Representation and Association of Chinese Financial Equity Knowledge Driven by Multilayer Ontology

Zhenghao Liu, Huakui Lv and Xi Zeng

March 31, 2021

Representation and Association of Chinese Financial Equity Knowledge Driven by Multilayer Ontology

LIU Zheng-hao[1,2,3] LV Hua-kui[1,2,3] ZENG Xi[1]
[1](School of Information Management, Wuhan University, Wuhan 430072, China)
[1](Institute of Big Data, Wuhan University, Wuhan 430072, China)
[2](Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China)

Abstract: Aiming at the current situation of complex financial ownership structure and isolated data organization, this study referring to the methods for multi-layer hierarchical construct domain ontology modeling. At the same time, the three dimensions of industry, company and internal environment were integrated, and the concept cube was designed and constructed based on knowledge extraction and text classification technology, so as to provide a multi-level and fine-grained knowledge representation and association method for financial equity knowledge. The experimental results show that conceptual cube structure represents semantic information as a dense low-dimensional representation vector, which greatly enhances semantic relevance and interpretability. The multi-layer ontology-driven ownership structure reflects a variety of knowledge association patterns, and in the "Intelligent Financial Big Data System" developed by the research team, the association query of three categories of association relationships in the field of industry, enterprise and internal environment is realized, as well as the dynamic analysis and supervision of typical financial management problems.
Key words: Multilayer domain ontology; Concept cube; Financial equity; Knowledge association; Knowledge representation; Triple extraction

1 Introduction

As the financial industry is increasingly closely related, the complex relationship between financial institutions, coupled with the complexity of financial knowledge system, undoubtedly increases the complexity and risk of the financial system and brings difficulties to the organization of financial knowledge. As the basis of financial market transactions and corporate governance system, equity structure not only includes medium and micro factors such as enterprise operation and management, but also involves macro-environmental factors such as national policies and regulations and market transactions. If we grasp the financial ownership association, we will grasp the root of the formation and transmission of systemic financial risks. Therefore, the effective organization and association of financial knowledge, especially equity structure, has attracted extensive attention from academia and the industry.

With the arrival of the third wave of artificial intelligence, the concept of ontology is frequently mentioned in the field of artificial intelligence and knowledge engineering. In financial applications, ontology, as the underlying abstraction, can provide data model basis for business system and user model construction, financial knowledge service and financial risk identification. Knowledge representation and organization based on domain ontology ensure the uniqueness of knowledge understanding, while adapting to the diversity of knowledge areas involved and the complexity of semantic relationships. Among them, the multi-layer domain ontology can describe and represent the static and dynamic knowledge systems of different levels from multiple levels, which has significant advantages over the traditional domain ontology knowledge representation model.

Based on specific equity knowledge structure, this study draws on the construction and representation methods of hierarchical domain ontology, and constructs the Concept Cube of financial equity instances by using natural language processing technologies such as information extraction and text classification on the basis of ontology modeling, which provides a new idea for the effective representation and organization of financial knowledge. At the same time, it elaborates the model of equity knowledge association and its application examples. Through the construction of multi-layer ontology model and concept cube, the internal structure and correlation of financial equity can be excavated from the bottom, from multiple angles and at multiple levels, and the root causes of the formation and transmission of systemic financial risks can be identified, bringing new enlightenment for the development of user-oriented financial knowledge services and the prevention and resolution of systemic financial risks.

2 Related Works

2.1 Knowledge Representation and Knowledge Association

Knowledge representation is the modeling of knowledge level, which emphasizes the normative and formal expression of the underlying knowledge subject. And knowledge association is an important link of constructing knowledge network and forming knowledge increment on the basis of knowledge representation. In the early studies, scholars paid attention to the theoretical basis and organizational form of knowledge representation and correlation, and the research focus was different in different fields. In the field of information management, some scholars focus on the study of knowledge itself, that is, meta-knowledge. For example, starting from the perspective of meta-knowledge association, Gao, Ding, Pan and Yuan (2015) found that the association forms among knowledge elements include six types: membership, cross, co-occurrence, citation, co-citation and coupling, and knowledge association is directional. Asghari, Sierra-Sosa, and Elmaghraby (2020) monitored and tracked spatiotemporal thematic information based on a meta-knowledge modeling framework. Other scholars focus on the study of the correlation connotation of knowledge and its carriers, such as Wen et al. (2010) systematically studied the characteristics and types of knowledge association and used knowledge association to classify, represent and organize knowledge and its carriers. In the field of computer science, it focuses on the discussion of the internal structure and form of knowledge correlation. With the development of Semantic Web technology, Ontology and Linked Data have been widely studied as new technical methods and implementation approaches for consistent semantic representation and association. Ontology originates from philosophy. It decomposes things in the objective world, finds out the relationship between their basic internal composition and external concepts, and then studies the abstract nature of objective things. Ontology-based knowledge representation uses RDF language standard as the underlying data model, and is semantically organized through the consistent semantic description method and the unified access mechanism of the associated data. Later, the Linked Open Data Project further promoted the growth and integration of RDF Data, laying a theoretical and practical foundation for large-scale knowledge association and sharing. With the advent of the era of big data, research has gradually turned to the integration, linking and organization of massive multi-source heterogeneous data, and focused on the systematic or innovative related knowledge required by users, in order to provide users with customized services of related knowledge. Li and Liu (2019) believed that knowledge association in big data not only contains the static association state of knowledge information, but also contains the correlation process of knowledge information with increasing density condensed from the big data with sparse value, which has the characteristics of networking, multi-level and multi-faceted.

With the acceleration of economic globalization, the global economy and financial system has become an open-linked system. However, the current situation of fragmented and isolated data organization in the financial field makes it impossible for regulators to rely on the correlation between knowledge, which affects the coordinated development and linkage innovation of the entire financial system and also brings potential systemic risks. Traditional knowledge of financial field is restricted by objective factors, and there are great limitations in the storage, presentation and organization of knowledge. In the study of financial knowledge representation and association, O'Riain, Harth and Curry (2012) believe that the application of knowledge association technology in the financial field can standardize the knowledge representation and organization mode and increase the data interoperability across multiple financial systems and financial instruments. Zhang, Tang, Wang and Pan (2018) emphasized the effective support of linked data integration and semantic association discovery to financial business by comparing traditional financial data with linked financial data and combining with visual cases of financial association knowledge. However, financial knowledge types of diversity and heterogeneity brings new challenge to the existing knowledge modeling, due to the association often located in different angles, different levels, so only through the basic situation of financial market information unified, standardized modeling said to form a specific correlation patterns, can avoid the generation of data island phenomenon.

2.2 Multi-layer ontology

Ontology, as a conceptual modeling tool that can describe information systems at semantic and knowledge levels, is widely used in the representation and association of knowledge in various fields. Ontology can be divided into general ontology and domain ontology according to the target range and application scenarios described by the ontology. General ontology is the standard description of general domain knowledge. In order to highlight the comprehensiveness and level of knowledge, the hierarchical structure is generally adopted for design and description. For example, Li, Ma, Sun and

He (2017) put forward a general model of knowledge representation based on multi-layer domain ontology by integrating ontology knowledge of various domains by referring to the upper semantic knowledge provided by the top-level ontology. Ren, Xu, Liu and Cao (2009) proposed a hierarchical domain ontology model to address the problem that the concept is lack of hierarchy in the process of ontology construction, and the domain knowledge cannot be represented in a systematic, standardized and clear way. The domain ontology was decomposed into modeling metadata layer, top-level ontology layer and sub-domain ontology layer, which made each layer have stronger correlation and expression ability. However, the concepts that constitute the general ontology are extensive and complex. Although they can be applied to a variety of scenarios, there are also problems that the representation of concepts and relations is not comprehensive and profound.

Domain ontology aims to build up the corresponding knowledge specification description for a specific domain and can establish a systematic and scientific knowledge representation model for domain knowledge. It not only defines the basic concepts in the domain, but also covers the relationships among concepts, and provides important terms and theories, examples and interrelated domain activities in the domain. However, due to the lack of scientific design framework and finer-grained research on the existing domain ontology, the ontology models constructed cannot well conform to the five principles proposed by Gruber (1995) (clarity and objectivity; Completeness; Consistency; Maximum monotone scalability; Minimum commitment and minimum code preference), resulting in poor reusability and effectiveness. Based on this, a lot of research on the general level of ontology modeling and representation of the structure, such as Zhu and Liu (2018) carried out ontology modeling for the field of emergency events. Although the concept of hierarchy was not explicitly mentioned, they divided the model into upper event class, lower event class and event instance, which can be considered to have carried out fine-grained construction according to different levels; Zhang, Zhao and Feng (2019) believe that ontology construction in the financial field can be divided into two types of concepts with conceptual hierarchy relationship. One is the change of economic indicators reflecting the state and activities of economic entities, and the other is the economic entities themselves. These two kinds of concepts are in different dimensions, there is a certain relationship between them, so we can construct the relationship between these two kinds of concepts and the mapping relationship between them.

Due to the different characteristics of different domain concepts, especially when the domain knowledge is too complex, the traditional ontology model is not suitable for the domain requirements. However, multi-layer ontology, based on the concept of each level of ontology, faces the knowledge system of different levels, and follows the top-down ontology construction principle, can describe the knowledge system of different levels in multiple dimensions and facets, so as to enhance the semantic correlation between concepts. In essence, the association of financial equity is the link of complex underlying transaction information, so the purpose of this study is to effectively model and represent the associated transaction information based on the multi-layer domain ontology structure model, and to apply it to the practice of knowledge service and risk identification to generate more value.

