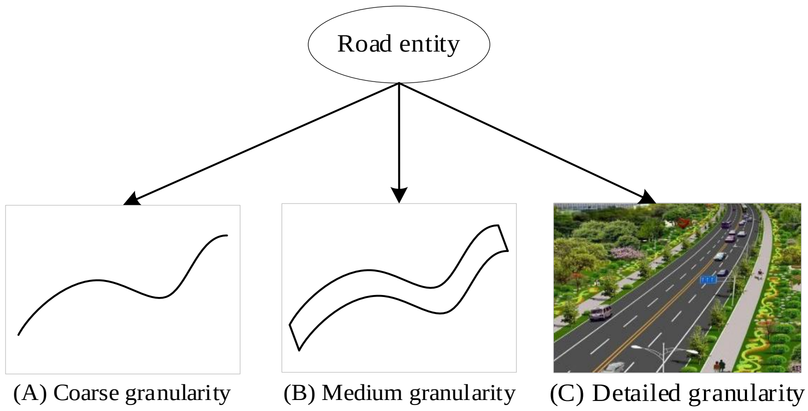


Figure 14. Different granular segmentation and representation of a road entity.

With advancements in computer simulation technology and geographical information science, the concept of geographical information classification granularity has emerged. This has led to a focus on constructing and managing multi-granularity spatio-temporal entities, which is now a key area of research in geographical information classification, and data management [3,37,41]. This study does not currently address how different departments may divide the same geographical entity into various granular levels. However, the granularity of entity divisions is a crucial factor affecting information exchange and sharing, just like classification systems. Further research is needed to explore how to express and manage geographical entities with multi-granular and to develop management models that accommodate these varied multi-granular.

5.2. Data Query Performance

Data query performance significantly impacts the value of a data management model. Often, the performance bottlenecks in business application systems are related to the speed at which the database responds to queries [42]. In this section, the performance of this study's geographical entity management model based on multi-classification is evaluated using a city's land use classification data. The data include 321,732 parcels. We compared two management approaches: the traditional hierarchical classification method and the multi-classification model proposed in this study. This comparison assesses the data query performance for each management approach. The database management system adopts PostgreSQL 13.14, and to avoid network impact, the database service and query client were deployed on the same computer with the following hardware environment: Windows 11 operating system, 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30 GHz processor, and 32 GB RAM.

In the traditional classification and hierarchical data management model, specific business application systems usually correspond to specific data classification systems. This often results in the need for separate databases—one for each land classification system. For example, managing three different land classification systems would require three distinct business databases. In contrast, the multi-classification model proposed in this paper simplifies this process. It uses a single data table to store and manage all geographical entities. The different classification systems are linked to these entities through data views, eliminating the need for multiple databases and streamlining data management. This approach allows for a single business database, where different data views are created based on various classification systems. Data operations are then handled through these views, simplifying management and reducing the complexity of data storage. In other words, in the example cited above, where managing three different land classification systems would require three distinct business databases, in the new model proposed in this study, only one business database needs to be constructed, where different data views are created based on various classification systems. Data operations are then handled through these views.

After constructing the above four databases in PostgreSQL, the data query performance test was conducted by querying information on land use types, specifically grasslands and community parks. We performed these tests with data volumes of 5000, 30,000, and 250,000 records, both with and without indexing. The results of these tests are presented in Table 6.

Table 6. Query performance.

Data Management Model	Classification System	Response Time (without Index)/ms			Response Time (with Index)/ms		
		3000	30,000	250,000	3000	30,000	250,000
categorization and hierarchization	Current land use classification	10.389	127.329	562.316	0.532	0.674	0.661
	Classification of third national land survey	10.763	128.935	549.648	0.586	0.639	0.675
	Classification of urban greenland	10.285	130.078	561.322	0.544	0.643	0.669
Multi-classification	Current land use classification	9.872	109.419	472.931	0.508	0.517	0.589
	Classification of third national land survey	9.365	108.197	480.129	0.515	0.508	0.593
	Classification of urban greenland	10.018	110.437	483.735	0.523	0.534	0.574

As can be seen from Table 6, when there are 30,000 data items in the data table, the response time of the traditional categorization and hierarchical data management method is 128.935 ms for querying "grassland" in the classification of the third national land survey without indexing, and the response time of the multi-classification data management model proposed in this paper is 108.197 ms. The query performance of the proposed method is 16.08% better than that of the traditional classification and hierarchical management method. In the case of querying "grassland" in the classification of the third national land survey with indexing, the response time of the two different geographical data management modes is 0.639 ms and 0.508 ms, respectively. The method proposed in this paper is 20.5% faster. In other cases, the overall trend is similar, and the query performance of this study's method is faster by 10% on average. Therefore, the geographical entity management method based on multi-classification proposed in this paper not only achieves the unified management of geographical entity data and reduces redundant database construction but also demonstrates strong query performance.

