publications

MDPI

Article

Knowledge Organization of Traditional Chinese Architectural Books: A Case Study of the *Yingzao Fashi*

Enmao Qiao [1,2], Jie Dong [1,*] and Shubin Zhou [1]

[1] Institute for the History of Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China
[2] School of Architecture and Planning, Inner Mongolia Technical University of Construction, Hohhot 010070, China
* Correspondence: 20204910001@mails.imnu.edu.cn

Abstract

Traditional Chinese architectural books transmit the craft of regulated timber-frame construction. They do so primarily through normative texts rather than descriptions of artifacts, making them highly historically and academically valuable among Chinese scientific and technical sources. The knowledge in these books, however, has long been dispersed across three separate research traditions—bibliography, terminological exegesis, and architectural history—which has limited both the depth of research and the breadth of its application. In this study, a regulation-centered semantic model for such books is proposed and evaluated. An ontology of 86 classes, 32 object properties, and 27 data properties was designed with a six-dimension procedure (scenario, reuse, entities, relations, constraints, and evaluation), formalized in Protégé, checked using the Pellet reasoner, and populated with the *damuzuo* (structural carpentry) chapters of the Northern Song *Yingzao Fashi* (1103), yielding 641 term records and 2553 relation assertions that were loaded into Neo4j. Unlike artifact-centered heritage models, the regulation entry is the unit of description, modular dimensions are stored in *fen* so that derived sizes remain derivable, and constructional relations are typed (joining, supporting, insertion, and part-of) rather than collapsed into a generic association. The model was evaluated quantitatively. Compared against an independently published Song–Qing glossary, the dataset records 75.5% of the Song *damuzuo* terms exactly (71 of 94); a further 17.0% match only partially, present as compounds or short forms of a recorded term, and are not considered covered. Supervised classifiers trained on the manual annotations reproduce the eight record categories of the annotation with an accuracy of 0.79 and a Cohen's kappa of 0.70 under five-fold cross-validation, which identifies the Position–Component boundary as the main source of disagreement. A distantly supervised character-level tagger recovers term mentions in the source text with an entity-level F1 of 0.94. Inverse and transitive reasoning adds 751 statements (29% of those asserted) that are implied but not stated in the text, including an eight-member support path from beam to flying rafter, and a pilot transfer to the Qing *Gongcheng Zuofa* (530 terms) shows that the structural class scheme transfers but surface vocabulary does not, so cross-text alignment must be performed at the concept level. The dataset, ontology, and evaluation scripts are provided alongside this article.

Keywords: traditional Chinese architectural books; *Yingzao Fashi*; semantic model; ontology; knowledge graph; digital humanities

Check for updates

Academic Editor: Andrew Kirby

Received: 5 October 2026
Revised: 12 November 2026
Accepted: 19 November 2026
Published: 23 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In Chinese premodern architecture, building frames are predominantly made of timber. The craftsmanship required to build these frames has been transmitted both through master–apprentice oral instruction and written treatises. In 2009, the traditional craftsmanship of Chinese timber-framed structures was included on the UNESCO Representative List of the Intangible Cultural Heritage of Humanity (UNESCO, 2009), confirming its historical and cultural significance internationally. The traditional written means of transmission is anchored in two state-issued building standards, the Northern Song *Yingzao Fashi* (Building Standards; Guo, 1998; J. Li, 1103) and the Qing *Gongcheng Zuofa* (Engineering Practice; Ministry of Works, Qing Dynasty, 1734), and is complemented by local and departmental craft regulations and private treatises such as Yao Chengzu's *Yingzao Fayuan* (Yao, 1986) and Xue Jingshi's *Ziren Yizhi* (Xue, ca. 1264). These texts share a compilational convention: each entry names a component, lists its permitted variant names, states its dimensions using a modular unit, and specifies its position and the components it is in contact with. These texts record the conventions of construction rather than describing existing buildings and therefore constitute a distinct genre within the Chinese scientific and technical literature (a fuller account of the *Yingzao Fashi* and the chapters used in this study is given in Section 2.1; a glossary of the Chinese terms used in this paper, with a note on the script in which they are printed, is provided in Appendix A).

Research on these books has long proceeded along three paths. Bibliographic studies examine their compilation, transmission, and editions, laying the groundwork for preservation and reuse. Philological studies explicate their terminology, with Liang Sicheng's annotated edition (J. Li, 1103/1983) and Chen Mingda's glossary (M. Chen, 2010) serving as the two standard reference works after the founding of the Society for Research in Chinese Architecture in the 1930s, while Pan and He (2005) further clarified the organization of the regulation chapters. Research on architectural history examines the correspondence between the recorded norms and surviving monuments through measured surveys and typological comparison. G. Wang's (2023) survey of terminological scholarship used since the founding of the Society for Research in Chinese Architecture shows that the same term is often interpreted differently across annotation traditions and that these divergences themselves represent knowledge that needs to be recorded. Each path has produced results, but disciplinary separation has generated knowledge barriers that restrict the integrated exploration and reuse of the historical knowledge these texts contain.

Digital humanities research has addressed such barriers with ontologies, linked data, and knowledge graphs (Cheng & Chou, 2022; W. Li et al., 2024b; C. Liu et al., 2024; W. Wang et al., 2022; Zeng, 2019). Large infrastructures such as Europeana/DM2E (Baierer et al., 2017), the Sampo portals (Hyvönen, 2023), Beyond 2022 (Debruyne et al., 2022), and Digital Dunhuang (F. Li & Xia, 2023) established how heritage collections can be modeled and linked, and the same techniques have since been applied to museum collections (J. Li & Bikakis, 2023), Chinese ancient architecture (J. Chen & Ou, 2020), Thai cultural practices (Chansanam & Tuamsuk, 2022; Jaitiang et al., 2023), temporal expressions in historical Chinese texts (L. Wang et al., 2024), and oral history (Vrachliotou & Papatheodorou, 2024). For ancient books in particular, work now spans the whole life cycle of the material, from catalog curation (W. Li et al., 2024a) and text mining (Joo et al., 2022; C. Liu et al., 2024) to knowledge organization of stele inscriptions (J. Liu et al., 2023), platform and repository building (C.-M. Chen & Chang, 2019; Zhao & Chen, 2023), and ontology-based organization of Chinese classics (L. Wang et al., 2023), medieval geographical works (Bartalesi et al., 2023), Greek travel texts (Foka et al., 2020), collection catalogs (Ortolja-Baird et al., 2019), and Dante's works (Bartalesi & Meghini, 2017). Two groups of prior work are closest to the present study. The first is architectural-heritage ontologies. CIDOC-CRM (Doerr, 2003) and

its building-oriented applications, ArCo (Carriero et al., 2019), the conservation-process ontology of Acierno et al. (2017), and semantically enriched HBIM (Quattrini et al., 2017) describe buildings, their parts, and the events that affected them; their unit of description is the physical object or the intervention, and the dimensions are recorded as measured values. J. Chen and Ou's (2020) semantic model for Chinese ancient architecture likewise organizes narratives about existing buildings. The second group comprises knowledge graphs of Chinese classical texts, which model persons, places, time, and textual units (W. Li et al., 2024b; C. Liu et al., 2024; L. Wang et al., 2023, 2024) or, for craft texts, the instruments and repertoires of a musical tradition (Zhou et al., 2024, 2025). None of these models takes a building *regulation* as its unit, represents the modular derivation of dimensions, or treats historically attested variant names as first-class data, and few report quantitative evaluation beyond a consistency check (Brank et al., 2005; Vrandečić, 2009).

These gaps matter because traditional architectural books present three distinctive features for knowledge organization. First, the unit to be organized is the regulation rather than the artifact: artifact-centered cultural heritage metadata asks for name, date, place, material, and size and cannot accommodate norms that determine dimensions before any object exists. Second, dimensions have no absolute value: the *Yingzao Fashi*, in the '*cai*' entry that opens *juan* 4, '*Damuzuo* Regulations I', states that 'in all building construction, the *cai* is the fundamental module', and since the *cai* comes in eight grades and all dimensions in the book are stated in *fen*, treating components as base units with sizes as attributes records the numbers while losing their derivation. Third, the terminology is historical and polysemous, with numerous variant names that the text itself registers in the *Kanxiang* section on variant names, in the two General Glossary *juan* (*juan* 1 and 2), and in the interlinear notes attached to entries in the regulation chapters. These features leave the concepts of traditional architectural books ill-defined and their structures unsystematic, creating difficulties in practical use and hindering their modern development and application.

This study therefore investigates how a regulation-centered semantic model for traditional architectural books can be designed, how well it represents the source text, and whether it transfers to a second text, taking the *damuzuo* chapters of the *Yingzao Fashi* as the case. This research makes four contributions: (1) It specifies an ontology in which the regulation entry, not the artifact, is the unit of description; modular dimensions are kept in *fen* so that absolute sizes remain derivable from the modular grade; and constructional relations are typed (joining, supporting, insertion, part-of, and sequence), with a documentary warrant for every category. (2) It reports the first quantitative evaluation of such a model: coverage against an independently published glossary, model–human agreement of the manual classification measured with supervised classifiers, distantly supervised recognition of term mentions in the source text, and the yield of inverse and transitive reasoning. (3) It analyzes the resulting dataset of 641 terms and 2553 relations for patterns in terminology, component connectivity, and section proportions and documents the modeling difficulties encountered. (4) It conducts tests of transferability with a pilot ontology for the Qing *Gongcheng Zuofa* (530 terms) and cross-text classification experiments. Section 2 describes the materials, the engineering method, and the evaluation protocol; Section 3 presents the ontology; Section 4 reports the knowledge graph and worked queries; Section 5 presents the evaluation and dataset analysis; and Sections 6 and 7 contain a discussion and the conclusions.

2. Materials and Methods

2.1. Source Text and Dataset

The *Yingzao Fashi* was compiled by Li Jie, then vice director of the Directorate of Construction, on imperial order. Building on the earlier Xining-era standards and on methods

collected from practicing craftsmen, the text was completed in 1100 and promulgated in 1103 (Feng, 2012; Guo, 1998; J. Li, 1103). Its thirty-four *juan* of main text, with a prolegomenon (*Kanxiang*) and a table of contents, cover terminology, work regulations, labor quotas, material quotas, and drawings. The *damuzuo* chapters record the names, dimensions, positions, and working methods of the timber-frame members and state that all dimensions derive from the *cai–qi* module.

The case selection of this study was based on the regulation entries and terms recorded in the *juan* relating to the *damuzuo* of the *Yingzao Fashi*, extended into a larger dataset. The *juan* involved are *juan* 4 and 5, *Damuzuo* Regulations I and II; *juan* 17 to 19, *Damuzuo* Labour Quotas I to III; the *damuzuo* portion of *juan* 26, Material Quotas for Various Works; *juan* 30 and 31, *Damuzuo* Regulation Drawings; the opening *Kanxiang* section on variant names; and *juan* 1 and 2, the two parts of the General Glossary. A further 36 records were taken from outside these chapters, where terms needed to describe the timber frame are defined or first attested: *juan* 3 (moats, fortifications, and stonework; 6 records), *juan* 7 and 8 (small carpentry; 16), *juan* 13 (tiling and plastering; 2), *juan* 28 (quotas for nails and glue; 11), and the prefatory memorial (1). In total, 641 term records, 86 classes, and 2553 relation assertions were collected. The records were drawn from 17 sources, namely, fifteen *juan* of the main text (*juan* 1–5, 7, 8, 13, 17–19, 26, 28, 30, and 31), the *Kanxiang*, and the prefatory memorial, and the source is noted for every record. Records carrying variant-name links number 57 with 96 variant-name assertions. Person data were cross-checked against reference works such as the Dictionary of Chinese Historical Personal Names; temporal data mainly follow the Chinese historical chronology compiled by the Shanghai Library. Liang Sicheng's *Annotated Yingzao Fashi* and Chen Mingda's *Glossary of the Yingzao Fashi* were used only for recall checking, not as term sources, to avoid confusing the source literature with later interpretations. Each record carries a formal term, its record category and subclass, its *juan*-level source, the original sentence in which it occurs, and, where the text gives them, its dimensions in *fen*, its variant names, and its relations to other members. The records were first collected in spreadsheets and then serialized to RDF/OWL (Section 3) and loaded into Neo4j (Section 4). Table 1 summarizes the dataset. Two classifications of the records must be kept separate. The *record category* (类名) is the working category under which a term was collected and annotated; it has eight values, subdivided into 67 record subclasses, and it is the label that the classification experiments described in Section 5.3 predict. The *ontology class* is the type of the corresponding individual in the ontology of Section 3.3, where Building Structure has six subclasses. Six record categories correspond one to one to these six subclasses. The other two, material quota (料例类) and quantity operator (增减替), are types of statement rather than types of structure and have no class of their own in the ontology; their records are categorized according to what they denote, as the last column of Table 1 shows.

Table 1. Composition of the *Yingzao Fashi damuzuo* dataset (641 term records) by record category, and the ontology class under which the records of each category are typed.

Record Category (Chinese Label)	Records	Share	Record Subclasses (Examples)	Ontology Class of the Individuals (Section 3.3)
Component (构件类)	285	44.5%	gong, dou, ang, beam, column, purlin, rafter, board, timu, shuatou (18 subclasses)	AB:Component and its 18 subclasses
Construction method (做法类)	172	26.8%	puzuo arrangement, timber dressing, erection, roof curve, mortise–tenon (27 subclasses)	AB:ConstructionMethod and its 27 subclasses

Table 1. *Cont.*

Record Category (Chinese Label)	Records	Share	Record Subclasses (Examples)	Ontology Class of the Individuals (Section 3.3)
Position (位置类)	68	10.6%	positions on gong, dou, ang, beam, column, eave; spatial loci (13 subclasses)	AB:Position and its 13 subclasses
Building form (形制类)	42	6.6%	palace-hall type, hall type, minor building type (3 subclasses)	AB:BuildingForm and its 3 subclasses
Modular system (材契类)	31	4.8%	the eight cai grades, cai, qi, zucai, dancai (1 subclass)	AB:ModularGrade
Material quota (料例类)	24	3.7%	timber stock (14 records), nails (7), material grades (3); juan 26 and 28 (3 subclasses)	No class of its own: AB:Dimension (timber stock), AB:Component (nails), AB:ModularGrade (grades)
Dimension (尺寸类)	10	1.6%	units of length and of labor (1 subclass)	AB:Dimension, subclass MeasurementUnit
Quantity operator (增减替)	9	1.4%	increase, decrease, 'as appropriate', substitution (1 subclass)	No class of its own: AB:Dimension

Note: Shares are of 641 records. The records come from 17 sources (fifteen *juan* of the main text, the *Kanxiang*, and the prefatory memorial); *juan* 4 (230 records), *juan* 5 (186), the *Kanxiang* (63), and *juan* 19 (48) contribute most. The eight record categories have 67 subclasses in all (18 + 27 + 13 + 3 + 1 + 3 + 1 + 1). Two records listed as components (绞昂栿, 实拍襻间) are typed as construction methods in the ontology, a discrepancy flagged for the next release; all other counts follow the correspondence shown (Section 5.1).

2.2. Ontology-Engineering Method and Tools

The ontology was built with a six-dimension procedure—scenario, reuse, entities, relations, constraints, and evaluation—that the third author has used for other Chinese craft texts (Zhou et al., 2024, 2025). The procedure is not an alternative to established ontology-engineering methodologies but a domain-specific instantiation of them. Its first two dimensions correspond to the specification and knowledge-acquisition activities and to the reuse of existing ontological resources prescribed by METHONTOLOGY (Fernández-López et al., 1997) and by NeOn scenario 3 (Suárez-Figueroa et al., 2012); entities and relations correspond to conceptualization; constraints to formalization; and the evaluation dimension to the evaluation activity that both methodologies apply throughout the life cycle (Brank et al., 2005; Vrandečić, 2009). What the six-dimension procedure adds for historical craft texts is the explicit scenario analysis of a body of literature whose users are philologists and architectural historians rather than curators, and the requirement that every class and relation be warranted by the text itself (Barité, 2018; Beghtol, 1986). OntoClean (Guarino & Welty, 2002) was used informally as a checking device for the class hierarchy: classes whose instances necessarily keep their identity (components, books, and persons) are rigid and form the backbone of the hierarchy, whereas positions, construction methods, and building forms are role-like classes that a member may acquire in a particular entry; they are therefore modeled as sibling subclasses of Building Structure rather than as subclasses of Component (Section 3.3).

