

## Article

# Research on the Biography Digital Storytelling Based on Ontology

Li Si<sup>1</sup>, Yao Liu<sup>1,\*</sup>
<sup>1</sup>Department of Library Science, School of Information Management, Wuhan University, 430072 Wuhan, Hubei, China

\*Correspondence: [852530653@qq.com](mailto:852530653@qq.com) (Yao Liu)

Academic Editor: Natália Tognoli

Submitted: 29 June 2024   Revised: 21 February 2025   Accepted: 7 March 2025   Published: 27 April 2025



Li Si is a Professor, PhD and Researcher of Center for the Studies of Information Resources of Wuhan University, Department of Library Science, School of Information Management, Wuhan University. She is the program leader of the Master's degree in Library and Information Science at Wuhan University. She is a member of the Identification and Description Technical Committee of National Information and Document Standardized Technical Committee (China) and member of the International Society for Knowledge Organization. Her research interests include information organization, knowledge organization and management, KOS, and scientific data management. She has authored and co-authored 16 books and published over 200 articles.



Yao Liu is a Ph.D. candidate in the Department of Library Science, School of Information Management, Wuhan University. She completed both his undergraduate and master's degrees at the same department. Her research focuses on information organization, knowledge organization, and digital humanities.

## Abstract

Biography digital storytelling is a narrative method rooted in reconstruction and deduction, designed to achieve the associative, semantic, situational, and visual representation of biographical content. Drawing from narratology theory, this research explores an ontology-based implementation pathway for biography digital storytelling through four key stages: selection and characteristics analysis of storytelling materials, extraction and semantic association of storytelling elements via ontology, clustering biographical themes and connecting narrative threads using Latent Dirichlet Allocation (LDA), and visualizing the biography through structured storylines. The study aims to deepen the mining of biographical knowledge, enable creative interpretation of storytelling content, and more effectively convey the biographical subject's spirit, offering both an innovative technical approach and theoretical perspectives on biography digital storytelling.

**Keywords:** biography digital storytelling; ontology; LDA

## 1. Introduction

Since the 1990s, the “digital turn” in narratology has driven the emergence of digital storytelling (Wang and Zhao, 2024, pp. 30), a practice or method conveying narratives through digital technologies and media (Dupagne, 2010). Digital storytelling involves the integration, reconstruction, relation, and presentation of existing resources using digital tools such as semantic web and computational techniques and facilitates various modes of expression and dissemination using multimedia, enhancing interactivity and constructing digital memory (Fu et al, 2023, pp. 1523). Biography, as a documented account of individual experiences and a repository of social memory, provides a vivid subject matter for digital storytelling. Biogra-

phy digital storytelling is inherently linked to the processing of biographical texts and historical documents, as well as the systematic organization and sequencing of storytelling elements, including characters, institutions, temporal settings, spatial contexts, and events. Knowledge organization through ontology can standardize and interrelate domain-specific knowledge, thus acting as a bridge between digital storytelling and biographies. Based on ontology, in conjunction with natural language processing (NLP) and visualization technology, this study employs the *Autobiography of Library Scientist Peng Feizhang at the Age of Ninety* (hereafter referred to as *Autobiography*) as an empirical case study. Drawing on narratology theory and grounded in the selection and analysis of storytelling materials, this



study aims to extract and link storytelling elements through biography ontology, organize storytelling threads using Latent Dirichlet Allocation (LDA), and visually present biography digital storytelling via storylines. This approach aims to enhance the comprehensibility of biographical knowledge while exploring a methodological framework for implementing biography digital storytelling.

## 2. Literature Review

### 2.1 Research on Biography Ontology

Ontology is a conceptual, semantic-level framework or a concrete artifact provided for a specific purpose (Guarino and Pierdaniele, 1995). It defines classes, properties, and relations among the members of classes that are relevant to model the knowledge of a domain (Gruber, 2009). This enables ontology to represent complex relationships between knowledge entities, integrate heterogeneous data resources, infer implicit knowledge, support knowledge integration, and enhance interoperability. Several studies have applied ontology to the knowledge organization of biography, mainly focusing on the construction principles and practical implementation of biography ontology.