5.3. Conclusions

This study addresses the issue of overlapping and redundant classification systems for the same geographical entity, which leads to duplicated database construction and difficulties in data sharing between different industry departments and business applications. To tackle this problem, the study focuses on creating a unified expression for multiple classifications of geographical entities. It analyzes the characteristics of multiple classifications and proposes a logical model based on directed hypergraph theory for integrating and unifying multiple classification systems. Additionally, a data structure is designed to support this multi-classification model. Building on this approach, the study uses PostgreSQL to manage the storage of geographical entities according to the multi-classification model. Using urban land use data as an example, an experiment is conducted to demonstrate unified management of entity data based on this multi-classification model. The performance of the proposed geographical entity management method is then compared with the traditional relational database management system. The results demonstrate that the proposed method not only reduces data duplication but also improves query performance. This confirms the feasibility and advantages of using this method for unified geographical entity management across different industries. Additionally, it enhances the transmission, sharing, and exchange of geographical entity information across various departments and applications.

Author Contributions: Lin Shi: Conceptualization, Methodology, Formal Analysis, Visualization, Writing—original draft. Xiaoji Lan: Methodology, Resources, Investigation, Supervision. Ming Xiao: Data curation, Visualization, Writing—review & editing. Ning Liu: Data curation, Validation, Software. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China [Grant No. 41561085].

Data Availability Statement: Data will be made available on request.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their constructive comments and suggestions.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Mao, Z. "Scientific Spirit" or "Scientific Obscurantism:"? An Analysis Starting from Xunzi's aradox of "not Seeking to Know Nature". *Stud. Dialectics Nat.* **2023**, *39*, 127–132. [CrossRef]
2. Zhu, Y. The Thoughts on New Science Education in New Education Experiment. *Educ. Sci. China* **2019**, *2*, 109–120. [CrossRef]
3. He, J.; Li, X. Research on the Ontology of Geographic Information Classification and Coding. *Geogr. Territ. Res.* **2002**, *18*, 1–7. [CrossRef]
4. Yu, T.; Li, R.; Wu, H. An Object-oriented Representation Method for Multi-granularity for Spatio-temporal Position of Geographical Entities. *J. Geo-Inf. Sci.* **2017**, *19*, 1208–1216. [CrossRef]
5. Li, J.; He, Z.; Yan, X.; Zhu, Q. A Semantic Categorization of Hydrological Domain Based on Formal Concept Analysis. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 976–981+988. [CrossRef]
6. Wang, J.; Lin, H.; Ran, Y.; Zhou, Y.; Song, J.; Du, J. A study of Earth System Science data classification for data sharing. *Adv. Earth Sci.* **2014**, *29*, 265–267+273–274.
7. Liao, S.; Jiang, L. Study on Classification System of Data for Earth System Science. *Prog. Geogr.* **2005**, *24*, 93–98. [CrossRef]
8. Zheng, Y. Design and Engineering Practices of National Fundamental Geo-Entity Relationship. *AEDMR* **2024**, *1*, 66. [CrossRef]
9. Chen, R.; Lei, J.; Yao, H.; Li, T.; Li, S. Anchor-Enhanced Geographical Entity Representation Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–15. [CrossRef]
10. Lv, G.; Yu, Z.; Zhou, L.; Lan, X. Construction of Geographical Entity Classification and Coding System. *Mod. Surv. Mapp.* **2019**, *42*, 1–6. [CrossRef]
11. Wang, X.; Pu, L. A review of the classification of natural resources. *Resour. Sci.* **2021**, *43*, 2203–2214. [CrossRef]
12. *GB/T 13923-1992[S/OL]*; Classify and Codes for the National Land Information. National Technical Committee for Geographic Information Standardization: Beijing, China, 1993.
13. *GB/T 13923-2006[S/OL]*; Specifications for Feature Classification and Codes of Fundamental Geographic Information. National Technical Committee for Geographic Information Standardization: Beijing, China, 2006.