Three tools were used: Protégé 5.6.9 (Gennari et al., 2003; Musen, 2015) for class, property, and axiom editing and for running the Pellet 2.2.0 reasoner (Khamparia & Pandey, 2017), which checks the ontology for unsatisfiable classes and inconsistent assertions; Neo4j Community Edition 5.26.16 (Angles, 2012) for storing the populated graph and supporting Cypher queries over paths of arbitrary length; and scikit-learn 1.9.1 (Pedregosa et al., 2011)

and NetworkX 3.7 (Hagberg et al., 2008) for writing the evaluation scripts (Section 2.4) in Python 3.12.

2.3. Workflow

The workflow followed three stages (Figure 1). (i) Ontology construction: This involved requirement analysis and concept extraction from the regulation chapters, class and property design in Protégé, reasoner checks, and revision. (ii) Knowledge graph construction: This entailed population of the ontology from the term records; serialization to RDF; mapping of classes to node labels, object properties to edges, and data properties to node properties; and importation into Neo4j. (iii) Application and evaluation: This involved Cypher querying, visualization, and the quantitative evaluation described in Section 2.4. Each later stage fed back into the earlier one; for example, the classifier confusion matrices of Section 5.3 prompted a re-examination of the Position–Component boundary in the class definitions.

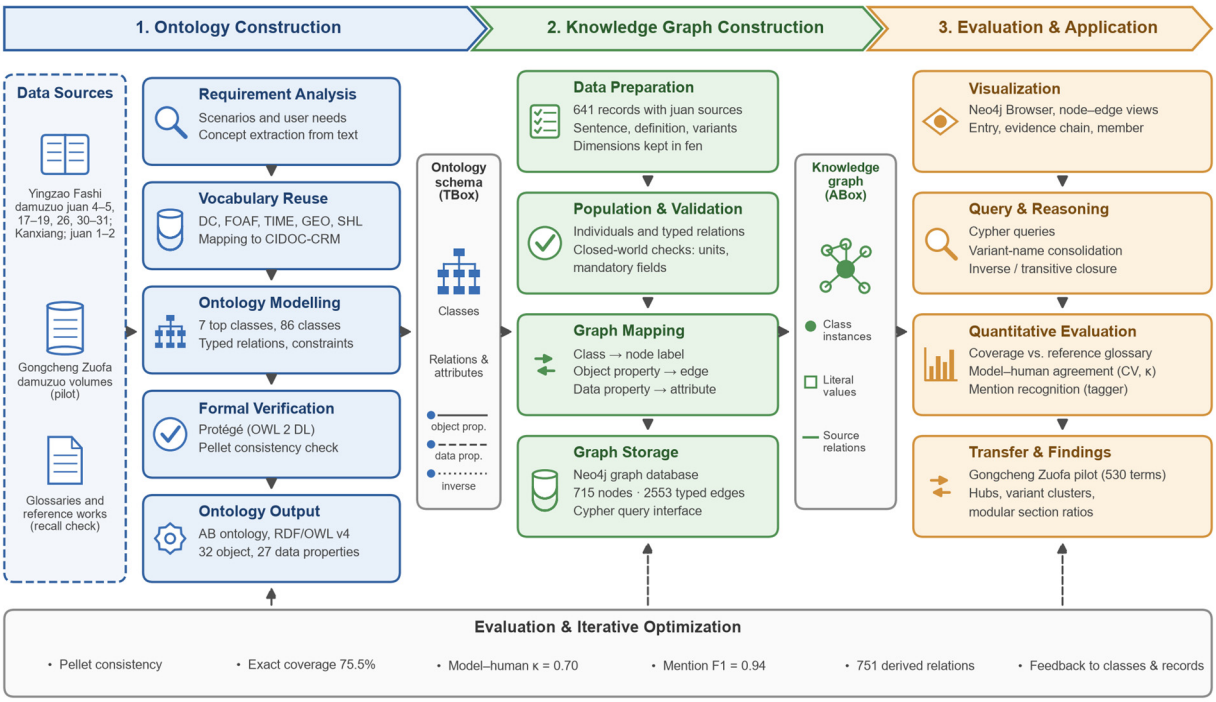


Figure 1. Workflow of the study. Column 1: ontology construction from the source texts to the released RDF/OWL ontology (TBox); column 2: population, validation, mapping, and storage of the knowledge graph in Neo4j (ABox); column 3: visualization, querying and reasoning, quantitative evaluation (Section 2.4), and transfer to the *Gongcheng Zuofa* pilot. The bottom band lists the evaluation results that feed back into the class definitions and records.

2.4. Evaluation Protocol

A reasoner's consistency check evaluates that an ontology is logically coherent, not that it is complete or accurate (Brank et al., 2005; Vrandečić, 2009). Five complementary measures were therefore defined before the analysis and are reported in Section 5. (1) Structural and logical checks: We conducted class and property counts, hierarchy depth, provenance completeness, and the Pellet consistency result. (2) Coverage against an external reference: The terms in the Song column (*Yingzao Fashi*) of the published 'Glossary of commonly used terms in traditional architecture' (古建筑常用名词对照表), from Appendix 6 of Wang's textbook on the construction technology of traditional Chinese architecture (X. Wang, 2019), which aligns Song, Qing, and Jiangnan (*Yingzao Fayuan*) terms, were matched against the 641 records. The main measure is exact coverage—a term is considered covered only when

the string matches after normalization of graphic variants (斗/枓, 板/版, 枋/方, 座/坐). A reference term that merely contains, or is contained in, a recorded term of at least two characters is reported separately as a partial match; because this containment rule is lenient, partial matches are not counted as covered. (3) Model–human agreement of the manual classification: Supervised classifiers (logistic regression, linear support-vector machine, random forest) were trained to predict the manually assigned record category of each term from character n-grams of the term and of its masked context sentences under five-fold stratified cross-validation; accuracy, macro-F1, and Cohen's kappa (Artstein & Poesio, 2008; Cohen, 1960) between prediction and manual label are reported, with the confusion matrix as a diagnostic of class boundaries. Because a second independent human annotator was not available, this model–human agreement is used as a reproducibility proxy and is reported as such (Hripcsak & Rothschild, 2005). (4) Distantly supervised recognition of term mentions: The 641 terms were projected onto the source text by longest-match annotation (Mintz et al., 2009), a character-level sequence tagger (logistic regression on a $\pm$2-character window) was trained under five-fold cross-validation over sentences, and entity-level precision, recall, and F1 were computed separately for mentions of terms unseen in the training folds. (5) Reasoning yield and transfer: The number of assertions added by inverse and transitive reasoning and the behavior of the class scheme on the Qing *Gongcheng Zuofa* were evaluated through both a pilot ontology (Section 5.6) and classifiers trained on one text and tested on the other.

3. Ontology Model Design

Following the procedure of Section 2.2, this section presents the ontology by dimension: scenario (Section 3.1), reuse (Section 3.2), entities (Section 3.3), relations and constraints (Section 3.4), and formalization (Section 3.5). The ontology is released as RDF/OWL in the Zenodo deposit described in the Data Availability Statement (Qiao et al., 2026).

3.1. Scenario: Scope and Requirements

Two scoping decisions define the scenario, the first of which concerns the categories of text. Traditional architectural books survive as transmitted editions (state-issued standards such as the *Yingzao Fashi* and the *Gongcheng Zuofa*; local and departmental craft regulations; private writings such as the *Yingzao Fayuan* and the *Ziren Yizhi*; drawings and illustrated catalogs) or only as fragments quoted in later works, as with Yu Hao's *Mujing*, which is known solely through passages in Shen Kuo's *Mengxi Bitan* (Shen, ca. 1088). An additional group of texts are mentioned by title alone and await verification. The ontology therefore distinguishes surviving, lost, recovered, and authenticated books so that provenance and evidential status can be stated for each text (Section 3.3).

The second decision concerns the practical requirements the model must fulfill. Three were identified with the philologists and architectural historians consulted during requirement analysis (Zhang et al., 2025): (i) retrieval of every regulation entry, term, and variant name with its *juan*-level source so that a reading of the text can be checked against the edition; (ii) fine-grained reorganization of each entry into its components, positions, methods, dimensions, and modular grade so that the internal structure of the regulations becomes queryable; and (iii) reasoning and analysis across entries, such as reconstructing the assembly order of a bracket set from typed relations or comparing the same term across texts. Requirement (i) governs the data properties of Section 3.4, requirement (ii) the Building Structure subclasses of Section 3.3, and requirement (iii) the typed object properties and the evaluation of Section 5.

3.2. Reuse: Existing Vocabularies and the Position of CIDOC-CRM

Reusing established vocabularies lowers the construction cost and improves interoperability, so the candidate vocabularies were examined before any class was created. CIDOC-CRM (Doerr, 2003) was the obvious candidate and was examined in detail. Its classes for persons (E21), places (E53), time-spans (E52), documents (E31), and dimensions (E54, E58) cover the general part of the present model, and mapping to them is reported in Table 2 so that the dataset can be exported to CRM-based aggregators. CIDOC-CRM was nevertheless not adopted as the backbone for two reasons. First, its description is event-centered and object-centered: a dimension (E54) is a measurement of an existing object, whereas the dimensions in the *Yingzao Fashi* are prescriptions in *fen* that exist before any object does. Second, the regulation entry itself has no close counterpart; E29 Design or Procedure is the nearest class but does not distinguish the internal structure (variant names, dimensions, position, and method) that the text gives every entry. Rather than extend CIDOC-CRM with a dozen specialized classes, the study reused the lighter-weight vocabularies that fit each general class directly and built a small domain vocabulary for the rest.

Table 2. Correspondence between the core classes of the Architectural Books (AB) ontology and CIDOC-CRM.

AB Class	Closest CIDOC-CRM Class	Relation	Note
AB:ArchitecturalBook	E31 Document/E33 Linguistic Object	narrower	A book as text; editions are handled by AB:version
AB:BuildingRule	E29 Design or Procedure	narrower	CRM has no class for a normative entry with internal structure
AB:Component	E22 Human-Made Object (type)	related	AB components are prescribed member types, not individual objects
AB:Dimension/AB:ModularGrade	E54 Dimension/E58 Measurement Unit	related	CRM dimensions are measurements of objects; AB dimensions are prescriptions in fen tied to a grade
AB:Position, AB:ConstructionMethod, AB:BuildingForm	E55 Type	broader	Role-like categories; CRM would express them as types
FOAF:Person	E21 Person	equivalent	—
GEO:SpatialThing (SHL:Place, SHL:Region)	E53 Place	equivalent	—
TIME:TemporalEntity (SHL:Dynasty, SHL:Reign)	E52 Time-Span/E4 Period	equivalent	Dynasties and reigns as named periods
AB:Sources	E31 Document	equivalent	Documentary warrant for a statement

The ontologies used in this study, therefore, are widely used—DC (Dublin Core Metadata Initiative, n.d.), Friend of a Friend (FOAF; Brickley & Miller, 2014), TIME (Cox & Little, 2017), GEO (W3C Semantic Web Interest Group, n.d.), and the Shanghai Library Ontology (SHL; Shanghai Library, n.d.)—as they provide concrete and fine-grained structures. Among them, the Shanghai Library Ontology's treatment of Chinese historical chronology and administrative geography suits the materials of this study better than comparable international vocabularies. In addition, a novel Architectural Books (AB) vocabulary was constructed for this study to supplement the existing frameworks. This treatment keeps the ontology consistent with mature standards while adapting it to the specialized semantic detail required by the architectural book domain.

3.3. Entities: Design of Classes and Hierarchical Structure

The seven top-level classes were derived from the text rather than from a prior framework. During requirement analysis, the six questions of digital narrative—when, where, who, why, what, and how—were used only as a completeness checklist to make sure that no element of knowledge in the books had been overlooked. Time and place attach to books, entries, persons, and sources; persons appear as compilers, annotators, and craftsmen. The source literature answers 'why' a statement can be trusted, while the book and its regulation entries answer 'how' and 'what'. The checklist did not add a modeling layer. The knowledge system structure of these traditional architectural books is divided into seven main entity classes: Architectural Book (AB:ArchitecturalBook), Building Regulation (AB:BuildingRule), Building Structure (AB:BuildingStructure), Person (FOAF:Person), Space (GEO:SpatialThing), Time (TIME:TemporalEntity), and Source Literature (AB:Sources). Each entity is further divided into subclasses. For example, Architectural Book includes the subclasses Surviving Architectural Book (AB:SurvivingBook), Lost Architectural Book (AB:LostBook), Recovered Lost Architectural Book (AB:CollectedLostBook), and Authenticated Architectural Book (AB:AuthenticatedBook). The specific hierarchical relations are shown in Figure 2. The hierarchy has 86 named classes in three levels: seven top-level classes, 14 second-level classes, and 65 third-level classes. Building Structure accounts for 68 of them, namely, its six subclasses (Figure 2) and the 62 third-level classes below them: 18 component families, 27 construction-method families, and 13 position families (*gong*, *dou*, *ang*, beam, column, purlin, rafter, and so on), three building forms, and the measurement unit under Dimension; Modular Grade is not subdivided. The subdivision by member family keeps the class of a term aligned with the chapter structure of the source.

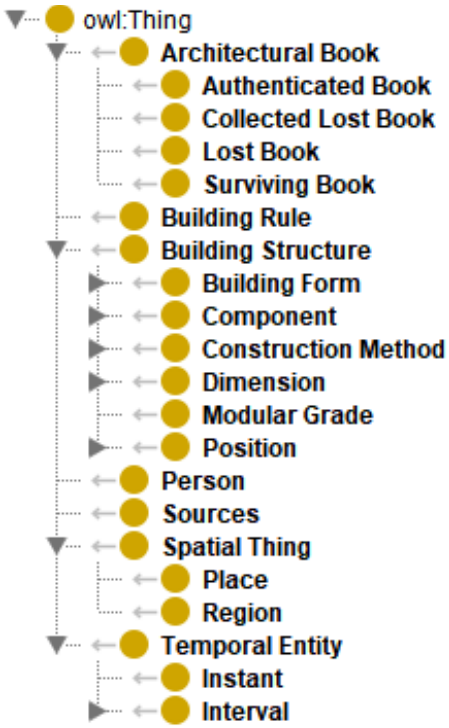


Figure 2. Traditional Chinese architectural book ontology and hierarchical structure. OWL, Web Ontology Language.

In creating dedicated classes for this ontology, the architectural book forms the core component, and all other entity elements are associated with this central class. A building regulation is a specific regulation entry recorded in an architectural book, such as the entries 'the rule for making *gong*', 'the rule for making *ang*', and 'the rule for making *dou*' in the

damuzuo chapters of the *Yingzao Fashi*. Building structure describes the internal structure of each regulation entry, including the component (AB:Component), position (AB:Position), construction method (AB:ConstructionMethod), dimension (AB:Dimension), modular grade (AB:ModularGrade), and building form (AB:BuildingForm). These six subclasses are the classes of the ontology. They are not the eight record categories reported in Table 1, under which the terms were collected; six categories coincide with the six subclasses, whereas the records of the two remaining categories, material quota and quantity operator, are typed under Dimension, Component, or Modular Grade according to what they denote (Table 1, last column). The source literature has many facets, including the recorded sources of architectural books, regulation entries, and biographies, and provides the basis for documentary verification and for constructing chains of descriptive evidence.