3. Multilayer Domain Ontology Modeling Representation

This study from academia and industry in general ontology classification thought, and on the basis of existing research, further separation and refinement of domain ontology, constructed by top ontology, the upper ontology, static son domain ontology, dynamic domain ontology and application ontology constitute multilayer financial stake in the ontology model, and in accordance with the ontology semantic function will be further ontology construction concept from abstract to concrete, in turn, divided into base layer, layer and instance, model architecture is shown in figure 1.

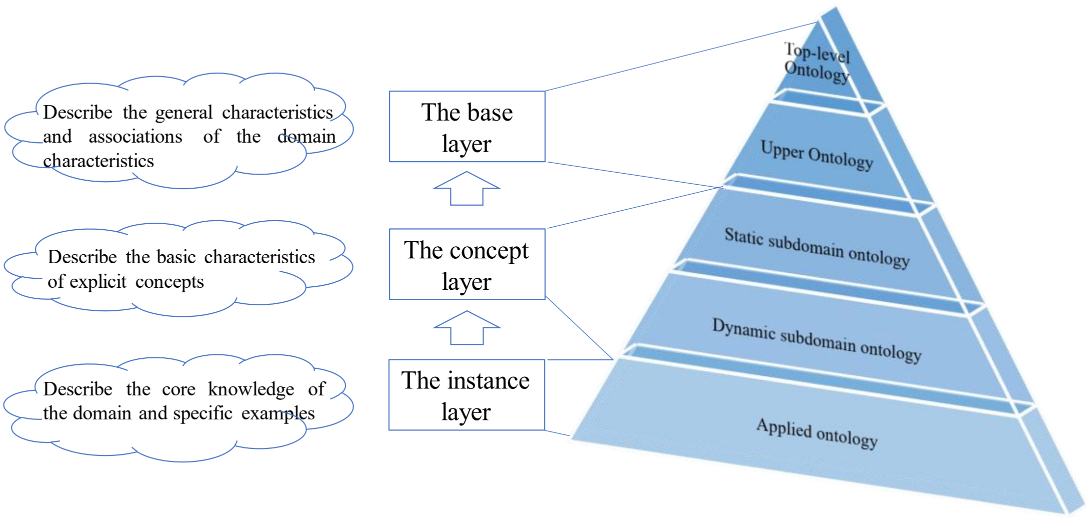


Figure 1 Multilayer Domain Ontology Architecture

The base layer, composed of the top-level ontology and the upper-layer ontology, provides a general connection between the characteristics of the financial domain and reveals the relationship between the domain knowledge at a higher semantic level. At the same time, the core class of the domain is described in a modularized way, and the subclass system is extended according to the granularity of the entity, which provides the low-level abstraction for the conceptual layer. The concept layer is composed of static subdomain ontology and dynamic subdomain ontology. As an intermediate level connecting abstract concepts and application cases, the concept layer can describe the basic characteristics of the securities market and describe the dynamic characteristics of the securities trading market. The instance layer, as the bottom layer, realizes the integration representation of specific instances in the domain by matching and mapping with the base layer and the concept layer.

3.1 The base layer

3.1.1 Top-level Ontology

Top-level ontology is a general description framework of domain Ontology, and specific terms cannot reflect the domain scope of knowledge concepts. Although domain ontology construction is only aimed at a certain domain, it has intricated cross-connection with other related fields. Due to the limitation of knowledge background, manpower and time of developers, it is unrealistic to construct each related domain ontology. However, the top-level ontology in this paper emphasizes the research on common concepts and the relations between concepts, such as space, time and event, within the limited domain, which not only has domain characteristics, but also can carry out knowledge sharing and reconstruction in a larger scope.

Top-level ontology can be abstractly represented by the quaternion $O = (C, P, R, X)$, where C represents the concept set in the ontology, P represents the set of concept attributes, R represents the set of relations among concepts, and X represents the set of ontology axioms and rules.

In financial field, the basic concept set of the top-level ontology can be expressed as C = {financial subject, financial contract, events, institutions, methods, time, spaces}. *Financial subject* refers to the individual participating in financial activities, such as shareholders, legal persons, creditors, etc. *Financial contract* is the basis for the implementation of financial activities, such as contract documents and oral contracts. *Events* specifically refer to the activities that financial entities or institutions participate in financial activities, such as corporate bankruptcy, foreign investment, etc. *Institutions* emphasize the group of participants, such as government agencies, various financial intermediaries, etc. Take the event as an example, P_{event} = {time, place, agent, recipient, event type}. The top-level ontology axioms and rules represent the facts existing in the domain body and can constrain the classes or relations in the body, such as institutions and events, which are within the scope of financial concepts.

3.1.2 Upper Ontology

In essence, Upper ontology is the same as top-level ontology, which provides a universal framework to describe the domain characteristics. However, the upper ontology inherits the reasonable logical structure of the top-level ontology on the basis of the upper framework system, which based on the normative relational definition and the axiomatic definition. The core purpose of this study is to establish a domain ontology rather than a general ontology, so the concept here refers to the object in a specific financial domain that can be abstractly described under the general component module provided by the top ontology.

In order to refine the granularity of ontology construction, a modular dynamic ontology construction method is adopted to complete the ontology reorganization and reuse based on the basic idea of ontology engineering (Kotis, Vouros, & Spiliotopoulos, 2020), so as to improve the construction efficiency and evolution ability of general ontology. On the basis of top-level ontology, the event upper ontology is taken as an example to form the Upper Ontology Module as shown in Figure 2.

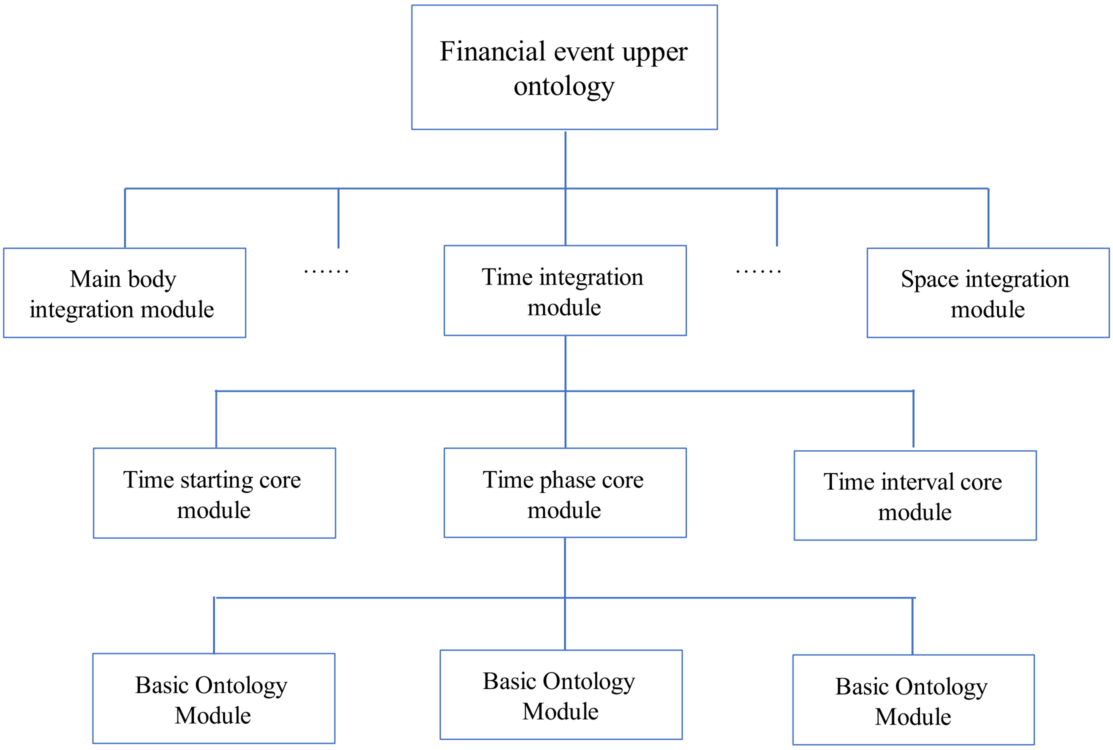


Figure 2 Financial event modularized upper ontology

The idea of ontology modularization comes from software modularization. Just like the modularization mechanism in the field of software engineering (Alobaid, Garijo, Poveda-Villalón, Santana-Perez, Fernández-Izquierdo, & Corcho, 2019), the modularized upper ontology has more semantic integrity and stability, and the concepts within the module are semantically related to each other, but the dependence on information outside the module is low. Compared with the top-level ontology, each general concept can be decomposed into modules, and the dynamic characteristics of the financial field can be described by adding time, space and other modules, such as the abstract description of the state and its transformation. It should be noted that the upper ontology is not the direct inheritance of the top ontology, but the integration of heterogeneous data and the gradual instantiation of modules should be carried out according to the basic concept set and relationship defined by the top ontology constructed.

