Formation of Knowledge Organization Ecosystem Framework and Research Progress: Postprint

Authors: evergreen, Yang Jing, Li Yongze

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] This study introduces the principles and methods of ecology into research on knowledge organization systems to better exert their functions and roles. [Method/Process] Based on ecological research achievements, ecological principles and methods are incorporated into the construction and application of knowledge organization systems, wherein biological individuals correspond to concept instances, species populations correspond to concepts, biological communities correspond to vocabularies, and biological environments correspond to literature environments. On the basis of complete and consistent correspondence in both structure and function, the working principles and methods of knowledge organization systems are investigated. [Results/Conclusion] The framework of the knowledge organization ecosystem is refined and improved across four levels: concept instances, concept populations, concept communities (vocabularies), and ecosystems, with the primary research content at each level being analyzed and summarized.

Full Text

Preamble

Vol. 63 No. 7 April 2019
ChinaXiv Cooperative Journal

Formation and Research Progress of the Knowledge Organization Ecosystem Framework

Chang Chun[1], Yang Jing[2], Li Yongze[3]
[1] Institute of Scientific and Technical Information of China, Beijing 100038
[2] Chinese Version Library, Beijing 100005
[3] School of Information Management, Nanjing University, Nanjing 210023

Abstract

[**Purpose/Significance**] This study introduces ecological principles and methods into knowledge organization system research to better fulfill the functions and roles of knowledge organization systems. [**Method/Process**] Based on ecological research findings, we introduce ecological principles and methods into the construction and application of knowledge organization systems. Biological individuals correspond to concept instances, species populations to concepts, biological communities to vocabularies, and biological environments to literature environments. On the basis of this complete structural and functional correspondence, we investigate the working principles and methods of knowledge organization systems. [**Result/Conclusion**] The knowledge organization ecosystem framework is refined and perfected across four levels: concept instances, concept populations, concept communities (vocabularies), and the ecosystem. The main research contents at each level are analyzed and summarized.

Keywords: Knowledge organization ecosystem; Ecology; Knowledge organization system; Thesaurus
Classification Number: G254
DOI: 10.13266/j.issn.0252-3116.2019.07.017

Introduction

Ecology is the science that studies the relationships between organisms and their surrounding environment [1]. In 2015, ecological principles and methods were introduced into knowledge organization system construction and application, and Chang Chun [2] proposed the concept of Knowledge Organization Ecosystem (KOES). After nearly three years of research and practice by his team—through studies on conceptual ecological niches [3], hierarchical relationships based on energy flow [4], and classification of conceptual relationships based on interspecies relationships [5]—the framework has been further refined. In terms of composition, the previous model of "species plus environment constituting an ecosystem" has been elaborated into a structure comprising biological individuals, species populations, biological communities, plus the biological environment. Correspondingly, the Knowledge Organization System (KOS, also referred to as "vocabulary" in this paper) has been similarly refined into instances, concepts, vocabularies, plus the literature environment to constitute a knowledge system. Applying ecosystem principles and methods to knowledge systems, with complete correspondence in both structure and function, forms the Knowledge Organization Ecosystem. Research and application content has been expanded from the previous three dimensions to four levels: individual, concept, vocabulary, and knowledge system, achieving complete correspondence with the disciplinary framework of ecology. This paper primarily introduces the specific elaborations and research content.

2. Overall Framework of the Knowledge Organization Ecosystem

2.1 Structure and Composition of the Knowledge Organization Ecosystem

The refined knowledge organization ecosystem formation process is shown in Figure 1 [Figure 1: see original paper]. In ecology, individuals of the same species form a population, which is the basic unit of a biological species. Different species form a community, which is the collective term for all organisms including plants, animals, and microorganisms. An ecosystem consists of a biological community and its entire living environment. Correspondingly, in knowledge organization systems, different instances form concepts, which are the basic units of knowledge organization systems. Different concepts form domain vocabularies, which contain various types of concept groups including common nouns and proper nouns. A knowledge system consists of domain vocabularies and the literature they index and retrieve. Biological individuals correspond to concept instances, species populations to concepts, biological communities to domain vocabularies, and ecosystems to knowledge systems. On the basis of this complete structural and functional correspondence, introducing ecological principles and methods into the functions and roles of knowledge organization systems forms a complete Knowledge Organization Ecosystem. In KOES, species correspond to concepts, environment to literature, and species relationships to conceptual relationships. All research directions are developed and completed based on these fundamental correspondences.