It should be noted that the division of the building structure subclasses is not a transplant of a modern structural engineering taxonomy but is taken from the textual conventions of the source literature itself. The entries in the regulation chapters of the *Yingzao Fashi* give, in order, the permitted variant names of a component, its dimensions in *fen*, its position, the components it intersects, and its working method; the subclass division of the ontology corresponds to this convention so that the design of the categories can be verified by the source literature itself. This practice is consistent with the documentary warrant discussion in the knowledge organization literature (Barité, 2018; Beghtol, 1986): the categories of a classification scheme are taken from the domain literature itself, and their rationality can then be tested by returning to that literature. Similarly, the determination of equivalence relations is based on the variant names recorded in the *Kanxiang* section on variant names and in the interlinear notes under the regulation entries; the form used in the main text of the regulation chapters is treated as the formal term and the forms recorded in the terminology chapters as variant names so that the distinction between formal and variant names becomes a verifiable textual fact.

From the perspective of general classes, persons in traditional architectural books primarily appear as compilers, annotators, and master craftsmen. Given that an individual, as a social agent, is necessarily constrained by space and time, a person may be associated with architectural books, space, time, and the source literature. Instances of spatial and temporal entities attach to and are associated with entities such as architectural books, building regulations, persons, and the source literature, constituting their spatiotemporal components. Two subclasses exist within the spatial element, region (SHL:Region) and place (SHL:Place), which describe the area and point features of physical space, respectively. Temporal entities are divided into instant (TIME:Instant) and interval (TIME:Interval) entities, describing the point and span features of time, respectively. Given the cultural particularity of temporal description in ancient China, intervals are further subdivided into reign title (SHL:Reign), dynasty (SHL:Dynasty), and specific time interval (TIME:ProperInterval).

3.4. Relations and Constraints: Design of the Property Tables

Relations exist pervasively among entity concepts and are expressed in an ontology as object properties and data properties. In the analysis of this study, relations among entities constitute object properties, and the inherent characteristics of entities constitute data properties. In addition, the concept of constraint in an ontology is made concrete by defining the domain and range of properties, thereby specifying the intension and extension of a relation. The object properties defined in this study are shown in Table 3, and the data properties are reported in Table 4.

Table 3. Object properties of traditional Chinese architectural books.

Property	Description	Domain	Range
DC:creator	Compiler/author	AB:ArchitecturalBook; AB:BuildingRule; AB:Sources	FOAF:Person
AB:relevantPerson	Related person	AB:ArchitecturalBook; AB:BuildingRule	FOAF:Person
AB:subordinate	Belongs to source	AB:ArchitecturalBook; AB:BuildingRule; FOAF:Person	AB:Sources
SHL:temporal	Dynasty	AB:ArchitecturalBook; AB:BuildingRule; AB:Sources; FOAF:Person	SHL:Dynasty
AB:relevantRule	Related regulation	AB:ArchitecturalBook; AB:BuildingRule; FOAF:Person	AB:BuildingRule
AB:recordRule	Records regulation	AB:ArchitecturalBook	AB:BuildingRule
AB:includeStructure	Includes structure	AB:BuildingRule	AB:BuildingStructure
AB:createdTime	Date compiled	AB:ArchitecturalBook; AB:BuildingRule; AB:Sources	TIME:TemporalEntity
AB:createdPlace	Place compiled	AB:ArchitecturalBook; AB:BuildingRule; AB:Sources	GEO:SpatialThing
AB:collectionPlace	Place held	AB:ArchitecturalBook; AB:Sources	GEO:SpatialThing
SHL:beginYear	Begins at	TIME:Interval	TIME:Instant
SHL:endYear	Ends at	TIME:Interval	TIME:Instant
AB:belongsToRegion	Belongs to region	SHL:Place	SHL:Region
SHL:birthDay	Year of birth	FOAF:Person	TIME:Instant
SHL:deathday	Year of death	FOAF:Person	TIME:Instant
SHL:nativePlace	Native place	FOAF:Person	SHL:Place
AB:hasVariantName	Variant name	AB:BuildingStructure	AB:BuildingStructure
AB:hasModularGrade	Belongs to modular grade	AB:BuildingStructure	AB:BuildingStructure
AB:hasLocus	Position	AB:BuildingStructure	AB:BuildingStructure
AB:hasConstructionMethod	Construction method used	AB:BuildingStructure	AB:BuildingStructure
AB:usedInBuildingForm	Permitted building form	AB:BuildingStructure	AB:BuildingStructure
AB:hasUnit	Unit of measure	AB:BuildingStructure	AB:BuildingStructure
AB:isPartOf	Part-of relation	AB:BuildingStructure	AB:BuildingStructure
AB:contains	Contains relation	AB:BuildingStructure	AB:BuildingStructure
AB:joinsWith	Joins with	AB:BuildingStructure	AB:BuildingStructure
AB:supports	Supports	AB:BuildingStructure	AB:BuildingStructure
AB:supportedBy	Supported by	AB:BuildingStructure	AB:BuildingStructure
AB:above/AB:below	Vertical position relation	AB:BuildingStructure	AB:BuildingStructure
AB:insertedInto/AB:receives	Insertion relation	AB:BuildingStructure	AB:BuildingStructure
AB:continueWith	Sequence relation	AB:BuildingStructure	AB:BuildingStructure

Note: DC, Dublin Core; AB, Architectural Book vocabulary; SHL, Shanghai Library Ontology; FOAF, Friend of a Friend; TIME, Time Ontology; GEO, Geographic Vocabulary.

Table 4. Data properties of traditional Chinese architectural books.

Property	Description	Domain	Range
DC:title	Name	OWL:Thing	RDFS:Literal
DC:description	Description	OWL:Thing	RDFS:Literal
DC:type	Type	OWL:Thing	RDFS:Literal
DC:identifier	Resource identifier	OWL:Thing	RDFS:Literal
DC:image	Image	OWL:Thing	XSD:anyURI
AB:content	Content	AB:ArchitecturalBook; AB:BuildingRule; AB:Sources; AB:BuildingStructure	RDFS:Literal
AB:volumeNumber	Number of juan	AB:ArchitecturalBook	RDFS:Literal
AB:juan	Juan-level provenance	AB:BuildingRule; AB:BuildingStructure	RDFS:Literal
AB:prologue	Preface	AB:ArchitecturalBook	RDFS:Literal
AB:epilogue	Postscript	AB:ArchitecturalBook	RDFS:Literal
AB:version	Edition	AB:ArchitecturalBook; AB:BuildingRule; AB:Sources	RDFS:Literal
AB:originalText	Original text	AB:BuildingRule; AB:BuildingStructure	RDFS:Literal
AB:translatedText	Translation	AB:BuildingRule; AB:BuildingStructure	RDFS:Literal
AB:definition	Definition	AB:BuildingStructure	RDFS:Literal
AB:prompt	Interlinear note	AB:BuildingStructure	RDFS:Literal
AB:termPinyin	Pinyin	AB:BuildingStructure	RDFS:Literal
AB:termEnglish	English heading	AB:BuildingStructure	RDFS:Literal
AB:widthInFen	Width (fen°)	AB:BuildingStructure	XSD:decimal
AB:thicknessInFen	Thickness (fen°)	AB:BuildingStructure	XSD:decimal
AB:lengthInFen	Length (fen°)	AB:BuildingStructure	XSD:decimal
SHL:courtesyName	Courtesy name	FOAF:Person	RDFS:Literal
SHL:pseudonym	Pseudonym	FOAF:Person	RDFS:Literal
TIME:inXSDDateTime	Date	TIME:TemporalEntity	RDFS:Literal
SHL:city	City	SHL:Region	RDFS:Literal
SHL:county	County	SHL:Region	RDFS:Literal
SHL:province	Province	SHL:Region	RDFS:Literal
SHL:country	Country	SHL:Region	RDFS:Literal

Note: RDFS, Resource Description Framework Schema; XSD, XML Schema Definition; XML, Extensible Markup Language. All dimension properties measured in *fen*° are governed by the unit-of-measure constraint.

In the relation design, this study typifies the associative relations among building structures. Hierarchical relations are borne by two mechanisms separately: the containment relation among classes forms the classification tree, while the whole–part relation among term instances is expressed by the two properties isPartOf (AB:isPartOf) and contains (AB:contains). These two are strictly distinguished because a part does not become the same type as the whole merely by belonging to it. Associative relations are not merged into

a vague 'related' category but, following the constructional relations recorded in the text of the *Yingzao Fashi*, are subdivided into joining (AB:joinsWith), supporting (AB:supports), supported-by (AB:supportedBy), vertical (AB:above/AB:below), insertion (AB:insertedInto/AB:receives), and sequence (AB:continueWith) types, with the corresponding inverse relations declared. The reason for this treatment is that although labeling a single vague related relation between two components is true, it cannot answer specific questions; only after distinguishing which one is inserted into and which supports the other can the system reconstruct the assembly order of the components.

Several domain and range declarations in Table 3 are deliberately broad. The structural properties (hasVariantName, hasLocus, hasConstructionMethod, hasModularGrade, hasUnit, usedInBuildingForm, and the part-of and constructional relations) are declared based on Building Structure rather than Component for two reasons. First, in OWL, a domain or range axiom is an inference rule, not a check; declaring hasLocus with the domain Component would not reject a position attached to a construction method but would silently classify that method as a component. Narrow OWL domains would therefore produce spurious inferences. Second, the text itself attaches these relations to more than one structural subclass: a construction method such as '*zucai* construction' has a modular grade, a position such as 'the outer projection' has a variant name, and a quantity operator attaches to dimensions. Properties whose subjects are genuinely heterogeneous (creator, temporal, createdTime, createdPlace, subordinate, relevantPerson, relevantRule, collectionPlace) use an explicit union of classes as their domain, as listed in Table 3, which is the standard OWL idiom for a shared property. The finer restrictions that the text does impose—for example, that a *fen* dimension must be attached to a member or grade and must carry the unit *fen*, or that every term record must have a formal term, a definition or original sentence, and a *juan*-level source—are enforced as closed-world checks on the populated data (implemented as validation scripts equivalent to SHACL shapes; Knublauch & Kontokostas, 2017) rather than as open-world OWL axioms. Records failing a check are flagged rather than silently completed; the part-of cycle and the seven part-of/variant-name conflicts reported in Section 5.5 are such flags. Table 5 lists the main modeling decisions, the alternatives considered, and the check performed on each.

Table 5. Principal modeling decisions, alternatives considered, and the check applied to each.

Decision	Alternative Considered	Reason for the Decision	Check Applied
Regulation entry (BuildingRule) as unit of description	Component as unit (artifact-centered metadata)	The text prescribes members before objects exist; entry structure (variant names, fen, position, method) is lost if the member is the unit	Every BuildingStructure instance reachable from a BuildingRule via includeStructure or isPartOf
Dimensions stored in fen with hasModularGrade	Absolute dimensions in cun/chi per grade	Eight grades would multiply every value by eight and hide the derivation	Unit-of-measure check: widthInFen/thicknessInFen/lengthInFen require hasUnit = fen
Typed constructional relations (joinsWith, supports, insertedInto, above, continueWith) with inverses	Single 'relatedTo' relation	Only typed relations answer assembly order and load-path queries	Inverse completion by reasoner; counts reported in Section 5.5
Variant names as hasVariantName between individuals	SKOS altLabel on one concept	Variant names are themselves attested terms with their own juan provenance and, sometimes, their own entries	Symmetric closure; cluster sizes reported in Section 5.5; SKOS export planned

Table 5. *Cont.*

Decision	Alternative Considered	Reason for the Decision	Check Applied
Position, ConstructionMethod, BuildingForm as siblings of Component under BuildingStructure	Subclasses of Component	Role-like, non-rigid categories (OntoClean) that a member acquires in an entry	Confusion analysis of Section 5.3 used to re-examine boundaries
Broad domains on BuildingStructure; union domains for shared properties	Narrow per-subclass domains	OWL domains infer rather than constrain; narrow domains would misclassify	Closed-world validation scripts (SHACL-equivalent) on the populated data
Mandatory fields: formal term, definition or original text, juan source	Default values	Missing evidence should be visible, not filled	641 of 641 pass

3.5. Formalization and Logical Verification

Once the classes and properties were fixed, the ontology was formalized in Protégé and checked with the Pellet reasoner (Khamparia & Pandey, 2017). The reasoner found no unsatisfiable class and no inconsistent assertion in the populated ontology (86 named classes, 32 object properties, 27 data properties, 715 named individuals, 9985 triples); it also output the inverse relations discussed in Section 5.5. A consistency check establishes only that the model is logically coherent, not that it is complete or accurate; those issues are addressed by the measures described in Section 2.4 and reported in Section 5. Figure 3 visualizes the ontology model.

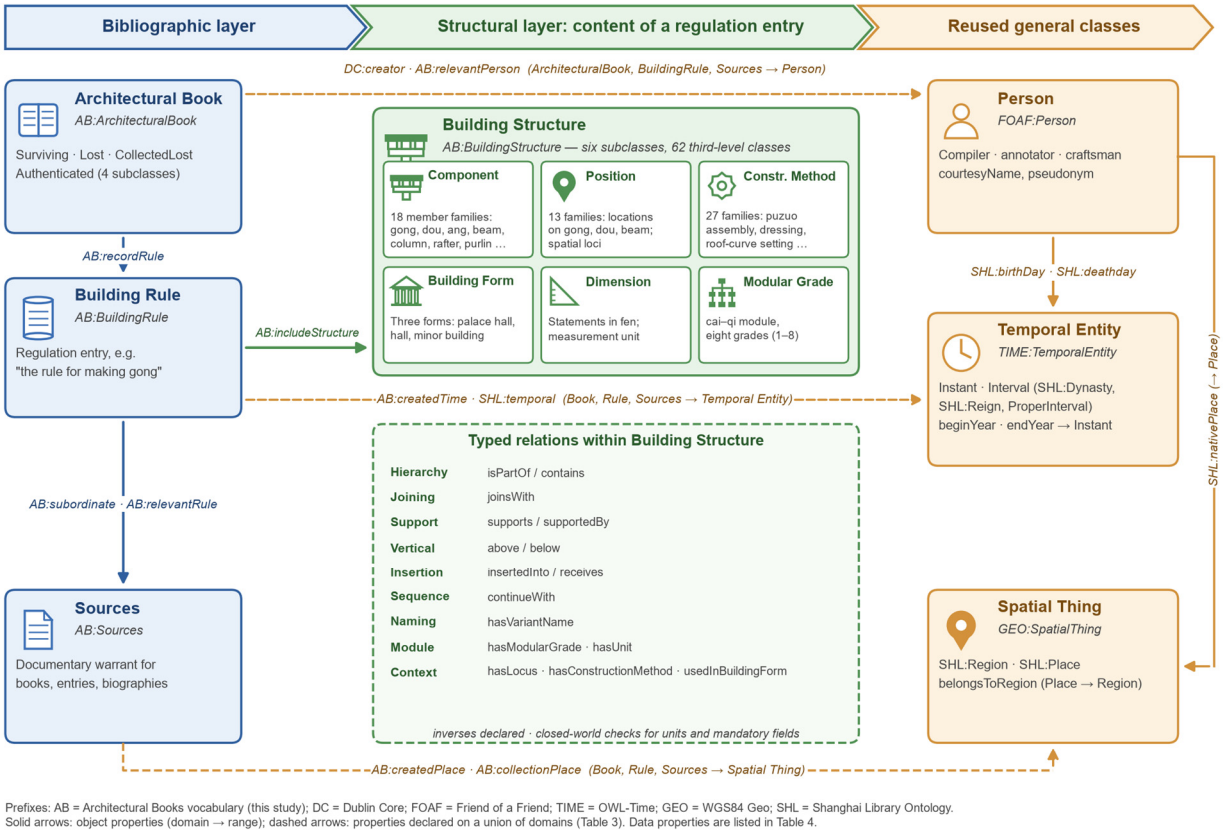


Figure 3. Traditional Chinese architectural book ontology model: the bibliographic layer (Architectural Book, Building Rule, Sources), the structural layer that holds the content of a regulation entry (Building Structure with its six subclasses; their 62 third-level classes are collapsed), and the reused general classes (Person, Temporal Entity, Spatial Thing). The typed relations within Building Structure are listed in the dashed box.