3.2 The concept layer

As the intermediate layer of multi-layer ontology, the concept layer is the core of ontology construction. In the actual construction process, in addition to the extraction and standard representation of a large number of entities, considering that mature financial domain ontology has been built in foreign countries and ontology described in different languages is basically the same in terms of basic concept definition, in order to improve the efficiency of ontology construction, the FIBO ontology is reused in the research (Petrova, Tuzovsky, & Aksenova, 2017). FIBO ontology mainly focuses on the classification hierarchy relationship in the financial field, which can be extracted and reused for the construction of financial static ontology (Patel, Dube, Tao, & Ning, 2015). Through the construction of the concept ontology, the information resources distributed in the financial field can be organized into an organic whole.

3.2.1 Static subdomain ontology

According to the modular framework of the top-level ontology and upper ontology, the static ontology mainly describes the explicit concept of the basic characteristics of the domain. The construction of concept system is the key work of static ontology construction, which mainly includes six parts: (1) Extraction and normalized representation of multi-source heterogeneous data; (2) Concept selection and specification based on abstract concepts of top-level ontology and upper ontology; (3) Define classes and their hierarchical relations; (4) Further clarifying the terms; (5) Create static subdomain ontology based on OWL ontology language; (6) Ontology evaluation and reconstruction. Figure 3 shows the conceptual ontology construction process and its connotation.

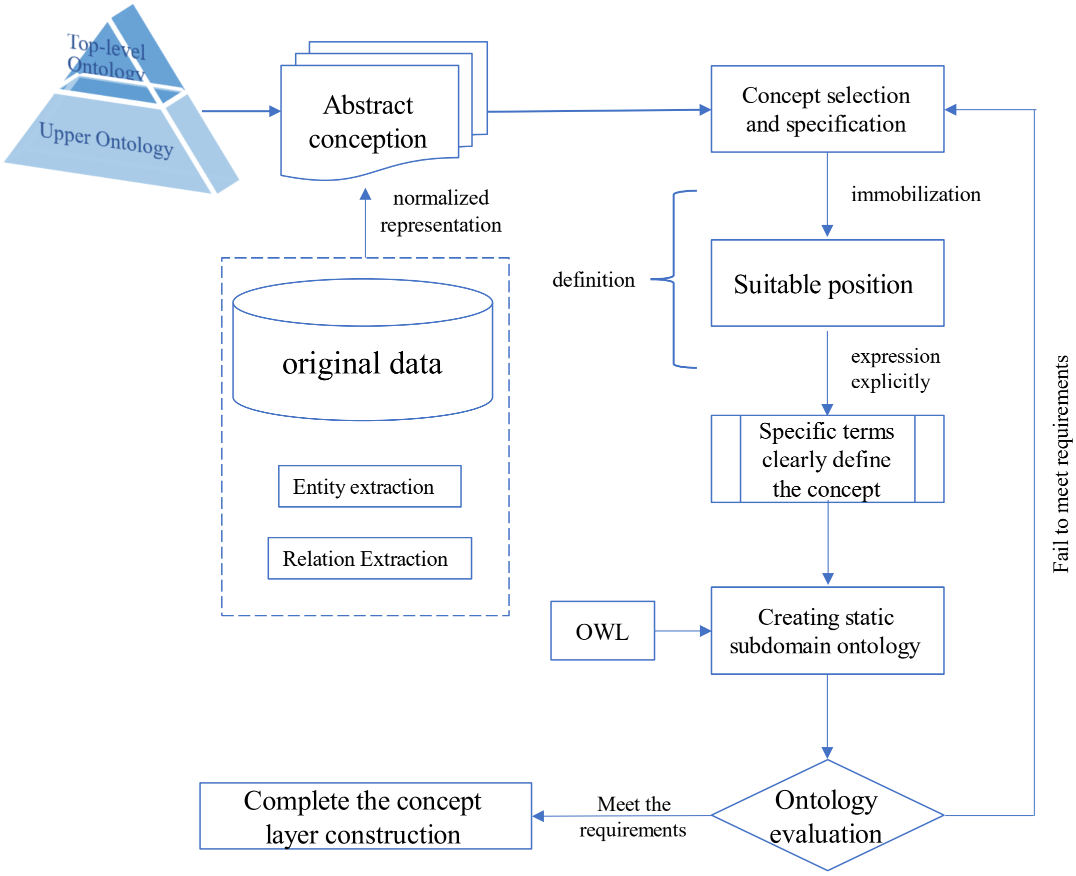


Figure 3 Construction process and connotation of static subdomain ontology

The core of constructing static subdomain ontology is to complete the representation of static knowledge, and the research adopts OWL language to complete the modeling of general concepts. In the process of knowledge representation of basic elements, information resources and knowledge resources need to be abstracted object-oriented in order to extract concepts and their relations. Secondly, the corresponding classes (including concepts, attributes, relationships, etc.) need to be constructed according to the syntax requirements of OWL and stored in the OWL type declaration document.

(1) Define the class

OWL defines and describes classes through association constraints. For example, *< owl:Class rdf :ID= "Financial Contract" > </owl:Class>* can stand for the general concept of "financial contract". The following indicates that "credit risk" is a subclass of "risk event", while "risk event" is a subclass of "event". In addition, the hierarchical relationship of OWL can be simply inferred, so that "credit risk" is also a subclass of "event".

```
<owl:Class rdf:ID="信用风险">
        <rdfs:sub Class Of>
                <owl:Class rdf:resource="#风险事件"/>
        </rdfs:sub Class Of>
</owl:Class>
<owl:Class rdf:ID="风险事件">
        <rdfs:sub Class Of>
                <owl:Class rdf:resource="#事件"/>
        </rdfs:sub Class Of>
</owl:Class>
```

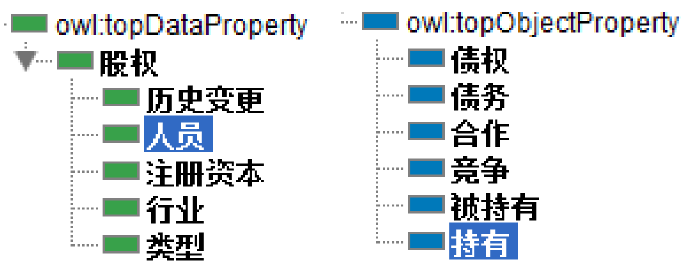
(2) Define attributes

In ontology language, semantic relationships are defined by attributes. OWL divides attributes into object attributes and data attributes. Object attributes are used to describe the characteristics and relationships of classes in the domain, while numerical attributes are used to associate class and data type values by describing class parameters.

```
<owl:ObjectProperty rdf:ID="参股">
      <rdf:type rdf:resource="&owl;TransitiveProperty"/>       # 定义传递性
     <rdf:type rdf:resoure="&owl;SymmetricProperty"/> # 定义对称性
     <rdf:domain rdf:resource="#股东"/>
   <rdf:range rdf:resource="#公司"/>
</owl:ObjectProperty>
```

On the concept layer, the expressive power of OWL was also supported by many concept constructors and axioms. In addition, it could form the hierarchy of concepts by "subClassOf" and "subPropertyOf", and describe the constraint relations between concepts by "domain", "range", "equivalentProperty", "hasValue", etc. We can equivalently form semantic association through "equivalentClass," "sameAs," and "inverseOf". Using "IntersectionOf", "UnionOf" and other concepts to form a logical combination; and axiomatic definition of the concept and its relationship is realized through "uniqueProperty", "transitiveProperty", etc. (Gruber, 1993). In addition, in order to strengthen the flexibility of semantic expression, the relationship between different entity concepts is artificially defined to achieve the preliminary correlation between concepts. For example, in the enterprise-enterprise relationship, $R_{enterprise}$ = {inter-bank capital exchange, holding, cooperation, competition}. The above definitions provide a normative representation framework for the construction of the instance layer, facilitate the normative representation and expansion of related concepts and entities, and thus guide the design and implementation of the concept layer and instance layer from a higher semantic level.

Financial ownership structure itself is relatively simple and involves less conceptual knowledge, but the entities are closely connected. The conceptual attributes of equity mainly include types, industries, registered capital, personnel, historical changes, etc. The relations in the equity structure mainly include the relationship between the shareholding and the shareheld, the competition and cooperation between companies, and the bond and debt relationship between shareholders, etc., as shown in Figure 4. The equity static ontology formed based on hierarchical association and relationship fusion is shown in Figure 5.