2.2 Correspondence at the Instance Level

Individual-level ecology primarily studies the relationships between biological individuals and their environment, including individual adaptability to the environment, which encompasses both biological and physical environments. Concepts and instances are important components of knowledge organization systems, where instances (entities) have a natural correspondence with biological individuals in ecology—even the English term “individual” is the same. For example, a clover plant, a sheep, a cow, a horse, or a tiger on a grassland are all different biological individuals that form their own species populations and interact with the physical environment (temperature, humidity, sunlight, water, etc.). Entity nouns such as personal names, place names, and institutions appearing in titles, abstracts, keywords, or full texts represent concepts, and each occurrence can be considered an instance individual. Instance individuals exist within the literature environment; different instance individuals form their respective concepts and have associative relationships with the literature environment in terms of word form, meaning, context, and semantics. Instance individuals originate from the literature environment while also constituting components of it.

2.3 Correspondence at the Concept Population Level

A population is a collection of individuals of the same species inhabiting the same area. A population is a group composed of individuals that exhibits characteristics not present at the individual level, such as birth rate, death rate, growth rate, sex ratio, intra-specific relationships, inter-specific relationships, age structure, and spatial distribution. For example, a sheep herd on a grassland has a lead sheep that determines the direction for the others. The herd exhibits population behaviors and characteristics, such as annual statistics on lamb birth and death rates, age structure, and the importance of the lead sheep. In knowledge organization, the counterpart to a population is the concept. Concepts are units of thought, collections composed of instance individuals. They can be concepts represented by preferred terms in a thesaurus, classes in a classification system, or classes in an ontology. Concepts similarly have characteristics of birth, hot research topics, and deactivation. They have equivalence relationships between preferred and non-preferred terms, as well as inter-concept relationships such as associative and hierarchical relationships. They exhibit features like concept maturity determination through word frequency fluctuations and statistical parameters such as equivalence rate, reference degree, and association ratio. Species populations and concepts in knowledge organization have a completely corresponding relationship.

2.4 Correspondence at the Vocabulary Community Level

A biological community refers to the regular assembly of different species populations within a specific space, comprising animals, plants, and microorganisms inhabiting the same area. At the community level, ecological characteristics include community structure, succession, diversity, and stability. On land, there are grassland communities, wetland communities, forest communities, alpine communities, etc. Correspondingly, KOS with varying degrees of semantic relationships align with biological communities. Vocabularies also have macrostructures, micro-structures, vocabulary updates, concept diversity, and concept stability features, with distinctions between domain vocabularies and comprehensive vocabularies. Domain vocabularies can be directly corresponded with biological communities.

2.5 Correspondence at the Ecosystem Level

An ecosystem is the combination of a biological community and its abiotic environment within a certain space. Research focuses on material cycling, energy flow, and information transfer. Corresponding to the systematic characteristics of knowledge organization activities, KOS construction and application also exhibit systemic features such as knowledge transfer and vocabulary application. Earth's energy flow manifests when green plants (producers) convert light energy into chemical energy stored in organic matter; herbivores (primary consumers) transfer part of this energy to the herbivore level through grazing; carnivores (secondary consumers) transfer part of the energy to the carnivore level through

predation. Energy in ecosystems exhibits unidirectional transfer and stepwise decrement. Similarly, knowledge created by humans in production and social activities is transferred to readers through publications, journals, conference proceedings, networks, and libraries. The hierarchical structure of knowledge organization systems, from generic to specific concepts through genus-species relationships, also exhibits energy transfer characteristics. In summary, the construction and application of KOS conform to ecosystem systematic characteristics.