4. Knowledge Graph Construction and Application

4.1. Knowledge Mapping and Storage

The populated ontology was exported in RDF format and loaded into Neo4j with rule-based mapping. Formally, the knowledge graph is written as $G = (V, E)$, where V is the set of nodes and E the set of edges. A node is a pair $v = (L, P)$, where L is the set of its labels (the classes of the entity) and P the set of its properties, the key–value pairs that describe the entity. An edge is a triple $e = (s, r, o)$, where r is the relation type and s and o are the start and end nodes that the edge connects. Under this mapping (Figure 4), each ontology class and subclass becomes a node label, and a node carries the labels of all its superclasses so that a query for Component also returns every *gong* or *dou*; each object property becomes a typed edge; each data property becomes a node property. The graph holds 715 nodes, 2553 typed edges, and 4546 node properties and was verified against the RDF by counting nodes per label and edges per type after import.

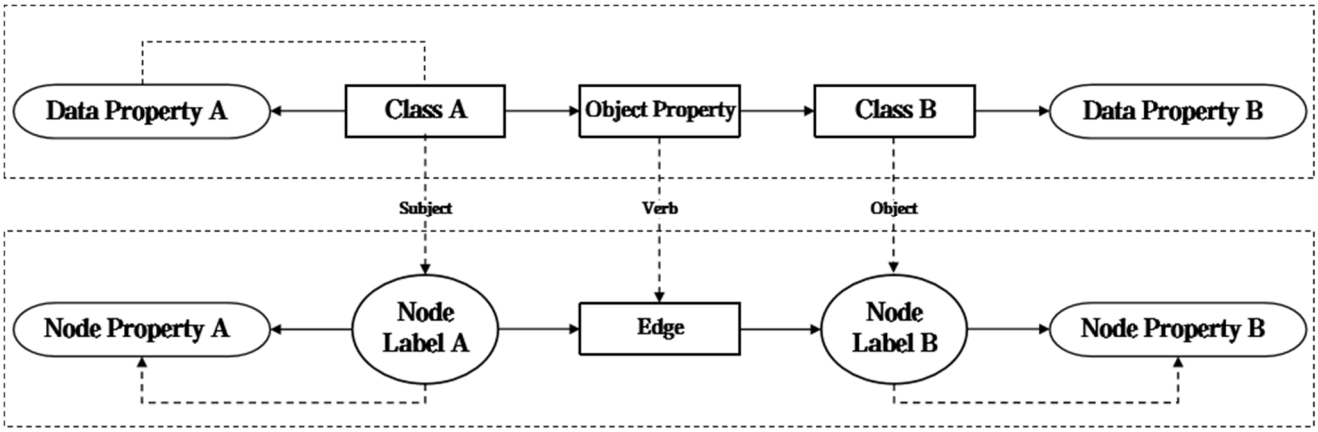


Figure 4. Ontology-to-graph database mapping process.

4.2. Worked Queries and Reasoning Results

This section reports what representative Cypher queries return. Table 6 lists five queries with their result counts; the full query texts are included in the Zenodo deposit (Qiao et al., 2026).

Table 6. Representative Cypher queries over the *Yingzao Fashi* graph and their results.

Question	Cypher (Abbreviated)	Result
Q1. Which components are governed by the entry 'the rule for making gong'?	MATCH (r:BuildingRule {title:'Gong'})-[:includeStructure]->(c:GongComponent) RETURN c	5 components (huagong, nidaogong, guazigong, linggong, mangong); see Figure 5
Q2. Which members have variant names, and how many labels collapse?	MATCH (a)-[:hasVariantName]->(b) RETURN a, collect(b)	57 formal terms carry 96 variant-name assertions; 127 attested labels resolve to 33 concept clusters
Q3. What does the huagong join, support, contain, or insert into?	MATCH (h {title:'Huagong'})-[r:joinsWith|supports|contains|isPartOf|insertedInto]-(x) RETURN type(r), x	38 typed structural links, the most of any member, to ludou, jiaohudou, nidaogong, and others
Q4. Which members are reached by following isPartOf transitively from a bracket-set element?	MATCH p = (x)-[:isPartOf*1..4]->(y) RETURN y	128 part-of statements not asserted in the text are derived; 1 cycle and 7 part-of/variant conflicts are exposed (Section 5.5)
Q5. What evidence links the lost Mujing to a datable source?	MATCH p = (m {title:'Mujing'})-[:subordinate|creator|temporal*1..3]-(x) RETURN p	Mujing → Mengxi Bitan → Shen Kuo → Northern Song; see Figure 6

For traditional architectural book resources, personalized queries can be carried out according to specific categories within the literature, periods, or regulation entries. For example, as shown in Figure 5, the '*gong*' entry in *juan* 4 of the *Yingzao Fashi*, *Damuzuo* Regulations I, records that 'there are five rules for making *gong*'; the graph accordingly returns the five components *huagong*, *nidaogong*, *guazigong*, *linggong*, and *mangong*, each linked to the '*gong*' regulation entry node by a containment relation. Such a query is presented alongside the classificatory information originally dispersed across the main text of the entry and its interlinear notes, making it easy for researchers to grasp the complete range of components governed by a given regulation entry. The value of the query lies less in the display than in what it makes checkable: each of the five components carries its *juan*-level source and its original sentence so that a reader can confirm from the graph alone that the five rules are those of *juan* 4 and not of the later glossaries.

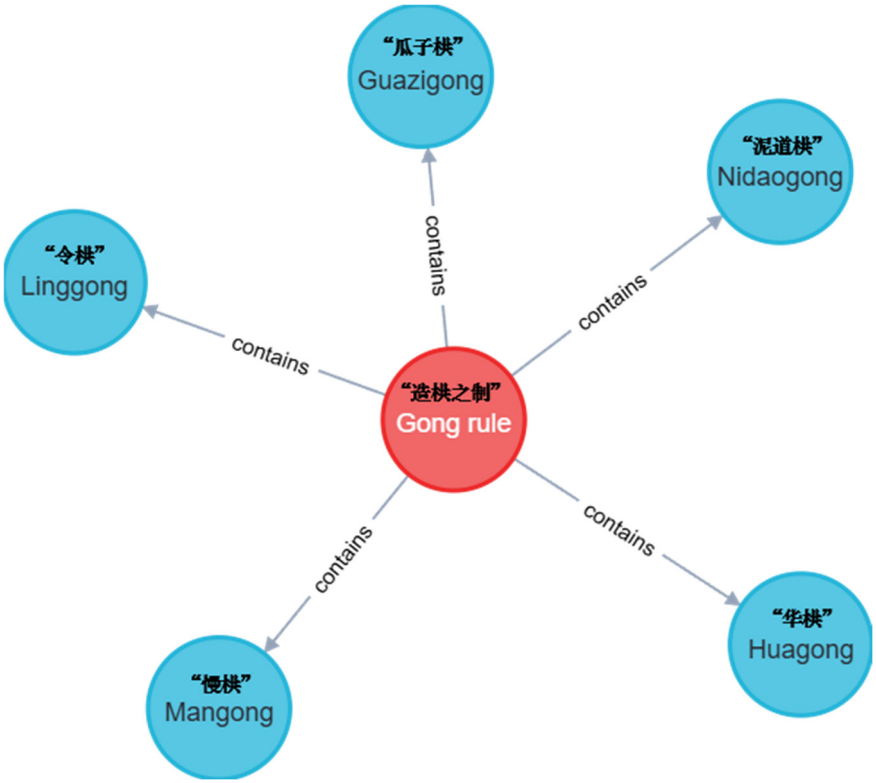


Figure 5. Query of the 'five rules for making *gong*' in the *damuzuo* chapters of the *Yingzao Fashi*. (The central node is the regulation entry '*gong*'; around it are the five component nodes *huagong*, *nidaogong*, *guazigong*, *linggong*, and *mangong*, each connected by a containment relation.).

The second use is the consolidation of variant names. Table 7 reports the before-and-after picture for the members with the most variant names. Before modeling, the labels in the left column are scattered across the *Kanxiang*, the two General Glossary *juan*, and the interlinear notes of *juan* 4 and 5, and a reader finding *juantou* or *tiaotou* in a labor-quota entry has no textual signal that both denote the *huagong*. After modeling, each label is an individual with its own provenance, linked to the formal term by hasVariantName, so that a query for the formal term returns every occurrence under every label, and a query for a variant name returns the formal entry. Across the dataset, 57 formal terms carry 96 variant-name assertions, and 127 attested labels resolve to 33 concept clusters (Section 5.5).

Table 7. Variant-name consolidation for the formal terms with the most attested variants (from the released RDF, version 4).

Formal Term	Class	Variant Names Attested in the Text (hasVariantName)	Source of the Formal Term
Dong (栋), ridge purlin	Purlin component	Fen (棼), Fu (桴), Yin (檃), Meng (甍), Ji (极), Tuan (槫), Lin (檩), Mian (榠)	Kanxiang (variant-name section)
Ang (昂)	Ang component	Jian (戬), Feiang (飞昂), Fei'ang (飞棉; identifier Feiju), Yingang (英昂), Xiejiao (斜角), Xia'ang (下昂)	juan 4
Chuan (椽), rafter	Rafter component	Jue (椑), Cui (榱), Liao (橑), Fangchuan (方椽), Yuanchuan (圆椽)	Kanxiang
Chashou (叉手), inverted V-brace	Inclined-strut component	Xiezhu (斜柱), Wu (梧), Wu (迕), Zhitang (枝橖)	Kanxiang
Dou (枓), bearing block	Dou component	Jie (楬), Er (栭), Lu (栌), Ta (楁)	Memorial presenting the Yingzao Fashi
Guleng (觚棱), hip rafter	Hip-rafter-beam component	Yangma (阳马), Quejiao (阙角), Jiaoliang (角梁), Liangmo (梁抹)	juan 5
Gong (栱), bracket arm	Gong component	Ji (槉), Luan (栾), Quji (曲枅); Bian (閞) and Bo (槫) are recorded as separate entries pointing to gong	Kanxiang
Huagong (华栱)	Gong component	Miaogong (杪栱), Juantou (卷头), Tiaotou (跳头)	juan 4 (interlinear note)

From the perspective of textual verification, the knowledge graph can provide a reliable chain of descriptive evidence for architectural books that require checking. The graph lists sources systematically to verify doubts or uncertainties regarding authorship, title, content, and date of compilation so that specific information about an architectural book can be clarified relatively objectively and accurately. For example, Yu Hao's *Mujing* was lost after the Song dynasty, and its content survives only in several passages quoted in Shen Kuo's *Mengxi Bitan*. As shown in Figure 6, the graph links the *Mujing* to the *Mengxi Bitan* by a descriptive relation, links the *Mengxi Bitan* to Shen Kuo by an authorship relation, and links both to the Northern Song by a dynasty relation, thereby forming a verifiable chain of evidence. Along this chain, it can be observed that the *Mujing* divides a building into three parts—upper, middle, and lower—so that 'the house has three divisions' and takes the length of a rafter as its proportional base, whereas the *Yingzao Fashi* states that 'in all building construction, the *cai* is the fundamental module' and takes the *cai–qi* module as its base. Although both originated in the Song dynasty and are both architectural texts, their dimensional bases are not the same, and they should therefore not be regarded as two versions of a single institutional system. This is evidence assembly, not inference; rather than deriving the conclusion, the graph assembles the evidence chain (lost book → quoting source → author → dynasty) that a scholar needs in order to draw their conclusion, and it does so from assertions each of which carries a source. The genuinely inferred statements in the graph are the inverse and transitive relations quantified in Section 5.5.

In this study, entry-level fine-grained analysis of specific regulation entries was performed so as to construct a knowledge graph of traditional architectural books from the professional standpoint of building craftsmanship. Figure 7 shows a fine-grained representation of the *huagong* item in the '*gong*' entry of *juan* 4 of the *Yingzao Fashi*. This presents not only external knowledge features such as the documentary source and compiler but also internal building structure features.

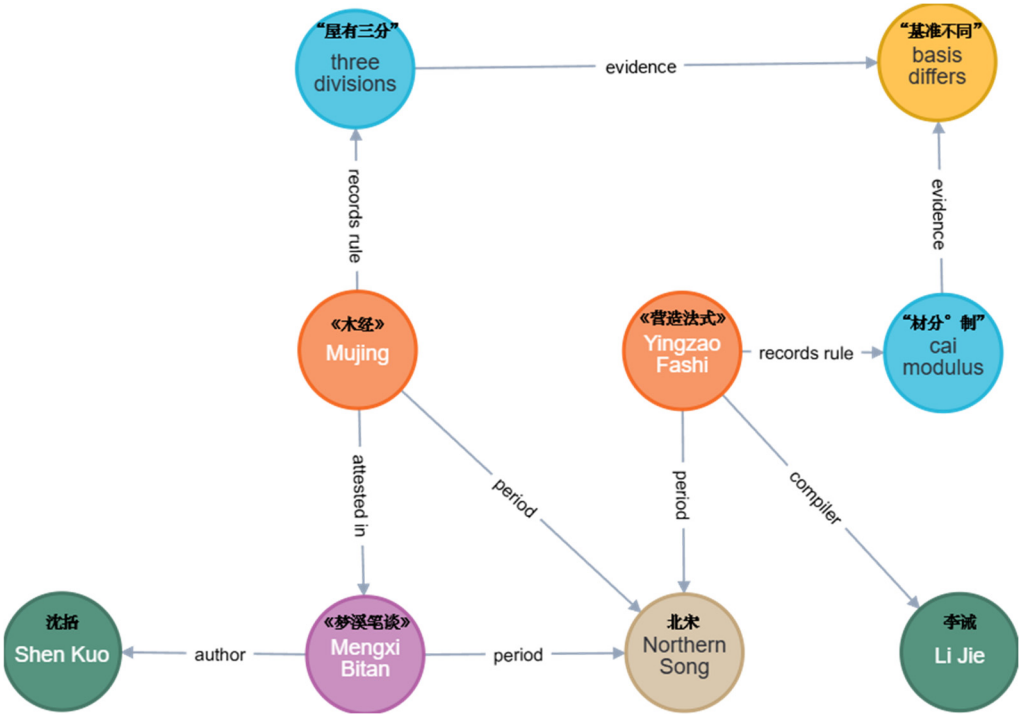


Figure 6. Evidence chain for the lost *Mujing* assembled from the graph. (Nodes include the *Mujing*, *Mengxi Bitan*, *Yingzao Fashi*, Shen Kuo, Li Jie, and the Northern Song; edges include descriptive, authorship, and dynasty relations.) The node ‘basis differs’ and its two edges labelled ‘evidence’ record the conclusion that a scholar draws from the two rules; they were entered by hand and are not derived by the graph (Section 4.2).