14. *GB/T 37118-2018[S/OL]*; Specification for Geo-entity Spatial Data. National Technical Committee for Geographic Information Standardization: Beijing, China, 2018.
15. *DB32/T 4292-2022[S/OL]*; Classification and Codes for General Geographical Entities of the Police. Public Security Bureau of Jiangsu Province: Nanjing, China, 2022.
16. *GB/T 21381-2008[S]*; Coding rules for Entity Identification Code of Traffic Management Geographical Information—City Road. National Technical Committee for the Standardization of Intelligent Transportation Systems: Beijing, China, 2008.
17. *GB/T 21010-2017[S]*; Current Land Use Classification. National Technical Committee for Standardization of Natural Resources and Territorial Spatial Planning: Beijing, China, 2017.
18. Chen, C.; He, J. Data Logic Study for the Resources and Environment Information System Design in a Feature based Approach. *Acta Geogr. Sin.* **1999**, *54*, 87–95. [CrossRef]
19. Di Gregorio, A.; O'brien, D. Overview of land-cover classifications and their interoperability. In *Remote Sensing of Land Use and Land Cover: Principles and Application*; CRC Press: Boca Raton, FL, USA, 2012; pp. 37–47.
20. Jepsen, M.R.; Levin, G. Semantically based reclassification of Danish land-use and land-cover information. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 2375–2390. [CrossRef]
21. Niu, J.; Xu, F.; Li, Z.; Hong, X. Land Use Zoning Model Considering Semantic Similarity of Geographic Entity. *Geomat. Inf. Sci. Wuhan Univ.* **2015**, *40*, 816–822. [CrossRef]
22. Huang, J.; Shi, W.; Zhang, X.; Li, D. Semantie Analysis between Different Classification Systems: Take Natural Resource Survey as an Example. *Geogr. Geo-Inf. Sci.* **2020**, *36*, 16–21. [CrossRef]
23. Peng, L.; Wang, Y.; Yin, Z.; Li, Z.; Yang, G. Current status of natural resources classification and unified classification of natural resources for the future. *Geol. Bull. China* **2022**, *41*, 2106–2113. [CrossRef]
24. Sun, X.; Liu, X.; Liu, X.; Gao, J.; Zhu, Z.; Zheng, W. Classification system of natural resources for integrated management. *Resour. Sci.* **2020**, *42*, 1860–1869. [CrossRef]
25. Mark, D.M. Toward a theoretical framework for geographic entity types. In *Spatial Information Theory a Theoretical Basis for GIS*; Springer: Berlin/Heidelberg, Germany, 1993; Volume 716, pp. 270–283. [CrossRef]
26. Feng, M.; Zhu, Y.; Wang, J.; Liu, R.; Song, J. Research and Implement of Distributed Multi-standard Geo-metadata Sharing. *Geogr. Geo-Inf. Sci.* **2007**, *23*, 8–13.
27. Guttenberg, A.Z. A Multiple Land Use Classification System. *J. Am. Inst. Plan.* **1959**, *25*, 143–150. [CrossRef]
28. Ji, S.; Wei, Y.; Dai, Q.; Gao, Y. Efcient hypergraph neural network on million-level data. *Sci. Sin. Inform.* **2024**, *54*, 853–871. (In Chinese) [CrossRef]
29. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph Neural Networks. *AAAI* **2019**, *33*, 3558–3565. [CrossRef]
30. Ouvrard, X. Hypergraphs: An introduction and review. *arXiv* **2020**, arXiv:2002.05014. [CrossRef]
31. Bertault, F.; Eades, P. Drawing Hypergraphs in the Subset Standard. In *Graph Drawing*; Springer: Berlin/Heidelberg, Germany, 2001. [CrossRef]
32. Antelmi, A.; Cordasco, G.; Polato, M.; Scarano, V.; Spagnuolo, C.; Yang, D. A Survey on Hypergraph Representation Learning. *ACM Comput. Surv.* **2024**, *56*, 1–38. [CrossRef]
33. Bai, S. Hypergraph convolution and hypergraph attention. *Pattern Recognit.* **2021**, *110*, 107637. [CrossRef]
34. Gao, Y.; Wang, M.; Tao, D.; Ji, R.; Dai, Q. 3-D Object Retrieval and Recognition With Hypergraph Analysis. *IEEE Trans. Image Process.* **2012**, *21*, 4290–4303. [CrossRef]
35. Wang, F.; Pena-Pena, K.; Qian, W.; Arce, G.R. T-HyperGNNs: Hypergraph Neural Networks via Tensor Representations. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *14*, 1–15. [CrossRef]
36. *CJJ/T 85-2017[S/OL]*; Standard for Classification of Urban Green Space. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Beijing, China, 2018.
37. Hua, Y.; Zhou, C. Description Frame of Data Model of Multi-granularity Spatio-temporal Object for Pan-spatial Information System. *J. Geo-Inf. Sci.* **2017**, *19*, 1142–1149. [CrossRef]
38. Tambassi, T. What a Geographical Entity Could Be. In *The Philosophy of GIS*; Tambassi, T., Ed.; Springer International Publishing: New York, NY, USA, 2019; pp. 177–205. [CrossRef]
39. Inkpen, R.; Wilson, G. Entities and classification. In *Science, Philosophy and Physical Geography*, 2nd ed.; Routledge: Oxfordshire, UK, 2013.
40. Hua, Y. The Core Problems and Key Technologies of Pan-SpatialInformation System. *J. Geomaties Sci. Technol.* **2016**, *33*, 331–335. [CrossRef]
41. Li, M. Research on multi-granularity division method based on quotient space geographic entities. In Proceedings of the Fourth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2023), Wuhan, China, 12–14 April 2024; Volume 12978, pp. 696–701. [CrossRef]
42. Song, Y. Research of database inquire performance optimization based on MySQL. *Electron. Des. Eng.* **2021**, *29*, 43–47. [CrossRef]