（a）Financial equity concept data attributes (b) Financial equity relation constraint

Figure 4 Data attributes and object attributes of financial equity ontology (Chinese)

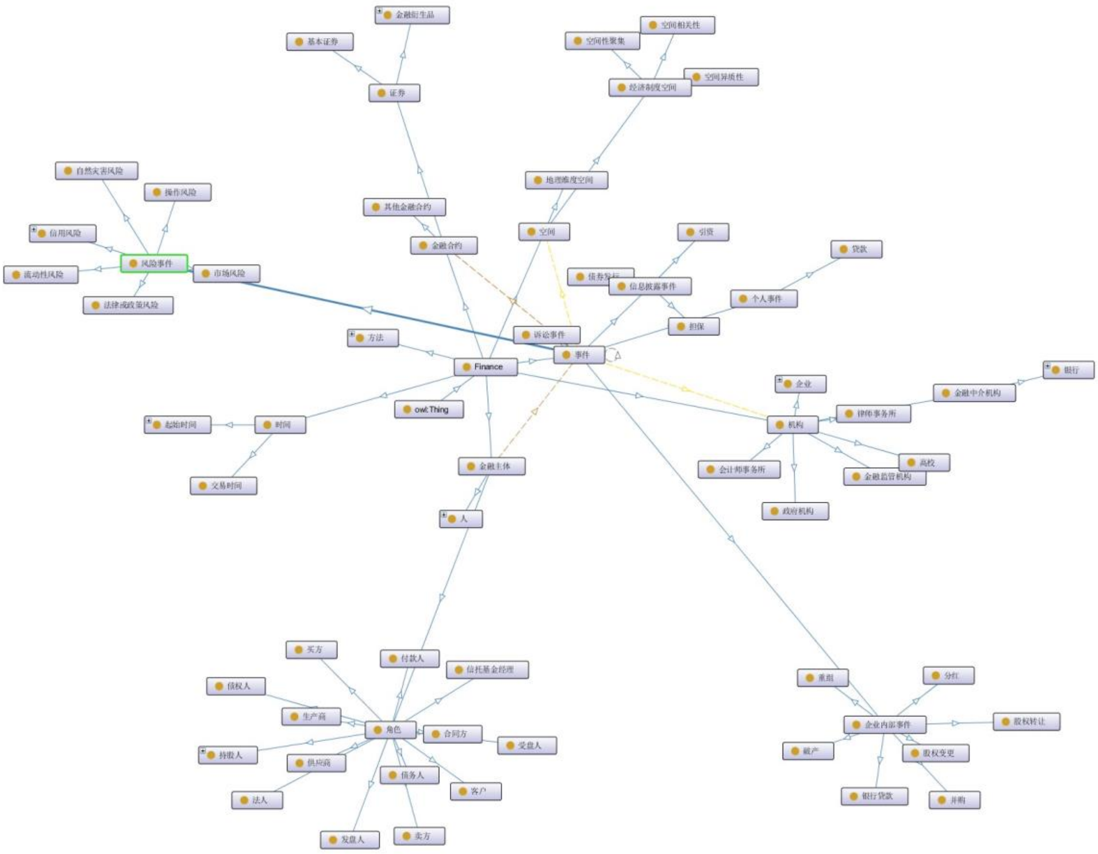


Figure 5 Equity static subdomain ontology

3.2.2 Dynamic subdomain ontology

The dynamic subdomain ontology can normalize the knowledge with dynamic characteristics in the domain. The main work is to classify and organize the concepts of the core domain and make clear attribute association through clear concept classification hierarchy. This section takes the core entity of equity change event as an example to construct the dynamic subdomain ontology as shown in Figure 6.

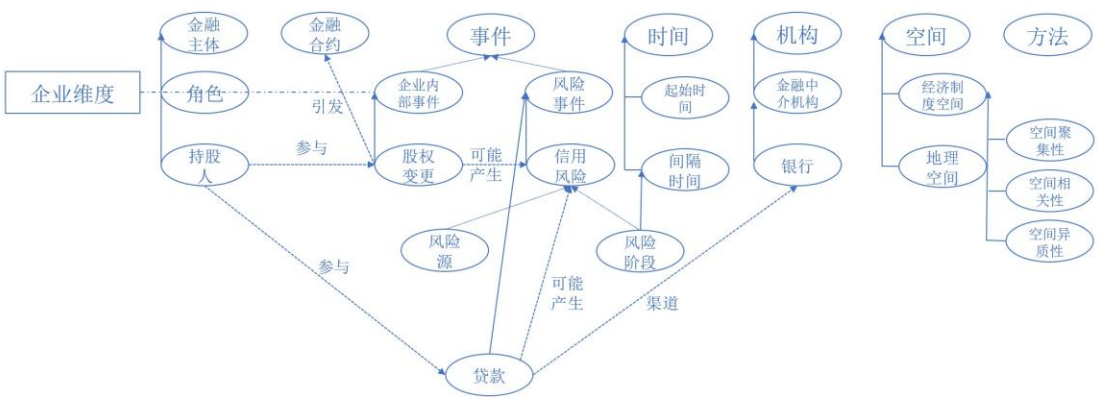


Figure 6 Dynamic subdomain ontology of ownership change

This ontology inherits the event module, time module and space module of the basic layer, and extracts the corresponding concepts and hierarchical relations from the static subdomain ontology. It mainly contains the description of the time dimension and space dimension of the equity change event, involving both static knowledge and dynamic knowledge. In the actual construction, basic concepts and relationships are obtained by analyzing the OWL of the static ontology model. For concept relations, for example, "cooperation-competition" establishes the antisense relation through inverse-of, and "mutual possession" establishes the symmetric relation through symmetric. Domain and range constrain the subject-object attributes. For example, to "participating" relationship, domain attributes are added as financial subjects and institutions, while ranges attributes constrain as events.

3.3 The instance layer

As the lowest level ontology, the application ontology can be matched and mapped with the upper level to realize the instance integration representation within the domain. The module of upper ontology integration is referenced and inherited by using ontology, and the static subdomain ontology and dynamic subdomain ontology are reified.

Hierarchical construction method reflects the categorical relevance of knowledge representation model. However, Shams and Barforoush (2004) found in their survey that most studies still focus on hierarchical relation of ontology construction and are often unable to extract and represent cross-hierarchical entities and non-hierarchical relations. For example, the risk object of "credit risk" in different contexts may be enterprises or individuals, or even the whole industrial chain. Another example is that the subject of the "cooperative" relationship may involve financial subjects and financial institutions at different levels. Thus it can be seen that the polysemy of the concept itself determines that it can represent different meanings in different dimensions, rather than being classified into a single dimension or category. On the other hand, Tang, Ma, Fu, & Zhang (2019) pointed out in their research that in addition to classification correlation, financial knowledge also has typical characteristics of time-series correlation, statistical correlation and event correlation, and the important characteristics of relational knowledge are not only hierarchical continuity, but also high integration and multifaceted.

Therefore, it is necessary to construct a cross-level knowledge instance representation structure that can integrate multi-dimensional information to further enrich the diversity and flexibility of semantic representation. The following article focuses on the knowledge instance representation and storage model of the "conceptual cube" structure. The conceptual cube can represent the entity as a space vector based on probability, so as to enhance the ability of semantic association.

3.3.1 Concept cube

In the field of database, data cube is a core data model in the field of data warehouse and online analysis and processing. it can be multi-dimensional data characterizing features, many of the existing studies draws this idea. Li, Chen, & Andrienko (2018) built Semantics-Space-Time Cube, which explores the interrelationships among the three heterogeneous information aspects of semantics, space and time, and derives the law of changes in text semantics with time and space. Escobar, Candela, Trujillo, Marco-Such, & Peral (2020) used the multidimensional model method based on the RDF data cube vocabulary to add values to the open link data and completed the analysis of the multidimensional characteristics of the data; Shi (2010) used FCA theory to study high-performance data cubes and their semantics based on formal concepts and conceptual levels. It can be seen that the essence of the data cube lies in the multi-dimensional, multi-faceted feature representation.

Ontology is one of the manifestations of the knowledge base, and the fusion of multi-dimensional information can show the hidden features of the ontology knowledge from different sides. Therefore, the structure of the data cube can be used to further enrich the multi-dimensionality and flexibility of its semantic expression. Relying on the concept of data cube, this study defines an ontology model that can multi-dimensionally represent associated knowledge instances as "Concept Cube". The specific definition is as follows:

Definition1 *Concept Cube refers to the multi-dimensional knowledge representation and storage space constructed from dimensions. It is a fact-based ontology established to meet the needs of users for knowledge query and analysis from* *multiple angles and multiple levels. The instance model contains all domain knowledge instances and relationships to be retrieved and analyzed, and all related knowledge operations are performed on the cube.*

Table 1 provides a detailed comparison and introduction to the basic concepts, storage objects, main problems and typical application scenarios involved in the data cube and the concept cube.

Table 1 *Concept comparison between data cube and concept cube*

		Data Cube	Concept Cube
Basic Concept	**Dimension**	A perspective for observing data, such as representing business data from three perspectives: time, region, and product	A perspective to describe domain knowledge, such as describing securities domain knowledge from three aspects: industry, enterprise, and internal environment
	Dimension Members	The coordinate value of the dimension axis, such as the dimension member of the time dimension in the first quarter	The basic conceptual attributes corresponding to each dimension axis, such as industry entities, corporate financial indicators, internal risks, etc.
	Measure	A point determined by a limited combination of dimensional members. In theory, only a limited number of data points can be represented by a cube	A point determined by a combination of finite dimensional members and an infinite semantic feature (probabilistic representation). In theory, there are infinitely many points used to represent domain knowledge instances
Storage object		Multidimensional data	Linked knowledge
Main problem solved		To solve the multidimensionality of data and enrich the integration of data	To solve the ambiguity of knowledge and enrich the relevance of knowledge
Application scenarios		Complex data query	Related knowledge retrieval, path penetrating query, etc.

3.3.2 Model design

To equity structure, which involves the industry/business/trading between internal, management, supervision, competition and cooperation, etc, thus can be divided into three dimensions: industry, enterprise, and internal environment, forming an ontology cube structure. The three have built a domain knowledge framework from different scopes and directions, and they themselves are included in the ontology as the relationship between category and category (industry-enterprise relationship, industry-internal environment relationship, and enterprise-internal environment relationship).