From the angle of external knowledge features such as documentary sources and compilers, the *huagong* appears in the ‘*gong*’ entry of *juan* 4, *Damuzuo* Regulations I, of the *Yingzao Fashi* and is the first of the ‘five rules for making *gong*’; its variant names are *miaogong*, *juantou*, and *tiaotou*, all recorded in the interlinear notes under that entry. With respect to internal building structure, the *huagong* is a *zucai gong*, 21 *fen* wide, 10 *fen* thick, and 72 *fen* long, projecting outward in the external and internal projection of the bracket set and forming insertion and support relations with the *ludou* and *jiaohudou*, respectively. This study marks its dimensions in ‘*fen*’ and represents the derivation of its absolute size through the modular-grade relation: the *cai* is divided into eight grades, with the *cai* 15 *fen* wide and 10 *fen* thick and the *qi* 6 *fen* wide and 4 *fen* thick. For the same component under different modular grades, the *fen* value is unchanged while the absolute size changes accordingly. Thus, a first-grade *huagong* (1 *fen* = 0.6 *cun*) is 43.2 *cun* long and a sixth-grade one (1 *fen* = 0.4 *cun*) 28.8 *cun*, and both are the same graph node. This fine-grained dimensional representation shows that the dimensions recorded in architectural books are derived rather than absolute and that the derivation of a value is as much knowledge to be organized as the value itself. The data properties of the graph also show that the method and permitted form of the *huagong* are constrained by the regulation text, such as the correspondence between *zucai* construction and the palace-pavilion form, reflecting the institutional feature by which a state-issued building standard prescribes the use of materials according to hierarchical grade.



Figure 7. Fine-grained representation of a regulation entry: the *huagong* item in the ‘*gong*’ entry, *juan* 4 of the *Yingzao Fashi* (the central node is the *huagong*; around it are its variant names *miaogong*, *juantou*, and *tiaotou*; dimension nodes 21 *fen* wide, 10 *fen* thick, and 72 *fen* long; position nodes external and internal projection; the construction-method node *zucai* construction; the form node palace-pavilion; and related component nodes such as *ludou* and *jiaohudou*).

5. Evaluation and Dataset Analysis

This section reports the five measures defined in Section 2.4. All numbers were computed from the released RDF (version 4), the 641 term records, and the source text with the scripts deposited at Zenodo (Qiao et al., 2026); the classifiers and the sequence tagger were implemented with scikit-learn 1.9 (Pedregosa et al., 2011) and the graph statistics with NetworkX (Hagberg et al., 2008).

5.1. Structural Metrics and Logical Consistency

Table 8 presents the structural profile of the *Yingzao Fashi* ontology and, for comparison, of the *Gongcheng Zuofa* pilot ontology described in Section 5.6. The *Yingzao Fashi* graph contains 715 individuals, of which 641 are term records; 647 individuals (90.5%) carry an explicit *juan*-level source, and every term record passes the mandatory-field check. Under Building Structure, there are 643 individuals: 641 term records and 2 further individuals that are not records, a board (背版) that occurs only as the object of a joining relation, and a second individual for the *ang* variant name 飞棖 that is flagged for merging in the next release. According to ontology class there are 292 components, 174 construction methods, 68 positions, 42 building forms, 34 modular-grade individuals, and 33 dimension individuals; Table 1 relates these six classes to the eight record categories. Of the 2553 object-property assertions, 2427 (95.1%) are structural relations among Building Structure individuals; joining (joinsWith, 634), unit of measure (hasUnit, 392), construction method (hasConstructionMethod, 317), part-of (isPartOf, 296), and position (hasLocus, 229) account for three quarters of all assertions, whereas the 29 links from regulation entries to their structures and the 97 bibliographic relations among books, regulation entries,

persons, places, and times make up the remainder. Pellet reported no unsatisfiable class and no inconsistency; as argued in Section 3.5, this result is a precondition for the remaining measures rather than evidence of quality in itself.

Table 8. Structural profiles of the *Yingzao Fashi* (YZFS) ontology and of the *Gongcheng Zuofa* (GCZF) pilot ontology.

Measure	YZFS (This Study)	GCZF Pilot (Section 5.6)
Named classes (hierarchy depth)	86 (3 levels)	50 (4 levels)
Object properties/data properties	32/27	12/17
Named individuals (term records)	715 (641)	597 (530)
Object-property assertions	2553	6403 (of which 5376 are attestedIn/attests provenance links)
Data-property assertions	4546	4592
Individuals with juan-level source	647 (90.5%); 641 of 641 term records	589 (98.7%)
Term records with variant-name links	57 (96 assertions, 127 labels, 33 clusters)	not modeled (reserved slot)
Members with at least one fen dimension	69 (width 46, thickness 33, length 30)	227 with a dimension expression (strings)
Structural network: terms/links/components	355/913/10 (largest 321 terms)	178/376/3
Pellet consistency check	passed	passed (HermiT/Pellet)

5.2. Coverage Against an Independent Reference Glossary

The reference is the 'Glossary of commonly used terms in traditional architecture' (古建筑常用名词对照表), Appendix 6 of Wang's *Zhongguo Gujianzhu Gouzao Jishu* (X. Wang, 2019). It aligns Song (*Yingzao Fashi*), Qing (*Gongcheng Zuofa*), and Jiangnan (*Yingzao Fayuan*) terms in 199 rows; its four *damuzuo* subsections (columns, beam frames, purlins and rafters, bracket sets) list 94 Song terms and 94 Qing terms and are independent of both the present dataset and the two glossaries used for recall checking in Section 2.1. Table 9 reports the match. The main result is exact coverage. Of the 94 Song *damuzuo* terms, 71 (75.5%) are recorded exactly: 59 as identical strings and 12 after the modern printed forms of four characters are normalized to the forms of the source (板/版, 斗/枓, 枋/方, 座/坐). Exact coverage is highest for bracket sets (33 of 40 terms, 82.5%) and beam frames (27 of 35, 77.1%), where the regulation chapters are richest, and lowest for columns (4 of 9). A further 16 terms (17.0%) match only partially under the containment rule of Section 2.4: nine are compounds that contain a recorded term (殿身檐柱, 柱头铺作, 单材慢栱, 角华栱), and seven are short forms for which only longer forms are recorded (铺作 beside 四铺作 and 八铺作, 内柱 beside 屋内柱). Because containment is a lenient criterion, these 16 terms are shown for information and are not counted as covered; counting them would give 92.6%. The seven terms without any match are two column terms that occur only in the stonework and plan chapters (分心柱, 虚柱), three roof-edge members recorded in the reference under their Qing-style spellings (橑风槫, 橑檐枋, 燕颔板), and two boards (地面板, 绰木方). When the roof and platform subsections are added, exact coverage falls to 54.6% (66.7% with partial matches), as expected, because tiles, ridges, and stone platforms lie outside the *damuzuo* chapters modeled here. For the Qing pilot, exact coverage of the 94 Qing *damuzuo* terms is 61.7% (58 terms), with eight further partial matches. The reference list was scanned from print, and ten obvious optical recognition errors (for example, 平梁 for 平梁, 四椽栿 for 四椽栿) were corrected before matching; the corrected list is included in the Zenodo deposit (Qiao et al., 2026).

Table 9. Coverage of the Song and Qing terms of the independent reference glossary (X. Wang, 2019) by the two datasets (*damuzuo* subsections); exact coverage is the main measure.

Glossary Subsection	Dataset	Ref. Terms	Exact	Exact Coverage	Partial	Missing	Exact + Partial
Columns	YZFS (641)	9	4	44.4%	3	2	77.8%
Beam frames	YZFS	35	27	77.1%	6	2	94.3%
Purlins and rafters	YZFS	10	7	70.0%	0	3	70.0%
Bracket sets	YZFS	40	33	82.5%	7	0	100%
All damuzuo subsections	YZFS	94	71	75.5%	16	7	92.6%
Columns	GCZF pilot (530)	6	5	83.3%	0	1	83.3%
Beam frames	GCZF pilot	36	26	72.2%	2	8	77.8%
Purlins and rafters	GCZF pilot	13	8	61.5%	3	2	84.6%
Bracket sets	GCZF pilot	39	19	48.7%	3	17	56.4%
All damuzuo subsections	GCZF pilot	94	58	61.7%	8	28	70.2%

Note: Exact = string match after normalization of graphic variants (the matching script maps 斗/枓, 拱/栱, 契/栔, 板/版, 枋/方, 座/坐, 攀/襻, 搏/博, 闇/暗, and 標/槕; 12 of the 71 exact matches of Song terms depend on it); partial = the reference term contains, or is contained in, a recorded term of at least two characters, a lenient rule, so partial matches are shown for information and are not counted as covered. Single-character fragments produced by splitting compound cells (上, 中 in '上, 中, 下平槕') were excluded.

5.3. Consistency of the Manual Classification: Model–Human Agreement

Table 10 summarizes the classification experiments and Figure 8 shows the confusion matrices. On the *Yingzao Fashi* records, the best configuration (linear support-vector machine on character n-grams of the term plus its masked contexts) reproduces the eight manually assigned record categories (Table 1) with an accuracy of 0.793, a macro-F1 of 0.784, and a Cohen's kappa of 0.704 under five-fold stratified cross-validation; at the subclass level (the 41 record subclasses with at least five records, n = 569), the accuracy is 0.703 and kappa 0.692. The term string alone already yields kappa 0.65, whereas context alone yields only 0.44, which shows that the manual category assignment is driven chiefly by the morphology of the term (the head character 栱, 枓, 柱, 栿 names the member family) and that context contributes a smaller, complementary signal. Per category (Table 11), the modular system (F1 = 0.92), building form (0.91), material quota (0.88), dimension (0.86), and component (0.83) categories are reproduced reliably; construction method (0.77) is intermediate; and position (0.54) and the small quantity-operator category (0.57, n = 9) are not. The confusion matrix locates the disagreement precisely: 27 of the 68 position records are predicted as components and 13 components as positions, and 51 records are exchanged between component and construction method. These are not random errors. A position such as '*gong* eye' (栱眼) or '*dou* mouth' (枓口) is named after the member it belongs to, and a method such as '*zucai* construction' is named after the member it produces, so the surface form that identifies the family also blurs the structural role. The result quantifies the OntoClean argument of Section 2.2: component is a rigid class, whereas position and method are roles, and the boundary between them is where human and machine disagree most. In the ontology, this boundary is kept explicit by the hasLocus and hasConstructionMethod properties; for future annotation, the rule adopted is that a term denotes a position only if the text uses it with a locative preposition (于, 在, 上, 内).

On the *Gongcheng Zuofa* records, the agreement is higher (accuracy 0.879, kappa 0.775 at category level; 0.850 and 0.838 at subclass level), largely because 61% of its records are components with transparent compound names (for example, 七架梁, 老檐枋), whereas its 18 position records are again poorly recovered (F1 = 0.26). Pooling both texts (n = 1171) gives accuracy 0.839 and kappa 0.749, so the eight-category scheme is learnable across the two texts when both vocabularies are seen. By contrast, a classifier trained on one text and applied to the other performs poorly (YZFS to GCZF: accuracy 0.596, kappa 0.352; GCZF to YZFS: 0.510 and 0.204): the Song modular vocabulary (材, 栔, 分) is never recognized from

the Qing model and the Qing 斗口 terms never from the Song model, and Qing components are frequently assigned to the Song construction-method category. The category scheme therefore transfers, but the surface vocabulary does not, which is the empirical reason for the concept-level alignment property introduced in the pilot ontology (Section 5.6).

Table 10. Model–human agreement of the manual classification (eight record categories unless stated). CV = five-fold stratified cross-validation; LR = logistic regression; SVM = linear support-vector machine; RF = random forest.

Setting	n	Categories	Features/Model	Accuracy	Macro-F1	Kappa
YZFS, CV, category level	641	8	name + context/SVM	0.793	0.784	0.704
YZFS, CV, category level	641	8	name only/LR	0.763	0.724	0.651
YZFS, CV, category level	641	8	context only/LR	0.585	0.583	0.438
YZFS, CV, category level	641	8	name + context/RF	0.671	0.597	0.496
YZFS, CV, subclass level	569	41	name + context/SVM	0.703	0.614	0.692
GCZF, CV, category level	527	6	name only/LR	0.879	0.709	0.775
GCZF, CV, subclass level	513	23	name + context/SVM	0.850	0.692	0.838
Pooled YZFS + GCZF, CV	1171	8	name + context/SVM	0.839	0.799	0.749
Train YZFS, test GCZF	641/530	8	name only/SVM	0.596	0.345	0.352
Train GCZF, test YZFS	530/641	8	name + context/SVM	0.510	0.263	0.204

Table 11. Per-category precision, recall, and F1 for the *Yingzao Fashi* records (SVM, name + context, five-fold CV).

Record Category	Records	Precision	Recall	F1	Main Confusion
Modular system (材契类)	31	0.933	0.903	0.918	construction method (3)
Building form (形制类)	42	0.973	0.857	0.911	position, method, component (2 each)
Material quota (料例类)	24	0.875	0.875	0.875	construction method (2)
Dimension (尺寸类)	10	0.818	0.900	0.857	component (1)
Component (构件类)	285	0.809	0.846	0.827	construction method (27), position (13)
Construction method (做法类)	172	0.763	0.785	0.774	component (24), position (10)
Quantity operator (增减替)	9	0.800	0.444	0.571	method, component (2 each)
Position (位置类)	68	0.576	0.500	0.535	component (27), method (6)

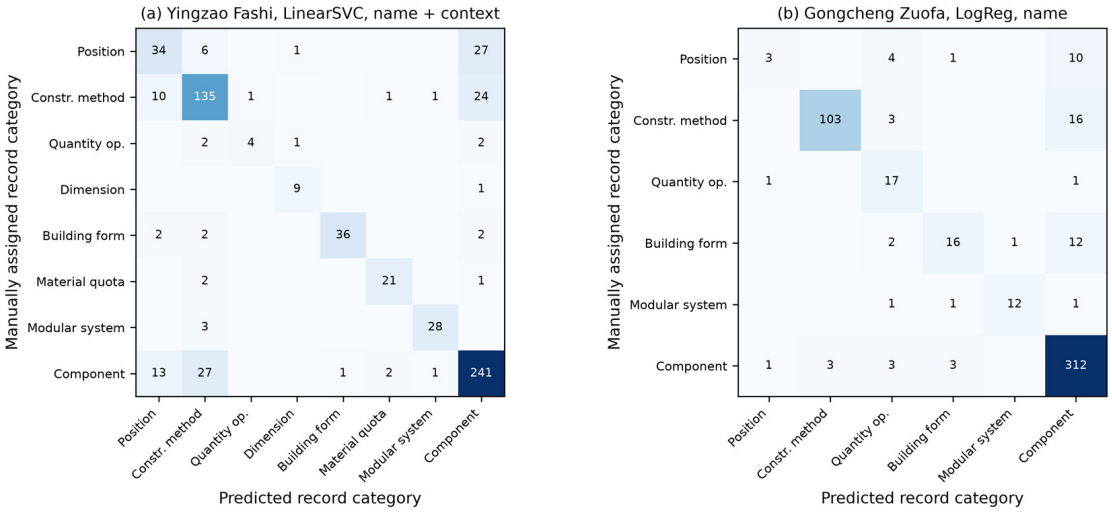


Figure 8. Confusion matrices of the best classifier against the manual category labels under five-fold cross-validation: (**a**) *Yingzao Fashi*, 641 records, eight record categories; (**b**) *Gongcheng Zuofa*, 527 records, six record categories. Rows are manual labels, and columns predictions; off-diagonal mass concentrates on the Position–Component and Component–Method boundaries.