3. Research Directions in the Knowledge Organization Ecosystem

3.1 Research Directions in the Knowledge Organization Ecosystem

Research directions in KOES are introduced across four levels: instance individual, concept population, vocabulary community, and ecosystem. Detailed contents are shown in Table 1 .

Table 1. Research Directions in the Knowledge Organization Ecosystem

Level	Research Directions
Instance Individual	Research on entity noun mining in literature environment Research on instance individuals originating from natural language Research on word frequency statistical changes of instance individuals
Concept Population	Research on characteristics of individual concept maturation process Research on intra-species relationships of synonymous and near-synonymous concepts Research on inter-species relationships of hierarchical and related concepts
Vocabulary Community	Research on measurable characteristics of vocabulary concepts Research on diversity characteristics of vocabulary concepts Research on stability characteristics of vocabulary concepts
Ecosystem	Systematic research on category hierarchical structure Systematic research on word family hierarchical structure Research on knowledge transfer and energy flow characteristics

3.2 Research Directions at the Instance Individual Level

Research at the instance individual level is limited to entities such as terms and words, focusing primarily on word form and meaning standardization, such as selection methods for preferred terms [6]. It includes natural language processing of literature texts to obtain professional terms through text segmentation,

automatic acquisition of proper noun entities such as personal names, place names, and institution names through data mining, and identification methods for synonyms [7] and homographs [8]. All word-level research can be categorized under this level. From the perspective of KOS application and services, research directions related to individual users, such as personalized services, also belong to this level.

3.3 Research Directions at the Concept Population Level

Biological taxonomy uses species populations as units, while KOS construction and application use concepts as units. At the concept level, numerous studies can be conducted based on species population characteristics, such as researching concept maturation cycles through word frequency statistics [9], determining specialized [10] and general concepts [11] through domain distribution, studying intra-species relationships of synonymous and near-synonymous concepts [12], establishing methods for inter-concept hierarchical relationships [13], and establishing methods for inter-concept associative relationships [5]. From the perspective of using species populations as units to constitute species, corresponding research using concepts as units for KOS construction and application offers numerous meaningful directions. Extending to KOS applications, this level can also include user ontology construction, user profiling, and other user classification-related research directions.

3.4 Research Directions at the Vocabulary Community Level

Biological communities focus on collections formed by various species populations, studying community structure and biodiversity characteristics. Correspondingly, each vocabulary contains different numbers of concepts. At the vocabulary level, research addresses concept diversity, conceptual niche spatial structure determining vocabulary stability characteristics [3], quantitative characteristics such as reference degree and association ratio, hierarchical [4] or network structures composed of conceptual relationships, and conceptual dynamic succession in vocabulary updating and maintenance. All associative and attribute characteristics among different concept groups belong to this level. Extending to KOS services, this level also includes research on different service characteristics of comprehensive, specialized, and university libraries or documentation institutions.

3.5 Research Directions at the Ecosystem Level

From a systematic perspective, ecosystem research examines material cycling, energy flow, and information transfer. Correspondingly, KOS also exhibits various systematic characteristics, such as automatic thesaurus construction based on natural language [14], systematic research on category systems and word family structures, inter-system interoperability [15], and hierarchical relationship research based on energy flow systematicity [4]. Additionally, besides systematic characteristics, the ecosystem level involves environmental factors. The

fundamental application of KOS is literature indexing and retrieval, making retrieval-related research part of this level [16]. From a KOS service perspective, this level also includes various systematic research directions related to users, such as cultural background, information infrastructure, and social environment affecting knowledge acquisition.