(1) Industry: The “industry” dimension or “market” dimension describes the entities, attributes and relationships of financial securities at a macro level. The main attributes of the financial securities industry/market include name, industry operating status, industry policy, industry capabilities (market capacity, output value, and number of companies in the industry), industry financial indicators, and industry life cycle (start-up period, growth period, mature period) And recession period) and industry systemic risks.

(2) Company: The "Company" dimension describes domain knowledge from a meso level. Its main attributes include the name and quantity of the company or organization, the ownership structure, management structure, trade alliance structure of the company or organization's governance structure, enterprise/organization competition and cooperation, corporate financial indicators, the company's life cycle, and external corporate risks. Among them, corporate financial indicators are a relatively broad concept with obvious numerical attributes. Financial indicators and their corresponding financial entities are usually used to reflect the status, changes and relationships of financial entities, and their attributes include update frequency, time, and data sources.

(3) Inner Environment: The "inner environment" dimension represents knowledge from a micro level. Its main attributes include company product structure, company personnel organizational structure, product financial indicators (including growth stage, capacity, sales, price, etc.), company internal culture (company values, company strategy, company philosophy, etc.), and internal company risks.

The base layer of domain ontology defines the basic elements such as concepts and relationships, and provides a general knowledge framework. The concept layer represents the core classes and attributes from the static and dynamic aspects respectively, inherits the entities and relationships in the base layer and extends the subclasses according to the granularity of entities. The instance layer is defined as a cube structure composed of three dimensions of industry, enterprise and internal environment, in which the instance set mapped by concept can form a specific sub-cube, and each sub-cube stores the entities and relationships with probability values as spatial coordinates. Figure 7 vividly shows the conceptual cube structure at the instance level and the hierarchical association and hierarchical mapping of the multi-layer ontology.

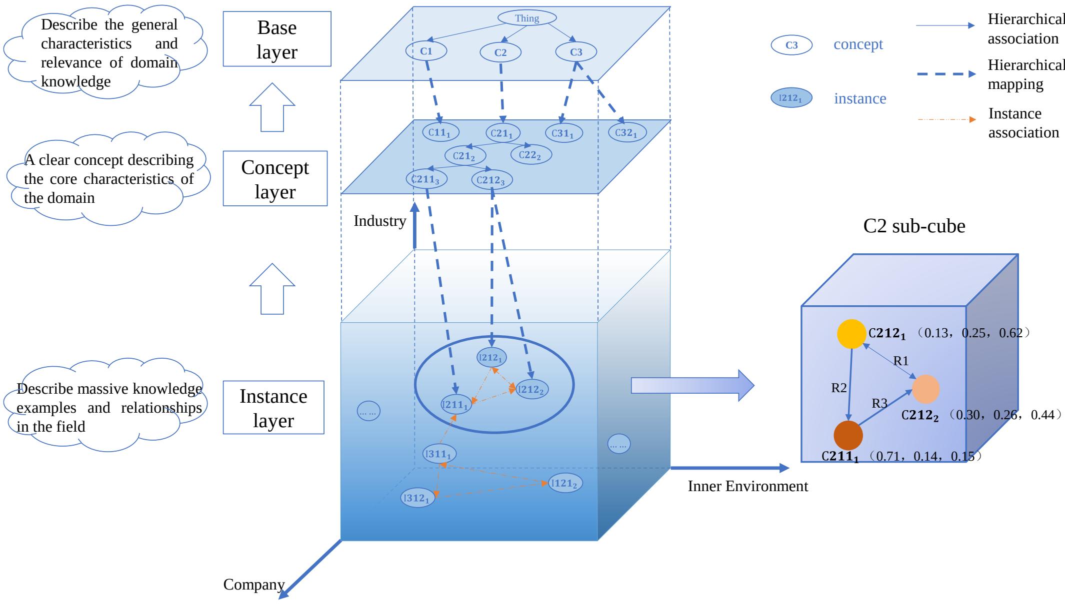


Figure 7 Schematic diagram of the multi-layer domain ontology cube model structure

4 Equity knowledge association based on multi-layer ontology

4.1 Realization of Equity Concept Cube

The equity knowledge representation model based on multi-layer ontology formalizes the association structure from the conceptual level. The instance layer needs to rely on specific data, starting from the knowledge elements of the basic and conceptual layers, and mapping the knowledge elements to the underlying *Entity-Relation-Property* structure. Driven by the multi-layer ontology, combined with the characteristics of the equity structure, select appropriate dimensions to represent the semantic information of the research object as dense low-dimensional representation vectors, which can efficiently calculate the semantic connections between entities and relationships in a low-dimensional space, and enable knowledge acquisition, The performance of fusion and inference has been significantly improved. In this section, on the basis of multi-layer ontology modeling, actual data is used to realize the concept cube structure of equity that can realize semantic association in three-dimensional space.

4.1.1 Data processing and triple extraction

The data sources mainly include listed company transaction data provided by Shenzhen Securities Information Co., Ltd. and automatically obtained public text information resources, such as securities industry reports, corporate research reports and announcements, and financial news short messages. For announcements and public opinion data, we form an initial corpus through preprocessing such as word segmentation and stop-word removal. Generally speaking, the instance layer can be automatically constructed with the help of natural language processing technologies such as knowledge extraction. This article extracts open domain entity relationships based on dependency syntax rules, and uses LTP natural language processing tools to achieve effective extraction of multi-source text data. Set the extraction rules such as Table 2 shown provide technical support for expanding ontology instances and their relationships. Dependency Parsing can judge the grammatical dependence of each component in a sentence based on the part of speech and the positional relationship between words. Therefore, the entity relationship extraction based on the dependency syntax mainly depends on the predicate in the sentence. When in the relationship represented by the predicate When it contains arguments, semantic triples can be extracted.

Table 2 *Semantic triples extraction rules of instance-level text*

number	**Dependency syntactic relation**	**Extraction rules**	**Annotation**	**Sample extraction**
1	Subject-verb-object relationship	postags[index] == 'v' 'SBV' and 'VOB' in dict (e_1, r, e_2)	Predicate as the core to identify subject-predicate and verb-object relations	(投资者，认购，公司发行股票)
2	Verb-object relationship, attributive postposition	postags[index] == 'v' relation == 'ATT' 'VOB' in dict (e_1, r, e_2)	Predicate as the core to identify the attribute relations	(股票价格，高于，面值)
3	Subject-predicate-verb-complement relationship	postags[index] == 'v' 'SBV' and 'CMP' in dict (e_1, r) 'POB' in dict (e_2)	Predicate as the core, first identify the head entity and relationship through the subject-predicate verb-complement relationship, and then identify the tail entity through the preposition-object relationship	(公司，致力于，软件基础平台核心技术研发能力)

For the extracted results, we manually eliminated the ideographically ambiguous entities and non-ideographic relationships, and finally got 32,627 entities and their associated 1,928 semantic relationships. Part of the extraction results are shown in Figure 8.

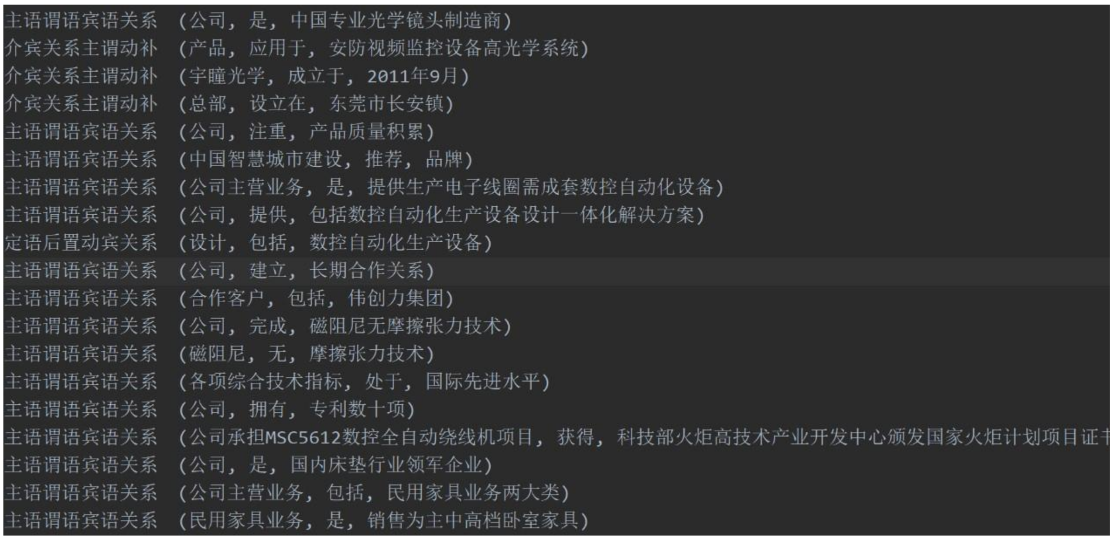

主语谓语宾语关系 （公司，是，中国专业光学镜头制造商）
介宾关系主谓动补 （产品，应用于，安防视频监控设备高光学系统）
介宾关系主谓动补 （宇瞳光学，成立于，2011年9月）
介宾关系主谓动补 （总部，设立在，东莞市长安镇）
主语谓语宾语关系 （公司，注重，产品质量积累）
主语谓语宾语关系 （中国智慧城市建设，推荐，品牌）
主语谓语宾语关系 （公司主营业务，是，提供生产电子线圈需成套数控自动化设备）
主语谓语宾语关系 （公司，提供，包括数控自动化生产设备设计一体化解决方案）
定语后置动宾关系 （设计，包括，数控自动化生产设备）
主语谓语宾语关系 （公司，建立，长期合作关系）
主语谓语宾语关系 （合作客户，包括，伟创力集团）
主语谓语宾语关系 （公司，完成，磁阻尼无摩擦张力技术）
主语谓语宾语关系 （磁阻尼，无，摩擦张力技术）
主语谓语宾语关系 （各项综合技术指标，处于，国际先进水平）
主语谓语宾语关系 （公司，拥有，专利数十项）
主语谓语宾语关系 （公司承担MSC5612数控全自动绕线机项目，获得，科技部火炬高技术产业开发中心颁发国家火炬计划项目证
主语谓语宾语关系 （公司，是，国内床垫行业领军企业）
主语谓语宾语关系 （公司主营业务，包括，民用家具业务两大类）
主语谓语宾语关系 （民用家具业务，是，销售为主中高档卧室家具）

Figure8 Partial triples extraction results (Chinese)

4.1.2 Text Classification and Spatial Vector Representation

The triples based on dependency syntax retain contextual and semantic information well and can be used as the original corpus for classification. Therefore, the extracted semantic triples are classified as a whole, and the classification results of head and tail entities are represented by probability. For the same entity, the average value is taken as the final spatial vector value. The results are further divided into different sub-cubes to complete the final positioning of the entity.