5.4. Recognition of Term Mentions in the Source Text

Projecting the 641 terms onto the 58,046 characters of the *damuzuo* source text by longest match yields 7606 term mentions in 3671 sentences; 56.7% of the mentions are single characters (材, 分, 栱, 枓, 柱), 27.0% two characters, and 16.3% three or more. A character-level tagger trained on these projections under five-fold cross-validation over sentences recovers the mentions with an entity-level precision of 0.955, recall of 0.917, and F1 of 0.936 (token accuracy 0.979; Table 12). Recall is higher for single-character mentions (0.953) than for multi-character ones (0.871), because the frequent heads are easy to memorize while rarer compounds (底版, 斢翅) are split. The decisive figure is the recall on mentions of terms that never occur in the training folds, 0.217 (272 mentions): the tagger learns the vocabulary and the local syntax of the regulation formula ('凡 X 之制', '以 X 为法', '广 N 分') but cannot invent new members from context alone. The most frequent misses are polysemous heads used in non-terminological senses (栿, 栱, 椽, 昂 in enumerations, 楹 and 依 in the General Glossary quotations), and the most frequent spurious hits are the same heads tagged inside compounds that the dictionary had split differently (柱 in 蜀柱, 昂 in 下昂). The experiment therefore supports the manual annotation on two counts. First, the 641 terms are consistently recoverable from the text, with an F1 that is in the range reported for inter-annotator agreement on entity annotation (Artstein & Poesio, 2008; Hripcsak & Rothschild, 2005). Second, the residual errors identify exactly the class of records that needs human judgment, namely, single-character polysemous heads, and 37 records (5.8%) whose recorded form is an editorial normalization that does not occur verbatim in the text (for example, 第一等材 for the text's 第一等, 上平槫 for the enumeration 上, 中, 下平槫). The same procedure applied to the *Gongcheng Zuofa* pilot (227,951 characters, 9629 sentences, 37,210 projected mentions of 527 terms) yields a precision of 0.987, a recall of 0.987, and an F1 value of 0.987, with an unseen-term recall of 0.459.

Table 12. Distantly supervised recognition of term mentions (character-level logistic-regression tagger, five-fold cross-validation over sentences).

Measure	YZFS Damuzuo Text	GCZF Damuzuo Text (Pilot)
Characters/sentences	58,046/3671	227,951/9629
Terms/projected mentions	641/7606	527/37,210
Single-character mentions	4314 (56.7%)	22,201 (59.7%)
Entity-level precision/recall/F1	0.955/0.917/0.936	0.987/0.987/0.987
Token accuracy	0.979	0.994
Recall, single-character/multi-character mentions	0.953/0.871	0.996/0.974
Recall on terms unseen in training folds	0.217 (272 mentions)	0.459 (194 mentions)
Terms not occurring verbatim in the text	37 (5.8%)	13 (2.5%)

5.5. Reasoning Yield and Patterns in the Dataset

Reasoning yield. The typed relations of Section 3.4 were declared with inverses (isPartOf/contains, supports/supportedBy, above/below, insertedInto/receives). Materializing the inverses adds 462 assertions that are not stated in the text (contains 284, isPartOf 99, supportedBy 59, below 16, receives 3, supports 1), because the text almost always reports relations ordered from the smaller member to the larger one. Treating part-of as transitive adds 128 further statements; for example, that the *juantou* (a *huagong* head) is part of the eight-tier bracket set (八铺作) through the *huagong* or that the *yuandou* is part of the *tiaowo* through the *ludou*. Treating support as transitive adds 161 statements, and the longest derived support chain runs over eight members from the beam to the flying rafter

(梁 → 绞昂栿 → 昂 → 枓 → 栱 → 承椽方 → 檐椽 → 飞椽). No single entry states this path, yet it is the complete eave load route implied by the regulations. In total, 751 derived statements (29.4% of the 2553 asserted) can be returned by a Cypher query without needing to be entered by hand (Table 13). Reasoning also audits the data: the part-of closure contains one two-node cycle (实拍襻间 ↔ 襻间) and seven part-of statements whose two ends are also linked as variant names (for example, *juantou* isPartOf *huagong*, where *juantou* is a variant name of the *huagong* head rather than a part of it); these eight cases were flagged for correction in the next release and are counted in the error budget of Section 6.

Table 13. Statements derived by reasoning over the asserted relations (version 4 of the graph).

Derivation	Input Assertions	Derived	Example
Inverse completion (4 inverse pairs)	486	462	Ludou contains Yuandou (from Yuandou isPartOf Ludou)
Part-of transitivity (isPartOf)	395	128	Juantou isPartOf Huagong isPartOf Bapuzuo
Support transitivity (supports)	60	161	Liang → Jiaoangfu → Ang → Dou → Gong → Chengchuanfang → Yanchuan → Feichuan
Variant-name closure (symmetric, transitive)	96	33 clusters, 127 labels	Dong ≡ Fen, Fu, Ji, Lin, Meng, Mian, Tuan, Yin
Audit: cycles and part-of/variant conflicts	—	1 + 7 flagged	Shipaipanjian ↔ Panjian; Juantou isPartOf Huagong

Regarding patterns in terminology, the 641 terms are short: 12.2% are single characters, 40.2% two characters, 23.9% three, and 23.7% four or more. The Qing vocabulary is markedly longer (28.9% of its 530 terms have five or more characters; for example, 九檩单檐庑殿周围廊), reflecting the shift from a glossary-like Song text to procedural Qing specifications. Variant names concentrate on the oldest members of the vocabulary: the eight attested names of the ridge purlin (栋: 栭, 桴, 槾, 甍, 极, 槫, 檩, 榜), the six of the *ang*, the five of the rafter, and the four each of the *dou*, the inverted V-brace (叉手), and the hip rafter (阳马) all come from the General Glossary *juan*, where Li Jie collected names from the classical literature, whereas the regulation chapters add only three or fewer variants per member (*huagong*: 杪栱, 卷头, 跳头). In all, 127 attested labels resolve to 33 concept clusters.

Patterns in component relations: The 355 terms connected by typed structural relations form 913 undirected links in ten components. The largest component contains 321 terms (90.4%), so the bracket-set vocabulary and the frame vocabulary are one connected system rather than two. The degree distribution is heavy-tailed (median 3, mean 5.1, maximum 38; Figure 9a). The hubs are the *huagong* (38 links), the column (33), the rafter (31), the *linggong* (27), the king-post and the *ludou* (25 each), the *nidaogong* and the hip rafter (23 each), and the *tuofeng* (21); they also have the highest betweenness; that is, they are the members through which shortest paths between other members run. A second group of hubs was unexpected: the timber-stock items of *juan* 26 (大料模方, 朴柱, 常使方, 松柱) have 21 to 34 links each because the material quotas list every member that a stock may be cut into, so the quota chapters have the highest cross-reference density in the whole text. Text frequency is only weakly related to structural connectivity (Spearman $\varrho = 0.13$, $p = 0.003$, n = 510; Figure 9b); the most frequent terms (材, 分, 加, 减) are modular and quantity words with few constructional relations, whereas the hubs are moderately frequent members. Frequency, in other words, measures the text's attention; degree measures the building's structure.

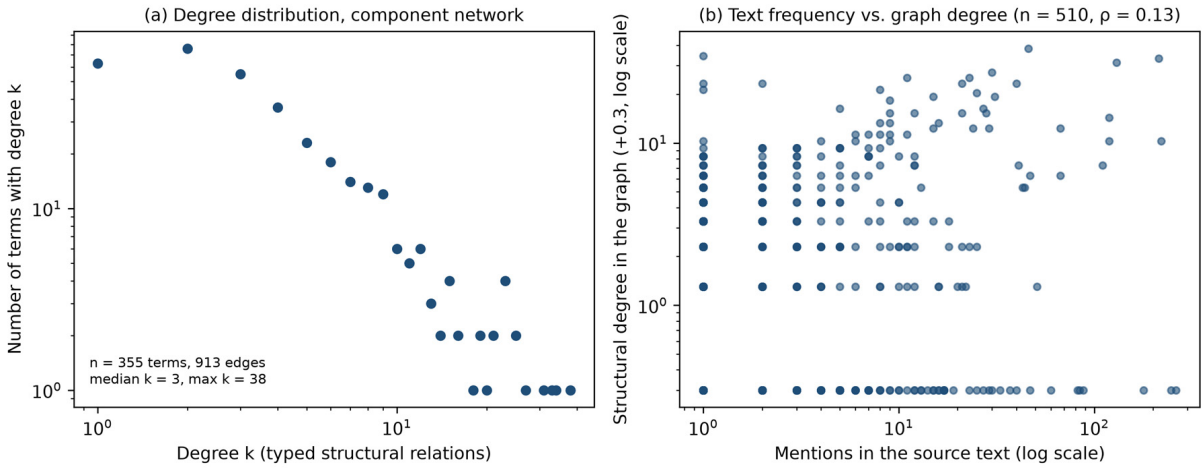


Figure 9. (**a**) Degree distribution of the 355 terms connected by typed structural relations (log–log axes). (**b**) Mentions of each term in the source text against its structural degree in the graph (n = 510 terms matched between text and graph); the weak rank correlation shows that textual frequency and structural connectivity measure different things.

Patterns in construction rules: Sixty-nine members carry at least one dimension in *fen*, and 26 carry both width (广) and thickness (厚). Ordinary least-squares regression of width on thickness gives width = 5.36 + 1.29 × thickness ($R^2 = 0.55$, $p < 0.001$, n = 26), but the scatter (Figure 10a) is not a cloud around a line: the ratios cluster at a few modular values (Figure 10b). Seven members have exactly the single-*cai* ratio 15:10 stated in the '*cai*' entry of *juan* 4, three have the *zucai* ratio 21:10, five have the two-*cai* ratio 30:10, and the remainder sit at 4:3 (16:12), 7:5, and 8:3. The regression therefore confirms the rule the text states—section depth scales with thickness—while the residuals show that the text prescribes a discrete set of modular proportions rather than a continuous law, a fact that a model storing absolute dimensions would hide and that the *fen*-plus-grade representation of Section 3.4 preserves. Length is only weakly predicted by width ($R^2 = 0.34$, n = 14), as expected, since lengths follow the number of tiers or rafters rather than the section.

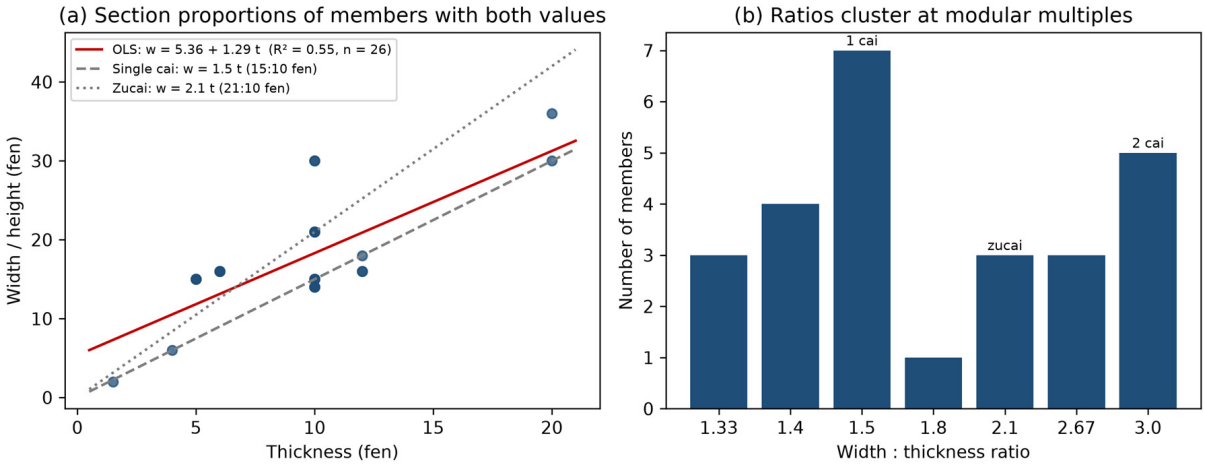


Figure 10. Section proportions of the 26 members with both width and thickness in *fen*: (**a**) ordinary least-squares fit against the single-*cai* (15:10) and *zucai* (21:10) rules; (**b**) distribution of width: thickness ratios, which cluster at modular multiples.

5.6. Pilot Transfer to the Qing Gongcheng Zuofa

To test transferability, the same procedure was applied to the *damuzuo* volumes of the Qing *Gongcheng Zuofa* (1734; Ma, 2003; Ministry of Works, Qing Dynasty, 1734): 530 terms

with their *juan*, original sentence, translation, and gloss were collected chiefly from the forty *damuzuo* and bracket-set volumes (519 terms from *juan* 1–40, three from *juan* 43, and eight without a *juan*-level source), classified with the same two-level scheme, and serialized as a pilot ontology (50 classes, 12 object properties, 17 data properties, 597 individuals, 6403 object-property assertions). The pilot instantiates the structural layer only: it contains the counterparts of Building Structure and its subclasses and one class for the sources (48 individuals: the work itself and *juan* 1–47, in which its terms are attested), whereas the bibliographic layer and the reused general classes (Architectural Book, Building Rule, Person, Spatial Thing, Temporal Entity) do not depend on the text modeled and are not instantiated again. Table 14 records how each Song class fared. All six subclasses of Building Structure have counterparts. The pilot arranges them into three intermediate branches (regulation-level categories: building form, the modular system, and quantity operations; measures and materials; and members: component, position, and construction method) below a root class that expresses the generative role of the module so that its hierarchy has four levels, while the Song ontology has three (Table 8). The member, position, method, and form subclasses transferred with relabeling. The modular class had to be redefined because the Qing text takes the *doukou* (the mortise width of the bearing block, graded in eleven steps of half a *cun*) as its module instead of the *cai–qi* pair, and the Song roof-curve method (*juzhe*) is replaced by the Qing *jujia*. These are two types of record that the Song ontology types belonging to Dimension (Table 1) receive classes of their own in the pilot, namely, the quantity operators (增减替, 19 terms), which the Qing text uses systematically (as in '以斗口七十份定高'; with subclasses for augmentation, reduction, and equivalence) and timber stock (3 terms). Two Song record subclasses (material grade, and nail stock) had no Qing instances. The dimension properties also changed: Song entries report 长, 广, 厚 in *fen*, whereas Qing entries give 长, 高, 宽, 厚, 径 as expressions in *doukou* or in absolute *chi* and *cun*, so a height property was added and the width label changed. Of the object properties, the pilot populates provenance (attestedIn and its inverse, 2688 assertions each), location (locatedAt, 307), joining (joinsWith, 243), containment (contains, 134), and the links to module, unit, and material (governs, 274; madeOf, 39; hasModularGrade, 16; hasMeasurementUnit, 14). The support, vertical, insertion, and sequence relations of the Song ontology were not populated, so the transfer of the typed constructional relations has so far been tested for joining and part-of only. Because the Qing dimensions are expressions ('斗口七十份', '面阔') rather than numbers, the regression analysis described in Section 5.5 could not be repeated for the pilot; converting these expressions to numeric *doukou* multiples was the next step.

Table 14. Transfer of the *Yingzao Fashi* ontology to the *Gongcheng Zuofa* pilot: the fate of each class group and Song–Qing term alignment recorded for the pilot.