Three methodologies guide biographical ontology construction:

- Character-centric. The ontologies encapsulate educational backgrounds, professional competencies, and interpersonal relationships. Adhi and Widodo (2019) developed an ontology focused on characters, detailing their relatives, partners, and skills. Motta et al (2009) developed an ontology for musicians, containing person, works, and time-related entities. Leskinen and Hyvönen (2020) introduced a model comprising seven entities: person, relationship role, family relationship role, label, occupation, event, and place.

- Event-driven. The ontologies utilize temporal and spatial dimensions to forge connections. Yeh's (2018) Biographic Knowledge-based Story Ontology (BKOnto) defined four entities and fifteen properties centered on events. Zou and Park (2018) developed the Event-based Ontology for Historical Description (EOHD), which encompassed six entities, such as spatiotemporal and object change events. Krieger and Declerck's (2015) Ontology for Biographical Knowledge (BIO) represented actions within the happening, abstract, and object classes.

- Integrative approach. Araújo et al (2018) designed an ontology for the Museum of the Person (OntoMP) that correlated people and events, detailing ancestry, progeny, and professional endeavors through 30 concepts and 11 properties, including education, migration, and accidents.

The practical applications of biography ontology include: (a) semantic annotation and association. The BiographyNet Schema (Ockeloen et al, 2013) extracted characters, historical events, places, and eras from 125,000 biographies to generate a semantic knowledge base. The Bio CRM Ontology modeled characters' roles, time spans, so-

cial relationships, and historical events in the University of Oxford's Early Modern Letters Online (EMLO) project; (b) data analysis and visualization. The Bio CRM ontology served as a descriptive framework for 13,100 Finnish biographies on the BiographySampo website (<https://seco.s.aalto.fi/projects/biografiasampo/en/>), facilitating life map visualization, field retrieval, knowledge inference, and demographic data computation. Additionally, it enriched and analyzed 13,000 biographies in the Finnish National Biography Semantic Network and 19,137 biographies of Nossi alums through semantic enrichment, quantitative analysis, and visualization and; (c) knowledge mining. The Ontology for Biographical Knowledge (BIO) (Krieger and Declerck, 2015) supported cross-linguistic data mining and integration in the EU TrendMiner project by organizing and semantically linking individuals' biographies and life experiences.

### 2.2 Research on Digital Storytelling

Digital storytelling, which originated at the Centre for Digital Storytelling in the 1990s (Lambert, 2013), refers to the methods and practices of using digital technologies and media for storytelling (Nosrati and Detlor, 2022). Advances in digital technologies and media integration have broadened the application of digital storytelling to cultural heritage (Valtolina, 2016), education (Wu and Chen, 2020), museums (Perouli, 2021), and social memory (Tie et al, 2023, pp. 73). Digital storytelling involves cross-media, multimedia, and interactive storytelling, characterized by narrativity, mediality, and interactivity (Miller, 2019). As a digital humanities method, digital storytelling promotes knowledge association, reconstruction, discovery, presentation, and dissemination through semantic web, computational, and multimedia technologies (Fu et al, 2023, pp. 1525). Digital storytelling enhances user experience and understanding, as well as value creation.

Current research in digital storytelling concentrates on its elements, frameworks, and assessment. Two dominant theories outline digital storytelling elements: the three-element theory—data, storytelling, and visualization (<https://powerbi.microsoft.com/en-us/data-storytelling/?cdn=disable>)—and the five-element theory—central idea, exploratory points, linear structure, storytelling techniques, and visualization (Dykes, 2015). Frameworks refer to methodologies for producing and presenting digital stories, including: ① Product design method, which mirrors the product design process and segments digital storytelling into ideation, investigation, reporting, and refinement (Planer and Godulla, 2021); ② Data-centered method, emphasizing data analysis insights and categorizing the storytelling process into digitization, datafication, and aestheticization—converting physical entities to digital, extracting data value, and disseminating through media and technology; ③ Storytelling-centered method, which involves selecting storytelling elements, choosing scenes,

designing plots, creating stories, and artistically presenting them (Meier, 2022). In addition, several methods are used to assess the quality of digital stories and define them as successful or unsuccessful. These methods focus on plot design, the pacing of storytelling, audio soundtrack and emotional content of the digital stories. Factors such as emotional content of the digital stories, and interactivity, quality of images and visual impact were also given attention (Malkawi et al, 2019).