In the experiment, we divided 8,000 randomly selected triples into three categories: industry (market), enterprise and internal environment according to the dimensions in Section 3.3, and then divided the data into training set and test set according to 8:2. In the experiment, Keras+Finbert deep learning framework was used to complete the classification task. BANCH SIZE =16, EPOCHS =5, and chose ADAM optimizer. The experimental results are shown in Table 3.

Table 3 *Evaluation results of dimension classification*

	Precision	***recall***	***F1-score***
Industry/market dimension	0.6518	**0.9614**	0.7769
Company dimension	0.9348	0.9577	**0.9416**
Inner environment dimension	**0.9797**	0.6636	0.7913
Macro avg	0.8554	0.8609	0.8381
Weighted avg	0.8972	0.8684	0.8681

The test results show the overall effect of the dimensional classification of triples is in line with expectations. In the final prediction task, the classification probability expressed by softmax activation function is directly used as the coordinate value of the head entity and tail entity of each triplet. For example, the probability that the triple (Bohai Property Insurance, signed, strategic cooperation agreement) is divided into "industry/enterprise/internal environment" is 0.2365/0.7611/0.0024, then the head entity "Bohai Property Insurance" and the tail entity "strategic cooperation". The relative coordinates of the “agreement” are (0.2365, 0.7611, 0.0024), but they are stored in different sub-cubes because of belonging to different conceptual categories; another example is (Kweichow Moutai, belonging to, the liquor industry). The overall output probability of the tuple is 0.0533/0.9446/0.0021, but the final relative coordinates of the two are represented in the form of an average value since head and tail entities appear multiple times in the corpus. Table 4 respectively presents the partial prediction examples and the space vector representation results based on probability when the head and tail entities in the corpus are unique (a) and non-unique (b).

Table 4(a) *Prediction examples and space vector representations with unique head and tail entities*

tuple	**head entities**	**category**	**Vector representation**	**Tail entities**	**category**	**Vector representation**
渤海财险，签订，战略合作协议	**渤海财险**	**公司**	**(0.2365, 0.7611, 0.0024)**	**战略合作协议**	**合约**	**(0.2365，0.7611 ，0.0024)**
北京国联视讯信息技术股份有限公司，简称，国联股份	北京国联视讯信息技术股份有限公司	公司	(0.0690, 0.9309, 0.0001)	国联股份	公司	(0.0690, 0.9309, 0.0001)

LS 系列发动机废气循环 EGR 系统荣获中国机械工业科学技术奖二等奖	LS 系列发动机废气循环 EGR 系统	业务/产品	(0.1392, 0.0085, 0.8522)	中国机械工业科学技术奖二等奖	企业成就	(0.1392, 0.0085, 0.8522)
钢化产品，销往，武钢股份	钢化产品	业务/产品	(0.2222, 0.0079, 0.7699)	武钢股份	公司	(0.2222, 0.0079, 0.7699)
运营总部，位于，深圳	运营总部	机构	(0.0005, 0.0007, 0.9988)	深圳	地理位置	(0.0005, 0.0007, 0.9988)

Table 4(b) *Prediction examples and spatial vector representations of non-unique head and tail entities*

tuple	**head entities**	**category**	**Vector representation**	**Tail entities**	**category**	**Vector representation**
中国人寿，持有，贵州茅台	中国人寿	公司	(0.6140, 0.3858, 0.0002)	贵州茅台	公司	*(0.1832, 0.5698, 0.2470)*
贵州茅台，秉承，诚信核心价值理念	贵州茅台	公司	*(0.1832, 0.5698, 0.2470)*	诚信核心价值理念	企业文化	(0.0105, 0.0050, 0.9844)
贵州茅台，收到，上交所监管函	贵州茅台	公司	*(0.1832, 0.5698, 0.2470)*	上交所监管函	机构	(0.0551, 0.9432, 0.0017)
白酒行业，拥有，汾三大中国驰名商标企业	白酒行业	主营行业	*(0.5264, 0.4726, 0.0010)*	汾三大中国驰名商标企业	机构	(0.9994, 0.0005, 0.0001)
贵州茅台，属于，白酒行业	**贵州茅台**	**公司**	***(0.1832, 0.5698, 0.2470)***	**白酒行业**	**主营行业**	***(0.5264, 0.4726, 0.0010)***

Three-layer ontology maps all concepts and entities one by one and fully associates them through the semantic data mapping model. In the construction process, the top-down construction model maps the knowledge elements to the underlying entities, relationships and attributes, which greatly enhances the stability and scalability of the ontology structure; and in the application process, bottom-up induction And integration can gradually refine the missing financial securities knowledge concepts and relationship patterns, and can further use semantic data mapping to supplement the upper-level ontology. In addition, connecting the relationships between different levels and different subcubes is especially important for complete and multi-dimensional conceptual knowledge description. Based on the examples in Table 4 and the space vector representation, Figure 9 visually shows the mapping, association and fusion results of the multi-layer ontology cube.

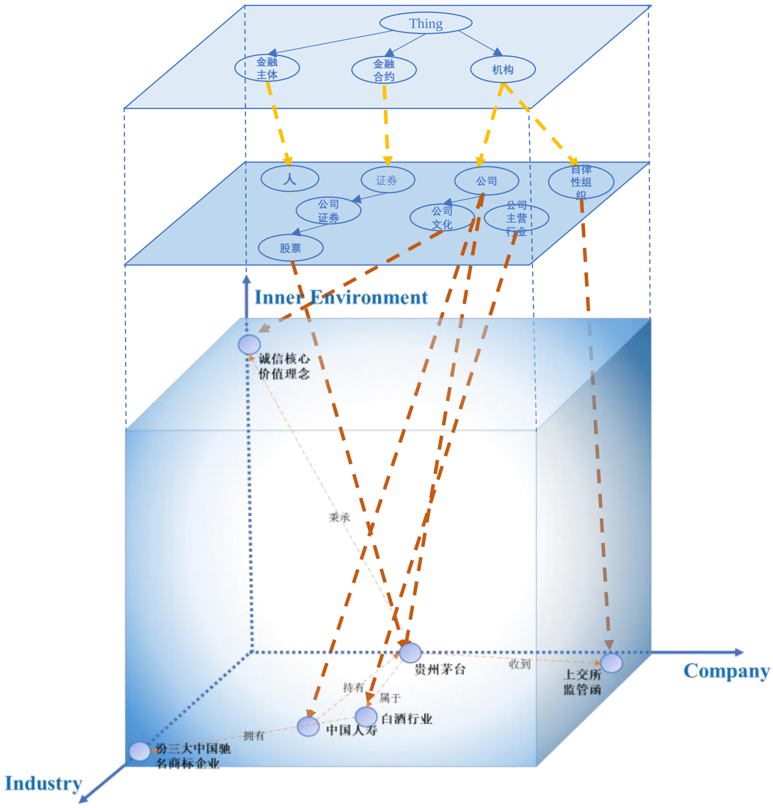


Figure9 Mapping, association and fusion of multi-layer ontology cubes

4.2 Analysis and Application of Ontology-based Equity Association Model

Equity is a comprehensive right of the personal and property rights enjoyed by the shareholders of a limited liability company or joint stock limited company. Equity data can fully reflect the relationship between organizations and reflect the transfer of rights through indirect shareholding behavior. Grasping the linkage of financial equity will grasp the source of the formation and transmission of systemic financial risks.

The financial equity structure itself is relatively simple, and the conceptual entities involved only include listed companies, various financial institutions, and individual roles such as shareholders, directors, supervisors, and senior executives. However, the connections between entities are close and complex. The static sub-domain ontology uses FIBO ontology to normatively describe the attributes and relationships of related concepts, and realizes the classification and association of domain knowledge. For a listed company, its attributes mainly include type, industry, registered capital, main products and businesses, personnel, historical change announcement information, etc. The relationship in the equity structure mainly includes the relationship between holding and held shares, and between companies The relationship between competition and cooperation, and the relationship between creditor's rights and debts among shareholders. By analyzing the semantic association between concepts and attributes in the domain, an RDF graph as shown in Figure 10 is constructed.