Song Class (YZFS)	Qing Class (GCZF Pilot)	Outcome and Reason
ArchitecturalBook, BuildingRule, Person, SpatialThing, TemporalEntity	not instantiated	Bibliographic and general classes that do not depend on the text; the pilot models the structural layer only
Sources	TextualSource (48 sources: the work and juan 1–47)	Retained for provenance
BuildingStructure, with six direct subclasses (three levels)	DoukouModularOrigin, with three intermediate branches that group the subclasses (four levels)	Root renamed after the Qing module; an intermediate level added

Table 14. *Cont.*

Song Class (YZFS)	Qing Class (GCZF Pilot)	Outcome and Reason
Component (18), Position (13), ConstructionMethod (27), BuildingForm (3) subclasses	Component (15), Position (5), ConstructionMethod (9), BuildingForm (3)	Transferred with relabeling; fewer subclasses because the Qing volumes modeled contain fewer distinct families; StructuralRelationMethod grows from 7 to 40 terms
ModularGrade: cai–qi, eight grades	DoukouModularSystem: doukou, eleven grades	Redefined: module changes from a height–thickness pair to a single width
RoofCurveMethod (juzhe)	JujiaRoofCurveMethod (jujia)	Renamed; 14 Qing terms from 五举 to 九举
Dimension, with MeasurementUnit	Measurement, with MeasurementUnit	Renamed
Quantity operators: 9 records typed under Dimension	QuantitativeOperation (increase 7, reduce 8, equal 4)	Class added: systematic in the Qing text
Timber stock: 14 records typed under Dimension	Material, with TimberStock (3 terms)	Class added
Record subclasses material grade (等第), nail stock (钉料)	not instantiated	No Qing instances in the volumes modeled
Dimensions: length, width (广), thickness in fen	length, height, width (宽), thickness, diameter as doukou or absolute expressions	Height added; width relabeled; values are expressions, not numbers
Typed relations: joining, part-of, location, support, vertical order, insertion, sequence	joinsWith (243), contains (134), and locatedAt (307) populated; support, vertical order, insertion, and sequence not populated	Transfer of the constructional relations tested for joining and part-of only
Song–Qing family alignment	8 same-name-same-meaning (14 pairs); 6 same-name-different-meaning (7 pairs); 17 different-name-same-meaning (39 pairs)	Recorded in a separate alignment table; hasSongCounterpart reserved, not asserted
Stability of the classification under revision (first versus final classification of the 530 terms; same annotator, not blind)	category level: 37 changes (7.0%), kappa 0.87; subclass level: 243 changes (45.8%), kappa 0.52, or 0.74 when three spelling harmonizations are discounted	Intra-annotator consistency, not inter-annotator agreement; most subclass changes harmonize labels with the Song scheme

Cross-text alignment was recorded at two levels. At the level of member families, the thirty-one correspondences between Song and Qing member families fall into three types: eight same-name-same-meaning families (14 term pairs, such as 月梁, 角梁, 檐柱, 替木), six same-name-different-meaning families (7 pairs, such as 宝瓶, which is a decorative column finial in the Song text and a structural hip-rafter support in the Qing text, or 庑, which shifts from eave to side hall), and seventeen different-name-same-meaning families (39 pairs, such as 华栱 → 翘, 泥道栱 → 正心瓜栱, 栌枓 → 坐斗, 铺作 → 斗科, 举折 → 举架, 材/分 → 斗口/份). The pilot ontology reserves a hasSongCounterpart property for these pairs but deliberately leaves it unasserted so that the Qing ontology contains no Song data; the cross-text classification results of Section 5.3 show why the alignment must be made explicitly at the concept level rather than inferred from shared labels. Finally, the pilot gives a measure of how stable the classification is under revision, which must be read for what it is. The 530 Qing terms were classified twice by the same annotator (the first author, who curated both datasets): first with working subclass labels when the terms were collected,

and again, three to four weeks later (February to March 2026), when the classification was revised against the scheme used for the Song text. The revision was not blind: the first labels were in view, and the changes were logged with the associated rationale. The figures that follow therefore represent intra-annotator consistency under revision, not agreement between independent annotators. Between the first and the final classification, 37 of the 530 terms (7.0%) changed categories (kappa: 0.87), and 243 (45.8%) changed subclass (kappa: 0.52). Most subclass changes harmonize the labels with the Song scheme without altering a judgment: 113 replace the Qing spelling of a family name with the Song one (枋 by 方, 板 by 版, 檩 by the composite 槫檩桁), and with these, discounted subclass agreement is 75.5% (kappa: 0.74). Fifty-seven distribute two generic bracket-set subclasses (斗栱) over the family subclasses used for the Song text (铺作, 栱, 枓, 昂, 耍头); twenty-nine reduce twelve building-form labels to the three Song forms; nineteen resolve a residual 'other components' subclass; and the remaining twenty-five are the eleven *doukou* sizes and fourteen individual reassignments. Of the 37 category changes, 19 move terms from component to construction method (17 of them bracket-set assemblies such as 斗科, 柱头科, and 品字科), 11 move the *doukou* sizes from material quota to the modular category, five cross the position boundary (斗底, 斗耳, and 套兽榫 from component to position; 搭交 and 斜搭交 from position to construction method), and two are single terms. The category layer is thus stable under revision, whereas the subclass layer depended on labeling conventions that were fixed only when the two texts were aligned. Once again, the terms that crossed a category boundary involve position and construction method, the role-like categories reported in Section 5.3.

6. Discussion

The results support three claims about the model (that it covers the vocabulary of its scope, that its classification is reproducible, and that it transfers) and they bound each of them. Exact coverage of an independent glossary (75.5% of its Song *damuzuo* terms, or 92.6% if partial matches are admitted) shows that the regulation chapters, modeled entry by entry, yield most of the vocabulary that architectural historians treat as the common vocabulary of the subject, with the remaining gaps located in compound forms and in chapters outside the present scope. Model–human agreement (kappa 0.70 at category level, 0.69 at subclass level) shows that the manual classification is reproducible from the term and its contexts are, to a degree, comparable with human inter-coder agreement on comparable tasks (Artstein & Poesio, 2008). The confusion matrices locate one boundary, position versus component, where the scheme is genuinely ambiguous. The distantly supervised tagger ($F1 = 0.94$) shows that the terms are recoverable from the text, but its unseen-term recall (0.22) shows equally clearly that automation can reproduce an annotation but not replace the philologist who created it. Reasoning adds 29% to the asserted relations and yields statements, such as the eight-member support path from beam to flying rafter, that are implied by the regulations but stated nowhere; the same closure exposed eight erroneous assertions, which is the kind of return that justifies formalization. Finally, the pilot shows that the structural classes transfer to a second state standard with one redefinition (the module) and two additions (classes for quantity operators and for timber stock), but that vocabulary does not transfer; transferability is therefore a property of the class scheme, not of the terms, and for the typed relations, it has so far been tested only in part (Section 5.6).

The novelty of this work is that, compared with the heritage ontologies reviewed in Section 1 (Acierno et al., 2017; Carriero et al., 2019; J. Chen & Ou, 2020; Doerr, 2003; Quattrini et al., 2017), the model takes the regulation entry rather than the object or the intervention as its unit, which is what allows prescriptive dimensions and permitted variant names to be first-class data; compared with knowledge graphs of Chinese classical texts (W. Li et al.,

2024b; C. Liu et al., 2024; L. Wang et al., 2023, 2024) and of other craft studies (Zhou et al., 2024, 2025), it adds typed constructional relations and modular derivation and is, to our knowledge, the first such model whose completeness and reproducibility are measured rather than asserted. The tools used (Protégé, RDF, Neo4j) are standard; the contribution lies in the modeling decisions reported in Table 5, in the evaluation protocol described in Section 2.4, and in the dataset and findings presented in Section 5, all of which are available.

Regarding difficulties encountered and resulting insights, four problems recurred during modeling and are worth reporting for similar projects. (1) Polysemous single characters: Twelve per cent of the terms are single characters that also occur as ordinary words (材 'timber' and 'module', 分 'divide' and '*fen*', 加 'add'); they inflate dictionary matches and are the main source of tagger errors. The solution was to record the regulation sentence with every term so that the terminological sense is attested and to treat the quantity words as a record category of their own (增减替) rather than as noise. (2) Editorial forms: Thirty-seven records are normalized forms that do not occur verbatim in the text (第一等材, 上平槫); they are needed for retrieval but must be marked as editorial, which is performed using the originalText property. (3) Position versus component: The same head character names a member, its positions, and its working methods. The Component, Position, and Construction Method subclasses of Building Structure separate them in the ontology, but the classifiers and the revision of the Qing classification show that terms are easily assigned across the boundary. A locative test (the term is used with 于, 在, 上, 内) was adopted as the operational criterion. (4) Variant names versus parts: Seven variant names were also entered as parts; the closure detected them, and the lesson is that the two relations must be checked against each other at the load time. A fifth difficulty is institutional rather than technical: a second independent annotator and a formal expert scoring panel were not available for this study, so the agreement figures reported here are model–human agreement and the intra-annotator consistency described in Section 5.6. They show that the category scheme is learnable and stable under revision, but they are not substitutes for inter-annotator agreement; the category definitions of Table 1, the locative rule, and the released term records are intended to make such a study possible.

The scope of this study also indicates that further integration of digital technologies remains necessary. Five limitations remain. First, the coverage is limited to the *damuzuo* chapters of the *Yingzao Fashi*; the chapters on small carpentry, stonework, brickwork, and polychrome painting in the same book can in principle be organized by the same logic but have not yet been recorded, and the coverage results reported in Section 5.2 decrease accordingly when roof and platform terms are included. Second, variant names are currently expressed as identity assertions between individuals, whereas a variant name is in fact a different label for the same concept; a later version could migrate to the label-level equivalence constructs of the Simple Knowledge Organization System (SKOS; Miles & Bechhofer, 2009), making the vocabulary more consistent with thesaurus practice; the 33 clusters of Section 5.5 are the natural SKOS concepts. Third, the image–text association between drawing-type documents and the text has not yet been incorporated into the ontology, and the correspondence between illustrations and entries remains to be represented. Fourth, the cross-lingual equivalence of terminology still requires further regulation, since most building terms have no established English equivalents and the available approximations often carry the presuppositions of other woodworking traditions; the glossary in Appendix A records the choices made here. Fifth, the Qing pilot stores dimensions as expressions, so the modular analysis described in Section 5.5 could not yet be repeated across the two texts; converting the Qing expressions to numeric *doukou* multiples and asserting the Song–Qing counterpart links are the immediate next steps, after which the diachronic question of how the same member changed its proportions between 1103 and 1734 becomes answerable with a single query.

7. Conclusions

In this study, a regulation-centered semantic model was designed, populated, and evaluated for traditional Chinese architectural books, using the *damuzuo* chapters of the *Yingzao Fashi* as the case. The ontology (86 classes, 32 object properties, and 27 data properties) treats the regulation entry as the unit of description, keeps modular dimensions in *fen* so that absolute sizes remain derivable from the grade, and assigns a type to the constructional relations that the text records; populated, it holds 641 term records and 2553 relation assertions, and it was mapped to a Neo4j graph for querying. This evaluation is quantitative: the dataset records exactly 75.5% of the Song *damuzuo* terms of an independent glossary (92.6% if partial matches are admitted); supervised classifiers reproduce the manual classification with kappa 0.70, locating the position–component boundary as the one systematic ambiguity. A distantly supervised tagger recovers term mentions with F1 = 0.94 but cannot generalize to unseen terms, and inverse and transitive reasoning adds 751 statements, including an eight-member support path implied but not stated by the text, while exposing eight erroneous assertions. Analysis of the dataset shows that variant names concentrate in the General Glossary; that *huagong*, column, rafter, *linggong*, and *ludou* are the structural hubs of the vocabulary; that textual frequency is only weakly related to structural connectivity; and that section proportions cluster at a small set of modular ratios rather than following a continuous law. A pilot ontology for the Qing *Gongcheng Zuofa* (530 terms) shows that the structural classes transfer with one redefinition and two additions, whereas the vocabulary does not, so cross-text alignment must be asserted at the concept level. The designated relations have, so far, been populated for the Qing text only in part. The ontology, the datasets, the alignment table, and the evaluation scripts have been made available so that the measures reported here can be reproduced and extended to the remaining chapters and texts.

Author Contributions: Conceptualization, E.Q. and S.Z.; methodology, E.Q. and S.Z.; software, E.Q.; validation, E.Q., J.D. and S.Z.; formal analysis, E.Q.; investigation, E.Q.; data curation, E.Q.; writing—original draft preparation, E.Q. and S.Z.; writing—review and editing, J.D.; visualization, E.Q.; supervision, J.D.; project administration, J.D.; funding acquisition, J.D. All authors participated sufficiently in the work to take public responsibility for the content. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Key Research and Development Programme of the Inner Mongolia Autonomous Region [grant number 2021GG0407] and the Doctoral Research Start-up Fund of Inner Mongolia Normal University. The APC was funded by the authors' institutions. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data Availability Statement: The data and code that support this study are openly available in the Zenodo repository at https://doi.org/10.5281/zenodo.24066263 (version 1.0, accessed on 10 November 2026; Qiao et al., 2026): the ontology (RDF/OWL, version 4); the 641 *Yingzao Fashi* term records (CSV); the 530 *Gongcheng Zuofa* pilot records and the log of the two classification rounds (CSV); the pilot ontology (RDF/OWL); the Song–Qing class alignment table; the corrected reference glossary list used for the coverage check; the five Cypher queries of Table 6; and the evaluation scripts (Python) with the complete result tables of all experiments. A helper script in the deposit recreates the folder layout that the evaluation scripts expect. The raw source texts are drawn from published editions of the *Yingzao Fashi* and the *Gongcheng Zuofa* and are therefore not redistributed here, and the same applies to the sentence contexts of the 6243 and 32,608 annotated term mentions, which are available from the corresponding author on reasonable request; the experiments that read these inputs (model–human agreement, mention recognition, and the text-frequency statistics) can be repeated once they are placed at the paths documented in the deposit.

Acknowledgments: During the preparation of this work, the authors used generative AI tools to assist with language editing and translation of the manuscript, to improve the readability of the English text, and to assist in drafting and debugging the evaluation scripts. The authors reviewed, executed, and verified all AI-assisted output, checked every reported number against the released data, and take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript: AB, Architectural Books (vocabulary); CIDOC-CRM, CIDOC Conceptual Reference Model; CSV, comma-separated values; CV, cross-validation; DC, Dublin Core; FOAF, Friend of a Friend; GEO, Geographic Vocabulary; GCZF, *Gongcheng Zuofa*; LR, logistic regression; OLS, ordinary least squares; OWL, Web Ontology Language; RDF, Resource Description Framework; RDFS, Resource Description Framework Schema; RF, random forest; SHACL, Shapes Constraint Language; SHL, Shanghai Library Ontology; SKOS, Simple Knowledge Organization System; SQL, Structured Query Language; SVM, support-vector machine; TIME, Time Ontology; XSD, XML Schema Definition; YZFS, *Yingzao Fashi*.

Appendix A. Glossary of Chinese Terms Used in This Paper

Romanization follows Hanyu Pinyin. Chinese terms are printed in simplified characters, as in the punctuated working text from which the records were collected (for example, 营造法式, 华栱, 栋). The craft-specific characters of the Song text are kept as shown in the source and not replaced by their modern counterparts: 枓 (not 斗), 栱 (not 拱), 栔 (not 契), 槫 (not 檩), 版 (not 板), and 方 (not 枋). Qing terms keep the Qing forms (斗口, 枋, 檩), terms quoted from the reference glossary of Section 5.2 are given as printed there, and rare characters without a simplified form in common use (for example, 槫, 閞, 欘) are printed as in the edition. English renderings are working glosses, not established equivalents (see Section 6).

Table A1. Glossary.