4. Related Discussions on the Knowledge Organization Ecosystem

4.1 Disciplinary Attributes of the Knowledge Organization Ecosystem

The knowledge organization ecosystem applies ecological principles, starting from ecosystem attributes and characteristics, to establish correspondences between species and concepts, and between literature information and biological environment. Concepts serve as the basic units of knowledge organization systems, just as species are the basic units of biological taxonomy. The knowledge organization ecosystem consists of interacting and interrelated factors including instance individuals, concepts, vocabularies, and literature environments. Drawing on ecological methods, it theoretically establishes a research system for KOS construction and application, explaining the working principles of KOS and providing theoretical foundations. The conceptual attributes of KOES belong to knowledge management, with research content encompassing various life characteristics related to knowledge. However, the research scope does not include actual living organisms, plants, animals, or microorganisms, nor natural matter and energy. It represents the application of ecological principles and methods to knowledge organization, falling within the scope of management science. Given the complexity of knowledge formation, development, and application, KOES imports and references ecological principles and methods rather than mechanically corresponding or artificially interpreting them. As research deepens, KOES will demonstrate even more complex characteristics in some aspects.

4.2 Correspondence Between Population/Individual and Concept/Instance

For biological populations, one species can form one population, and a population typically contains multiple individuals. Individual scale directly relates to the species' role and status in the ecosystem. For example, sheep on a grassland constitute one species, with each sheep being an individual. The herd size significantly impacts the prosperity of sheep and the grassland ecosystem. Species correspond to concepts in knowledge organization systems, and concepts should have multiple instances. How, then, are concept instances manifested in KOES? Through research, our conclusion is that each occurrence of a concept's preferred term in literature counts as one instance. For example, in a paper on "ontology," if the term appears in the title, abstract, and keywords, and 20 times in the main text, then in this paper, "ontology" is one concept with 23 instances. Each occurrence's position differs in attributional significance—

occurrences in the title, abstract, keywords, or conclusion contribute more to thematic expression than those in literature reviews or future work sections. For large databases, we can treat each document thematically related to "ontology" as one concept instance. For instance, in the CNKI database, 47,486 documents contain the subject term "ontology" [17], indicating that the concept "ontology" has 47,486 instances. This approach was used in the study "Research on Concept Word Frequency Changes Based on Logistic Population Growth Law" [9].

4.3 Correspondence Between Species Relationships and Concept Relationships

"Animal - Mammal - Cow" represents a conceptual inclusion relationship. Biological inclusion relationships correspond completely to hierarchical relationships in KOS, specifically genus-species relationships: part of "mammal" is "cow," and all "cows" are "mammals." Biological species exhibit predation, competition, parasitism, and symbiosis relationships, distinguishable through directionality and dependency. Predation is a unidirectional process from predator to prey, but food is substitutable; competition is mutual and non-directional without dependency; parasitism is unidirectional dependency of the parasite on the host; symbiosis is bidirectional and beneficial to both parties. These interspecies relationships can correspond to types of associative relationships. We analyzed and tested different types of associative relationships in the ISO 25964-1 thesaurus international standard and found they could all be uniquely categorized into one of these interspecies relationships. Empirical verification using the "Chinese Thesaurus (Engineering Volume)" confirmed that all associative relationships can be uniquely classified as predation, competition, parasitism, or symbiosis [5]. For example, unidirectional dependency corresponds to parasitism: the relationship between "electric current" and "ampere" is associative—"electric current" exists objectively first, and "ampere" as its unit depends unidirectionally on "electric current." Without electric current, the unit "ampere" would not exist, making it a parasitic relationship. Whether there are universal terms for KOS to describe such relationships as alternatives to biological interspecies relationships remains an exploratory research direction.