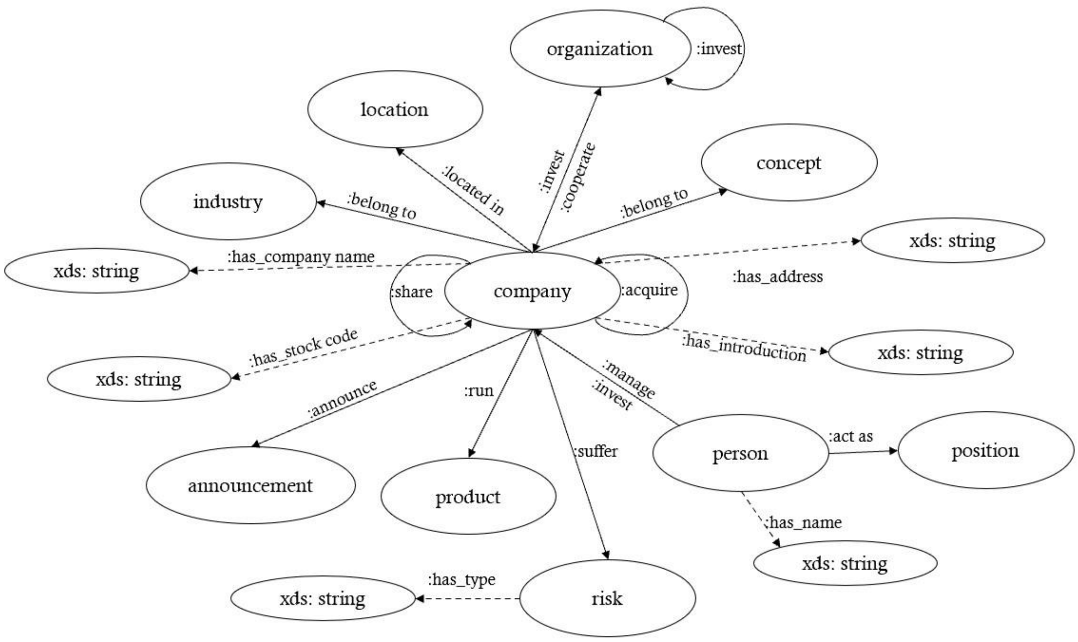


Figure 10 Ontology RDF Diagram of Static Subdomain of Financial Equity

On the other hand, according to the financial knowledge association model proposed by Li and Fan (2019), it can be considered that financial equity knowledge association is a dynamic evolution process driven by equity-related events. The dynamic sub-domain ontology inherits the top-level concepts such as event, time and space of the top-level ontology, and the event module and time module in the upper-level ontology, which clearly portray the dynamic characteristics of the trading market, and can deal with some vicious behaviors and financial risks that lead to potential risks. A reasonable description of the contagion phenomenon can reflect the event relevance and time relevance of the equity knowledge structure. Figure 11 shows four related behaviors (single shareholding, cross shareholding, related transaction and actual shareholding). Obviously, traditional knowledge representation methods cannot reveal the implicit relationship between entities. Only RDF triples form a network structure. And follow the evolution of the event to analyze the investment and financing behavior from the currency market to the capital market, and the behavior of shareholders to increase or decrease their holdings, so as to truly discover the flow characteristics and potential risks of the internal and external trading markets.

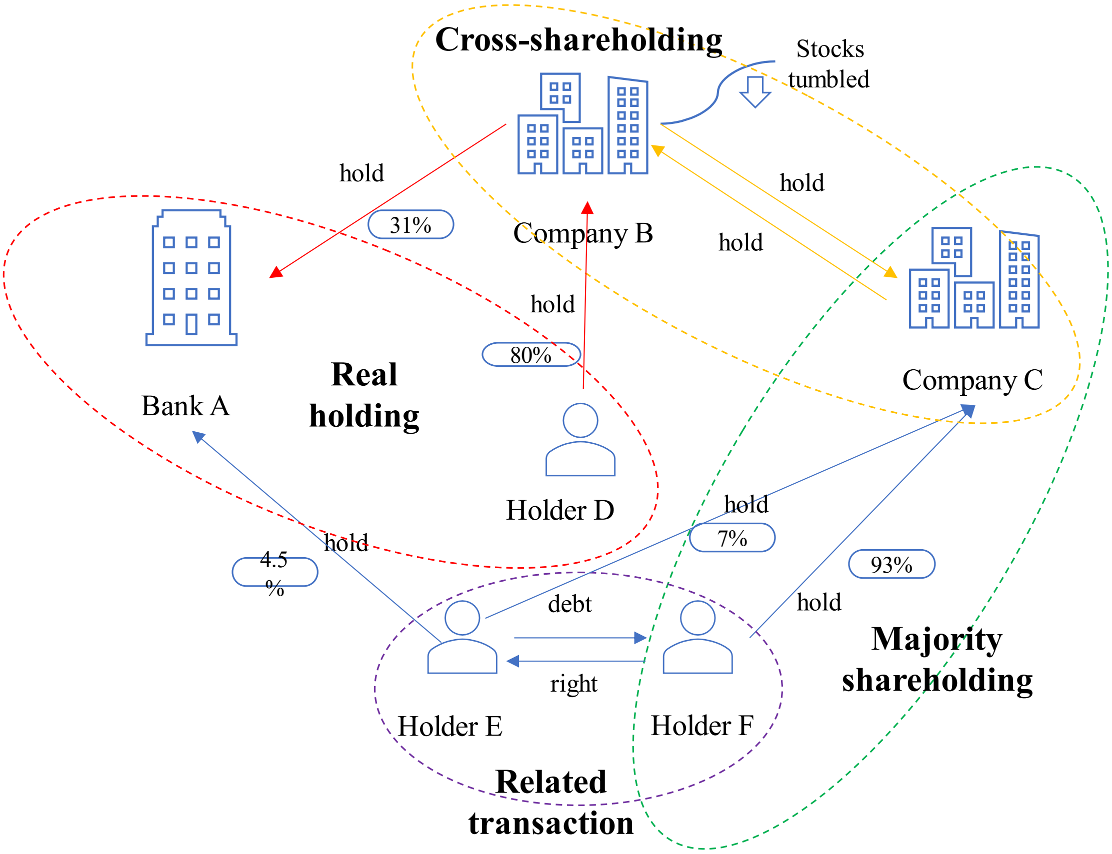


Figure 11 Equity association characteristics and behavior based on RDF triples

Three-dimensional space network of equity formed by the concept cube covers a large number of holding relationships. We can easily locate the unique position of the company based on the coordinates represented by probability, quickly find the number of outgoing chains between companies, and use PageRank algorithm to suspect that the company is holding Correlation analysis between people and potential risks. The algorithm determines the importance of nodes through iterative calculations. The importance PR value of a company is the iterative result of the importance of all chains to its own company. Since the advantages of concept cubes are hierarchical interconnection and spatial positioning, the PageRank algorithm is used to analyze key nodes and it is easy to judge their dimensional tendencies and the scope of their concepts. Once a company or organization is at risk, this spatial correlation structure can be very effective. To a large extent, it helps to judge the potential risk types (financial market systemic risk, corporate transaction risk, internal risk, etc.), and find the root cause of the risk based on the relationship. In addition, we can also treat the spatial coordinates as node embedding, use the Floyd algorithm to analyze the company's associated investment path and the shortest associated path, and explore the industry background behind the company.

It can be seen that there are various modes of complex equity knowledge association based on multi-domain ontology, such as classification association driven by static sub-domain ontology, event association and time association driven by dynamic sub-domain ontology, and spatial association embodied in conceptual cube structure.

Figure 12 shows "*Zhirong*" financial big data platform developed in cooperation with Shenzhen Securities Exchange. The platform relies on multi-layer ontology knowledge modeling, through the integration and fusion of equity knowledge data sets, and uses D2RQ technology to realize relational data The conversion to RDF format retains spatial semantic information, so it can complete tasks such as related entity query and financial dynamic analysis. In the association query, with the help of the dimensional characteristics embodied by the concept cube, the query of the three types of association relationships in the field, industry, enterprise, and internal environment, can be realized, and the association path between them can be displayed; in the dynamic analysis, specific The bank’s equity structure can realize the correlation analysis and risk identification of typical financial issues such as cross-shareholding, related transactions, capital family and actual controllers through equity penetration query analysis, in order to achieve the ultimate goal of risk management.