Term (Pinyin)	Characters	Gloss
Yingzao Fashi	营造法式	Building Standards, state building code compiled by Li Jie, promulgated 1103
Gongcheng Zuofa (Zeli)	工程做法 (则例)	Engineering Practice, Qing state building code of the Board of Works, 1734
Yingzao Fayuan	营造法原	Source of Building Methods, private Jiangnan treatise recorded by Yao Chengzu
Ziren Yizhi	梓人遗制	Traditions of the Joiner's Craft, Yuan-dynasty treatise by Xue Jingshi
Mujing/Mengxi Bitan	木经/梦溪笔谈	Timber Classic (lost, by Yu Hao)/Dream Pool Essays by Shen Kuo, which quotes it
Kanxiang	看详	Prolegomenon of the Yingzao Fashi; its section on variant names (诸作异名) lists alternative terms
juan	卷	Fascicle or chapter of a traditional book
damuzuo	大木作	Structural ('large') carpentry: the timber frame and bracket sets
xiaomuzuo	小木作	Non-structural ('small') carpentry: doors, windows, ceilings

Table A1. *Cont.*

Term (Pinyin)	Characters	Gloss
cai	材	The modular timber section (15 fen high, 10 fen thick) graded in eight ranks; also ‘timber’
qi (also read zhi)	栔	The gap module (6 fen by 4 fen) between two cai
fen (fen°)	分	Modular unit, one-fifteenth of the cai height; written fen° to distinguish it from the length unit
zucai/dancai	足材/单材	‘Full’ section (one cai plus one qi, 21 fen)/‘single’ section (one cai, 15 fen)
doukou	斗口	Qing module: the mortise width of the bearing block, graded in eleven steps
puzuo	铺作	Bracket set; also the count of its tiers (e.g., bapuzuo, eight-tier set)
gong	栱	Bracket arm
huagong (variants miaogong, juantou, tiaotou)	华栱 (杪栱, 卷头, 跳头)	Projecting bracket arm, perpendicular to the wall plane
nidaogong/guazigong/linggong/mangong	泥道栱/瓜子栱/令栱/慢栱	The four transverse bracket arms: inner, intermediate, outermost, long
ang (xia’ang, shang’ang)	昂 (下昂, 上昂)	Slanting cantilever arm of the bracket set (descending, ascending)
dou; ludou; jiaohudou; sandou	枓; 栌枓; 交互枓; 散枓	Bearing block; large bearing block on the column; cross block at a tier head; small block
shuatou (juetou)	耍头 (爵头)	Projecting arm head at the top of the bracket set
fu (liang)	栿 (梁)	Beam; rufu, two-rafter beam; pingliang, top beam
tuan (dong)	槫 (栋)	Purlin; the ridge purlin has eight attested names
chuan/feichuan	椽/飞椽	Rafter/flying (eave) rafter
zhu/shuzhu/zhuruzhu	柱/蜀柱/侏儒柱	Column/short post on a beam/king-post
fang	方 (枋)	Horizontal tie or lintel timber
timu/tuofeng	替木/驼峰	Bearing plate under a purlin/camel-hump block on a beam
chashou/tuojiao	叉手/托脚	Inverted V-brace under the ridge/inclined strut under a purlin
jiaoliang (yangma)	角梁 (阳马)	Hip rafter beam
juzhe/jujia	举折/举架	Song/Qing methods of setting the roof curve
zengjianti	增减替	Quantity operators in the text: increase, decrease, ‘as appropriate’, substitute
cun, chi	寸, 尺	Absolute length units (about 3.1 cm and 31 cm in the Song)

References

Acierno, M., Cursi, S., Simeone, D., & Fiorani, D. (2017). Architectural heritage knowledge modelling: An ontology-based framework for conservation process. *Journal of Cultural Heritage*, *24*, 124–133. [CrossRef]

Angles, R. (2012, April 1–5). *A comparison of current graph database models*. 2012 IEEE 28th International Conference on Data Engineering Workshops (pp. 171–177), Arlington, VA, USA. [CrossRef]

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*(4), 555–596. [CrossRef]

Baierer, K., Dröge, E., Eckert, K., Goldfarb, D., Iwanowa, J., Morbidoni, C., & Ritze, D. (2017). DM2E: A linked data source of digitised manuscripts for the digital humanities. *Semantic Web*, *8*(5), 733–745. [CrossRef]

Barité, M. (2018). Literary warrant. *Knowledge Organization*, *45*(6), 517–536. [CrossRef]
Bartalesi, V., & Meghini, C. (2017). Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study. *Semantic Web*, *8*(3), 385–394. [CrossRef]
Bartalesi, V., Pratelli, N., Lenzi, E., & Pontari, P. (2023). Using Semantic Web to create and explore an index of toponyms cited in medieval geographical works. *Journal on Computing and Cultural Heritage*, *16*(2), 26. [CrossRef]
Beghtol, C. (1986). Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, *30*(2), 109–125.
Brank, J., Grobelnik, M., & Mladenić, D. (2005, October 17). *A survey of ontology evaluation techniques*. Conference on Data Mining and Data Warehouses (SiKDD 2005) (pp. 166–170), Ljubljana, Slovenia.
Brickley, D., & Miller, L. (2014). *FOAF vocabulary specification 0.99*. Available online: http://xmlns.com/foaf/0.1/ (accessed on 20 January 2026).
Carriero, V. A., Gangemi, A., Mancinelli, M. L., Marinucci, L., Nuzzolese, A. G., Presutti, V., & Veninata, C. (2019). ArCo: The Italian cultural heritage knowledge graph. In *Lecture notes in computer science: Vol. 11779. The Semantic Web—ISWC 2019* (pp. 36–52). Springer. [CrossRef]
Chansanam, W., & Tuamsuk, K. (2022). Development of Thai culture ontology-based metadata: Heet Sib Song (Twelve Months Festival). *Humanities, Arts and Social Sciences Studies*, *22*(3), 725–737. [CrossRef]
Chen, C.-M., & Chang, C. (2019). A Chinese ancient book digital humanities research platform to support digital humanities research. *The Electronic Library*, *37*(2), 314–336. [CrossRef]
Chen, J., & Ou, S. (2020). Research on the construction of the semantic model for Chinese ancient architectures based on architectural narratives. *The Electronic Library*, *38*(4), 769–784. [CrossRef]
Chen, M. (2010). *Glossary of the Yingzao fashi*. Tianjin University Press. (In Chinese)
Cheng, Y.-J., & Chou, S.-L. (2022). Using digital humanity approaches to visualize and evaluate the cultural heritage ontology. *The Electronic Library*, *40*(1/2), 83–98. [CrossRef]
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. [CrossRef]
Cox, S., & Little, C. (2017, October 19). *Time ontology in OWL* (W3C Recommendation). Available online: https://www.w3.org/TR/2017/REC-owl-time-20171019/ (accessed on 20 January 2026).
Debruyne, C., Munnelly, G., Kilgallon, L., O'Sullivan, D., & Crooks, P. (2022). Creating a knowledge graph for Ireland's lost history: Knowledge engineering and curation in the Beyond 2022 project. *Journal on Computing and Cultural Heritage*, *15*(2), 25. [CrossRef]
Doerr, M. (2003). The CIDOC Conceptual Reference Module: An ontological approach to semantic interoperability of metadata. *AI Magazine*, *24*(3), 75–92. [CrossRef]
Dublin Core Metadata Initiative. (n.d.). *Dublin core (DC)*. Available online: https://www.dublincore.org/ (accessed on 20 January 2026).
Feng, J. (2012). *Chinese architecture and metaphor: Song culture in the Yingzao fashi building manual*. University of Hawai'i Press.
Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997, March 24–26). *METHONTOLOGY: From ontological art towards ontological engineering*. AAAI-97 Spring Symposium on Ontological Engineering (pp. 33–40), Stanford, CA, USA.
Foka, A., Barker, E., Konstantinidou, K., Mostofian, N., Demiroglu, O. C., Kiesling, B., & Talatas, L. (2020). Semantically geo-annotating an ancient Greek "travel guide": Itineraries, chronotopes, networks, and linked data. In *4th ACM SIGSPATIAL workshop on geospatial humanities, Seattle, WA, USA, 3 November 2020* (pp. 1–9). ACM. [CrossRef]
Gennari, J. H., Musen, M. A., Fergerson, R. W., Grosso, W. E., Crubézy, M., Eriksson, H., Noy, N. F., & Tu, S. W. (2003). The evolution of Protégé: An environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, *58*(1), 89–123. [CrossRef]
Guarino, N., & Welty, C. (2002). Evaluating ontological decisions with OntoClean. *Communications of the ACM*, *45*(2), 61–65. [CrossRef]
Guo, Q. (1998). *Yingzao fashi*: Twelfth-century Chinese building manual. *Architectural History*, *41*, 1–13. [CrossRef]
Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008, August 19–24). *Exploring network structure, dynamics, and function using NetworkX*. 7th Python in Science Conference (SciPy 2008) (pp. 11–15), Pasadena, CA, USA.
Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the F-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, *12*(3), 296–298. [CrossRef] [PubMed]
Hyvönen, E. (2023). Digital humanities on the Semantic Web: Sampo model and portal series. *Semantic Web*, *14*(4), 729–744. [CrossRef]
Jaitiang, T., Kabmala, M., & Chansanam, W. (2023). Development of ontology for Thai country songs. *Journal of Information Science Theory and Practice*, *11*(1), 79–88. [CrossRef]
Joo, S., Hootman, J., & Katsurai, M. (2022). Exploring the digital humanities research agenda: A text mining approach. *Journal of Documentation*, *78*(4), 853–870. [CrossRef]
Khamparia, A., & Pandey, B. (2017). Comprehensive analysis of Semantic Web reasoners and tools: A survey. *Education and Information Technologies*, *22*(6), 3121–3145. [CrossRef]

Knublauch, H., & Kontokostas, D. (2017, July 20). *Shapes Constraint Language (SHACL)* (W3C Recommendation). Available online: https://www.w3.org/TR/shacl/ (accessed on 5 November 2026).

Li, F., & Xia, S. (2023). International cooperation in the digital preservation of the cultural heritage of the Dunhuang Mogao Grottoes. *Library Trends*, *71*(3), 345–363. [CrossRef]

Li, J. (1103). *Yingzao fashi* [Building standards] (Siku Quanshu ed., *juan* 4, *Damuzuo* regulations I). Jiangzuo Jian. (In Chinese)

Li, J. (1103/1983). *Yingzao fashi* [Building standards] (S. Liang, Annot.). China Architecture & Building Press. (Original work published 1103) (In Chinese)

Li, J., & Bikakis, A. (2023). Towards a semantics-based recommendation system for cultural heritage collections. *Applied Sciences*, *13*(15), 8907. [CrossRef]

Li, W., Wang, J., & Wang, F. (2024a). Curating the Chinese ancient book catalogs: Leveraging the dual roles of humanities scholars as experts and users in collaborative practice. *Journal of the Association for Information Science and Technology*, *75*(12), 1331–1349. [CrossRef]

Li, W., Wang, Z., Yu, H., Lai, Q., & Wang, J. (2024b). Book cluster: A new approach to explore the evolution of knowledge organization in ancient China. *Digital Scholarship in the Humanities*, *39*(4), 1064–1074. [CrossRef]

Liu, C., Wang, D., Zhao, Z., Hu, D., Wu, M., Lin, L., Liu, J., Zhang, H., Shen, S., Li, B., & Zhao, L. (2024). SikuGPT: A generative pre-trained model for intelligent information processing of ancient texts from the perspective of digital humanities. *Journal on Computing and Cultural Heritage*, *17*(4), 53. [CrossRef]

Liu, J., Bian, J., Song, F., Shen, S., & Lai, X. (2023). Knowledge organization initiative for digital resources of ancient Chinese stele inscriptions. *Library Trends*, *71*(4), 592–628. [CrossRef]

Ma, B. (2003). *Zhongguo gujianzhu muzuo yingzao jishu* [Timber construction technology of traditional Chinese architecture] (2nd ed.). Science Press. (In Chinese)

Miles, A., & Bechhofer, S. (2009, August 18). *SKOS Simple Knowledge Organization System reference* (W3C Recommendation). Available online: https://www.w3.org/TR/skos-reference/ (accessed on 5 November 2026).

Ministry of Works, Qing Dynasty. (1734). *Gongcheng zuofa zeli* [Rules and regulations for engineering construction] (Woodblock). Gongbu. (In Chinese)

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009, August 2–7). *Distant supervision for relation extraction without labeled data*. Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (pp. 1003–1011), Suntec, Singapore.

Musen, M. A. (2015). The Protégé project: A look back and a look forward. *AI Matters*, *1*(4), 4–12. [CrossRef] [PubMed]

Ortolja-Baird, A., Pickering, V., Nyhan, J., Sloan, K., & Fleming, M. (2019). Digital humanities in the memory institution: The challenges of encoding Sir Hans Sloane's early modern catalogues of his collections. *Open Library of Humanities*, *5*(1), 44. [CrossRef]

Pan, G., & He, J. (2005). *Interpretation of the Yingzao fashi*. Southeast University Press. (In Chinese)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Qiao, E., Dong, J., & Zhou, S. (2026). *Knowledge organization of traditional Chinese architectural books: Ontology, datasets, and evaluation scripts for the Yingzao fashi case study* (Version 1.0) [Data set]. Zenodo. [CrossRef]

Quattrini, R., Pierdicca, R., & Morbidoni, C. (2017). Knowledge-based data enrichment for HBIM: Exploring high-quality models using the Semantic-Web. *Journal of Cultural Heritage*, *28*, 129–139. [CrossRef]

Shanghai Library. (n.d.). *Ontology and vocabularies*. Shanghai Library Open Data Platform. Available online: https://data.library.sh.cn/ (accessed on 18 November 2026).

Shen, K. (ca. 1088). *Mengxi bitan* [Dream pool essays]. (In Chinese)

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2012). The NeOn methodology for ontology engineering. In M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, & A. Gangemi (Eds.), *Ontology engineering in a networked world* (pp. 9–34). Springer. [CrossRef]

UNESCO. (2009). *Chinese traditional architectural craftsmanship for timber-framed structures*. Representative List of the Intangible Cultural Heritage of Humanity. Available online: https://ich.unesco.org/en/RL/chinese-traditional-architectural-craftsmanship-for-timber-framed-structures-00223 (accessed on 20 January 2026).

Vrachliotou, M., & Papatheodorou, C. (2024). Interoperability of oral history metadata: An ontological model. *The Oral History Review*, *51*(2), 446–474. [CrossRef]

Vrandečić, D. (2009). Ontology evaluation. In S. Staab, & R. Studer (Eds.), *Handbook on ontologies* (2nd ed., pp. 293–313). Springer. [CrossRef]

W3C Semantic Web Interest Group. (n.d.). *Basic Geo (WGS84 lat/long) vocabulary* (D. Brickley, Ed.). W3C Semantic Web Interest Group. Available online: https://www.w3.org/2003/01/geo/ (accessed on 11 November 2026).

Wang, G. (2023). The Society for Research in Chinese Architecture, Liang Sicheng and *Yingzao fashi*: A scholarly journey that has endured for more than half a century. *Journal of Architectural History*, *4*(4), 128–137. (In Chinese) [CrossRef]

Wang, L., Wang, J., & Tong, W. (2023). Using ontology to organize Chinese ancient books in the digital age. *Proceedings of the Association for Information Science and Technology*, *60*(1), 712–716. [CrossRef]

Wang, L., Wang, J., & Wei, T. (2024). Using ontology to model time description in historical Chinese texts. *Digital Scholarship in the Humanities*, *39*(1), 423–429. [CrossRef]

Wang, W., Ning, H., Shi, F., Dhelim, S., Zhang, W., & Chen, L. (2022). A survey of hybrid human-artificial intelligence for social computing. *IEEE Transactions on Human-Machine Systems*, *52*(3), 468–480. [CrossRef]

Wang, X. (2019). *Zhongguo gujianzhu gouzao jishu* [Construction technology of traditional Chinese architecture] (2nd ed., Appendix 6, Gujianzhu changyong mingci duizhaobiao [Glossary of commonly used terms in traditional architecture], pp. 422–429). Chemical Industry Press. (In Chinese)

Xue, J. (ca. 1264). *Ziren yizhi* [Traditions of the joiner's craft]. (In Chinese)

Yao, C. (1986). *Yingzao fayuan* (Z. Zhang, Ed.; D. Liu, Rev.). China Architecture & Building Press. (In Chinese)

Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. Review article. *El Profesional de la Información*, *28*(1), e280103. [CrossRef]

Zhang, Y., Yao, J., & Feng, X. (2025). Analysis of the theory of data production and reproduction of cultural heritage digital resources. *Library Journal*, *44*(11), 4–14, 39. (In Chinese) [CrossRef]

Zhao, W., & Chen, Y. (2023). The construction and open sharing of ancient books resources in the National Library of China. *Library Trends*, *71*(4), 466–491. [CrossRef]

Zhou, S., Meng, F., & Huang, Y. (2025). Semantic modeling of ancient Chinese Guqin books. *Knowledge Organization*, *52*(3), 38541. [CrossRef]

Zhou, S., Zhang, J., & Chen, H. (2024). Research on knowledge representation and semantic association construction of Sichuan Qingyin. *Research on Library Science*, (2), 63–74. (In Chinese) [CrossRef]

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