4.4 Focus and Relativity of Research Units

Species is the basic unit of biological taxonomy (e.g., horse, cow, sheep are different species), while concept is the basic unit of KOS (e.g., concept is the fundamental component of a thesaurus). In KOES, basic units and research units can be defined at different levels. For instance, at the concept individual level, the research unit is the instance individual—such as preferred term selection methods [6] addressing term standardization and identification of synonyms [7] and homographs [8]. At the concept level (species), research focuses on concept attributes like maturation cycles [9]. At the vocabulary level (community), research examines conceptual niche spatial structure determining vocabulary sta-

bility [3] and quantitative characteristics like reference degree and association ratio. At the knowledge system level (ecosystem), research includes systematic issues like interoperability [15] and retrieval studies [16]. Thus, different research unit scopes define different research problems.

References

[1] Niu Cuijuan, Lou Anru, Sun Ruyong, et al. *Fundamentals of Ecology* [M]. 3rd ed. Beijing: Higher Education Press, 2015.
[2] Chang Chun. Research on Knowledge Organization Ecosystem for Thesaurus Construction [J]. *Library and Information Service*, 2016, 60(15): 101-107.
[3] Yang Jing, Chang Chun. Research on Concept Stability Based on Niche Theory [J]. *Library and Information Service*, 2016, 60(13): 27-32.
[4] Li Yongze, Chang Chun. Research on Word Family Hierarchical Structure Based on Ecological Energy Transfer [J]. *Journal of Intelligence*, 2017, 36(3): 161-165, 172.
[5] Li Yongze, Chang Chun. Classification of Thesaurus Associative Relationships Based on Ecological Interspecies Relationships [J]. *Library and Information Service*, 2018, 62(8): 123-129.
[6] Chen Baixue, Chang Chun, Liu Chunyan, et al. Research on Concept Preferred Term Selection Methods [J]. *Journal of Intelligence*, 2015, 34(12): 170-175.
[7] Wang Liuan, Chang Chun. Research on Synonymous Term Identification in Use/For (UF) Relationships [J]. *Information Studies: Theory & Application*, 2014, 37(9): 97-100, 91.
[8] Chen Baixue, Chang Chun. Research on Machine-Aided Homograph Identification Methods [J]. *Digital Library Forum*, 2015(5): 8-13.
[9] Yang Jing, Chang Chun. Research on Concept Word Frequency Changes Based on Logistic Population Growth Law [J]. *Information Science*, 2017, 35(8): 15-18.
[10] Chang Chun, Lai Yuangen. Research on Machine-Aided Classification Methods for Specialized Concepts [J]. *New Technology of Library and Information Service*, 2011(10): 34-39.
[11] Chang Chun, Lai Yuangen. Method for Acquiring General Concepts in Digital Environments [J]. *Library and Information Service*, 2011, 55(23): 22-25.
[12] Zhang Bing, Chang Chun. Research on English Synonym Group Construction Based on Term Lemmatization [J]. *Journal of Intelligence*, 2014, 33(7): 171-175.
[13] Deng Panpan, Chang Chun. Research on Concept Multiple Hierarchical Relationships in Thesaurus Construction [J]. *Information Science*, 2015, 33(5): 59-62.
[14] Du Huiping, He Lin, Hou Hanqing. Automatic Construction of Natural Language Thesaurus Based on Cluster Analysis [J]. *Journal of the National Library of China*, 2007(3): 44-49.
[15] Chang Chun, Zeng Jianxun, Wu Wenna, et al. Design of Concept Mapping

Framework Between Chinese Thesaurus and English Super-Science Thesaurus [J]. *Digital Library Forum*, 2012(12): 28-33.
[16] Chang Chun. Research on Thesaurus Term Service Methods [J]. *Library and Information Service*, 2012, 56(22): 12-15.
[17] Ontology [EB/OL]. [2018-06-10]. http://kns.cnki.net/kns/brief/result.aspx?dbprefix=SCDB.

Author Contributions

Chang Chun: Responsible for conceptualization and writing;
Yang Jing: Responsible for data collection and empirical analysis;
Li Yongze: Responsible for data collection and empirical analysis.

Note: Figure translations are in progress. See original paper for figures.