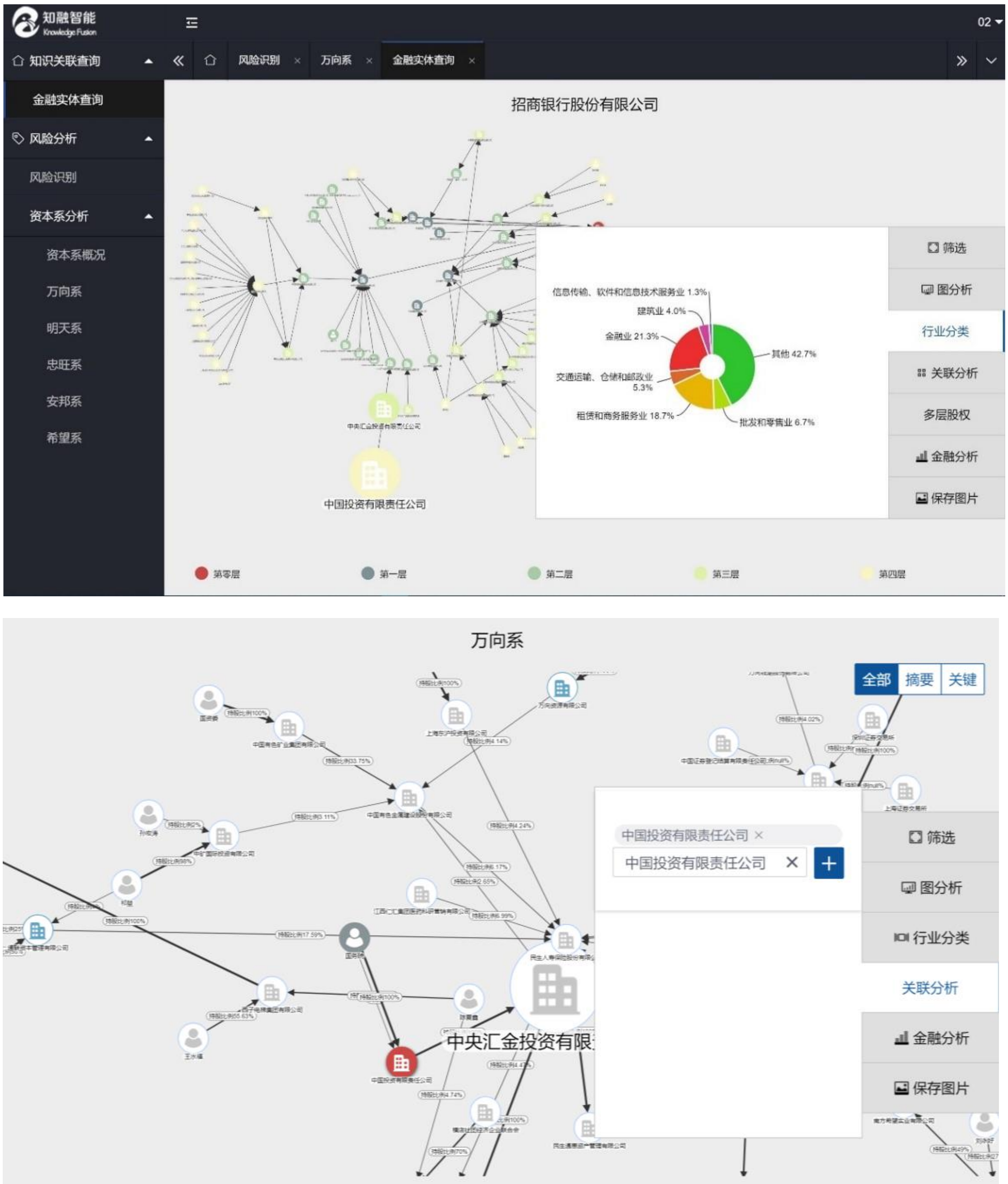


Figure12 Knowledge correlation analysis of "Zhirong" financial big data system

The field of financial equity covers a wide range and complex content. It not only involves the operation and governance of enterprises, but also includes environmental factors such as national policies and regulations and macroeconomic control guidelines. It is the core link of financial market transactions and the foundation of the corporate governance system. Through the link aggregation of different data resources, the platform forms a variety of equity association models, and discovers the characteristics of the equity structure between companies and some root causes that may lead to risk contagion, superposition and spread. As the number of examples populated in the ontology expands, it is possible to determine the global equity association model of all institutions in the financial system, reason and dig out more complex industry backgrounds, capital family and risk characteristics, and realize new values.

5 Results and discussion

Finance is a strong data-driven industry. Obtaining and organizing information in a timely and efficient manner and providing support for decision-making is the focus of the financial sector. The value of financial big data stems from the extensive knowledge associations it implies, while the traditional flat organization of financial big data ignores the internal connection of the data, nor does it consider the effective organization and integration of multi-source heterogeneous data. Multi-layer domain ontology describes and expresses different levels of knowledge systems from multiple levels, which can better realize knowledge modeling and association in the financial field.

Relying on the theory and method of multi-layer ontology representation, this study realizes the standardized representation of financial equity structure by constructing a multi-layer domain ontology model including the basic layer, the conceptual layer and the instance layer; at the same time, it considers the three levels of "industry-enterprise-inner environment". The impact of the concept of dimensionality on domain knowledge has constructed a "conceptual cube" data model with spatial correlation characteristics to provide a new research perspective for the organization of financial big data. At the application level, the research analyzes in detail the classification association, event association, time association and spatial association modes revealed by the equity network structure driven by multi-layer ontology. At the same time, it provides an example demonstration of financial equity knowledge association, which can be improved based on platform examples. A good discovery of typical financial management issues such as cross-shareholding and capital family groups provides a theoretical and practical basis for financial knowledge innovation services and effective identification of financial risks.

The research can further study the spatial correlation characteristics of concept cubes, and deeply analyze the differences between the model structure based on knowledge correlation and data cubes in the future, so as to expand the application scenarios of concept cubes. What's more, with the development of knowledge graphs, how to map knowledge graph directly by using multi-layer ontology is also an issue worthy of attention to truly realize the organic unity of the model-level knowledge representation and the instance-level association analysis.

References

Alobaid, A., Garijo, D., Poveda-Villalón, M., Santana-Perez, I., Fernández-Izquierdo, A., & Corcho, O. (2019). Automating ontology engineering support activities with OnToology. *Journal of Web Semantics*, *57*, 100472.

Asghari, M., Sierra-Sosa, D., & Elmaghraby, A.S. (2020). A topic modeling framework for spatio-temporal information management. *Information Processing & Management, 57*(6), 102340.

Escobar, P., Candela, G., Trujillo, J., Marco-Such, M., & Peral, J. (2020). Adding value to Linked Open Data using a multidimensional model approach based on the RDF Data Cube vocabulary. *Computer Standards & Interfaces*, *68*, 103-378.

Gao, J., Ding, K., Pan, Y., & Yuan, J. (2015). A Review on the Research of Knowledge Relevance. *Information Theory and Practice*, *38*(08), 139-144.

Gruber, R. (1993). *Towards Principles for the Design of Ontologies Used for Knowledge Sharing*. In International Workshop on Formal Ontology, Padova, Italy, 1993. Available as technical re-port KSL-93-04.

Kotis, K.I., Vouros, G.A., & Spiliotopoulos, D. (2020). Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations. *The Knowledge Engineering Review*, (04), 35.

Li, J., Chen, S., Chen, W., Andrienko, G., & Andrienko, N. (2018). Semantics-space-time cube: A conceptual framework for systematic analysis of texts in space and time. *IEEE transactions on visualization and computer graphics*, *26*(4), 1789-1806.

Li, J., Ma, Y., Sun, L., & He, Y. (2017). Research on Multi-level Ontology Metamomodel Construction Based on Semantic Consension. *Journal of Information Science*, (05), 66-74.

Li, X., & Fan, M. (2019). Knowledge association in big data. *Information Theory and Practice*, *42*(02), 72-77+111.

Li, X., & Liu, Y. (2019). A review of spatiotemporal data modeling methods. *Modern Library and Information Technology*, *003*(003), 1-13.

O'Riain, S., Harth, A., & Curry, E. (2012). Linked data driven information systems as an enabler for integrating financial data. In *Information Systems for Global Financial Markets: Emerging Developments and Effects* (pp. 239-270). IGI Global.

Patel, K. , Dube, I. , Tao, L. , & Ning, J. . (2015). Extending OWL to Support Custom Relations. IEEE International Conference on Cyber Security & Cloud Computing. IEEE.

Petrova, G. G., Tuzovsky, A. F., & Aksenova, N. V. (2017). Application of the financial industry business ontology (FIBO) for development of a financial organization ontology. *Journal of Physics: Conference Series, 803*, 012116.

Ren, S., Xu, H., Liu, X., & Cao, W. (2009). Research on Domain Analysis and Design Method of Hierarchical Ontology Model. *Computer and Applied Chemistry*, *26*(011), 1385-1388.

Shams Fa Rd, M., & Barforoush, A. A. (2004). The state of the art in ontology learning: a framework for comparison. *Knowledge Engineering Review*, *18*(4), 293-316.

Shi, Z. (2010). *Research on High Performance Data Cube and Its Semantics.* (Doctoral dissertation, Beijing Jiaotong University). Received from https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFD0911&filename=2010040954.nh

Tang, X., Ma, F., Fu, W., & Zhang, R. (2019). Financial knowledge representation and risk identification from the perspective of knowledge relevance. *Journal of Information Science and Technology*, *038*(003), 286-298.

Wen, T., Liu, X., & Liu, J. (2010). The Theoretical Basis of Knowledge Relevance.

Library (04), 12-14.

Wen, T., Gong, J., Zhang, R., Liu, X., & Liu, J. (2011). Knowledge Relevance: Connotation, Characteristics and Types. *Library, 000*(004), 32-35.

Zhang, L., Zhao, M., & Feng, Z. (2019). Research on knowledge discovery and stock forecasting of financial news based on domain ontology. *International Journal of Information Technology & Decision Making.* 953-979. doi: 10.1142/S0219622019500160.

Zhang, R., Tang, X., Wang, D., & Pan, J. (2018). *Information Theory and Practice*, *41*(10), 135-140.