In summary, existing research primarily conceptualizes biographical knowledge with the core concepts of characters and events; however, the definitions of entities and properties are simple. Moreover, there has been insufficient attention to other types of historical materials related to the biographical subject, such as photographs, manuscripts, and certificates. The current biography ontologies are primarily applied to semantic association, data analysis, and knowledge mining, yet their role in advancing biographical narrative and presentation remains limited. Accordingly, this study employs the *Autobiography* as a case study, articulating its requirements for knowledge description and extracting biographical knowledge elements in order to construct an ontology that enhances the granularity of knowledge representation and interrelates diverse material types. Building on this foundation, the study integrates thematic clustering and visualization techniques to explore the methodological pathway for biography digital storytelling, involving the thematic clustering of biographical content, the linkage of storytelling threads, and the visualization of storylines. The goal is to foster innovation in the narrative structures of biographies, enhance the user experience, and enrich users' comprehension of the evolution of library science in China.

### 3. Research Design

This research is grounded in Ryan's storytelling theory (Ryan, 2018) and structuralist narratology (Yang et al, 2022), which together provide the theoretical foundation for the study's design. Ryan proposed a data-centered and phased theory with a focus on data analysis and insights, dividing storytelling into four stages: data, analysis, insight, and story. Structuralist narratology, focusing on data processing, knowledge organization, and visualization, outlines three essential elements: fact, organization, and presentation. These frameworks integrate data objectivity with storytelling narrativity, offering significant theoretical and practical contributions to digital storytelling.

Based on the above theories, this study divides the biography digital storytelling into four steps: data, analysis, insight, and storytelling. Specifically, the data phase is to analyze the characteristics of multimodal biographical data. The analysis phase employs semantic web technologies, such as ontology, to reconstruct biographical knowledge entities. The subsequent insight phase clusters biographical topics using LDA, and the storytelling phase

employs visualization technologies to deepen narrative engagement and foster innovation in user experience. Following these steps, this study segments biography digital storytelling into four key steps: selection and characteristics analysis of storytelling materials, extraction and semantic association of storytelling elements based on ontology, biography themes and storytelling threads connection utilizing LDA, and visualization of biography digital storytelling using storylines, as depicted in Fig. 1.

## 4. Selection and Characteristics Analysis of Biography Digital Storytelling Materials

### 4.1 Selection of Storytelling Materials

The *Autobiography* was selected as the primary storytelling material for this study. *Autobiography*, published by the National Library of China Publishing House in 2020, was derived from Mr. Peng Feizhang's oral accounts, as transcribed and edited by Peng et al (2020). Drawing on in-depth interviews with Mr. Peng, the collators collected and digitized a range of physical materials, such as photographs, letters, certificates, and manuscripts. After verifying the historical accuracy of the materials, they organized them chronologically, resulting in a comprehensive manuscript of approximately three hundred thousand words, including ninety images.

Peng Feizhang is widely recognized as a distinguished Chinese library scientist, educator, and bibliographer with significant contributions spanning several decades. Beginning his career as a junior academic during the early years of the People's Republic of China, Peng later studied in the Soviet Union. Upon returning to China, he became the inaugural professor of library science, and subsequently held multiple pivotal roles, including serving as the first dean of China's first School of Library and Information Science, the country's first doctoral supervisor in information science, and the convener of the disciplinary review group of the State Council Academic Degrees Committee. Mr. Peng's contributions have profoundly influenced the academic landscape and institutional development in library science. The digital storytelling of his *Autobiography* not only captures the development of Chinese librarianship but also serves as a valuable means of studying the history of library science education and promoting the ideals embodied by exemplary figures. With the publisher's permission, the complete verbatim manuscript of the *Autobiography* was obtained and utilized in this study.

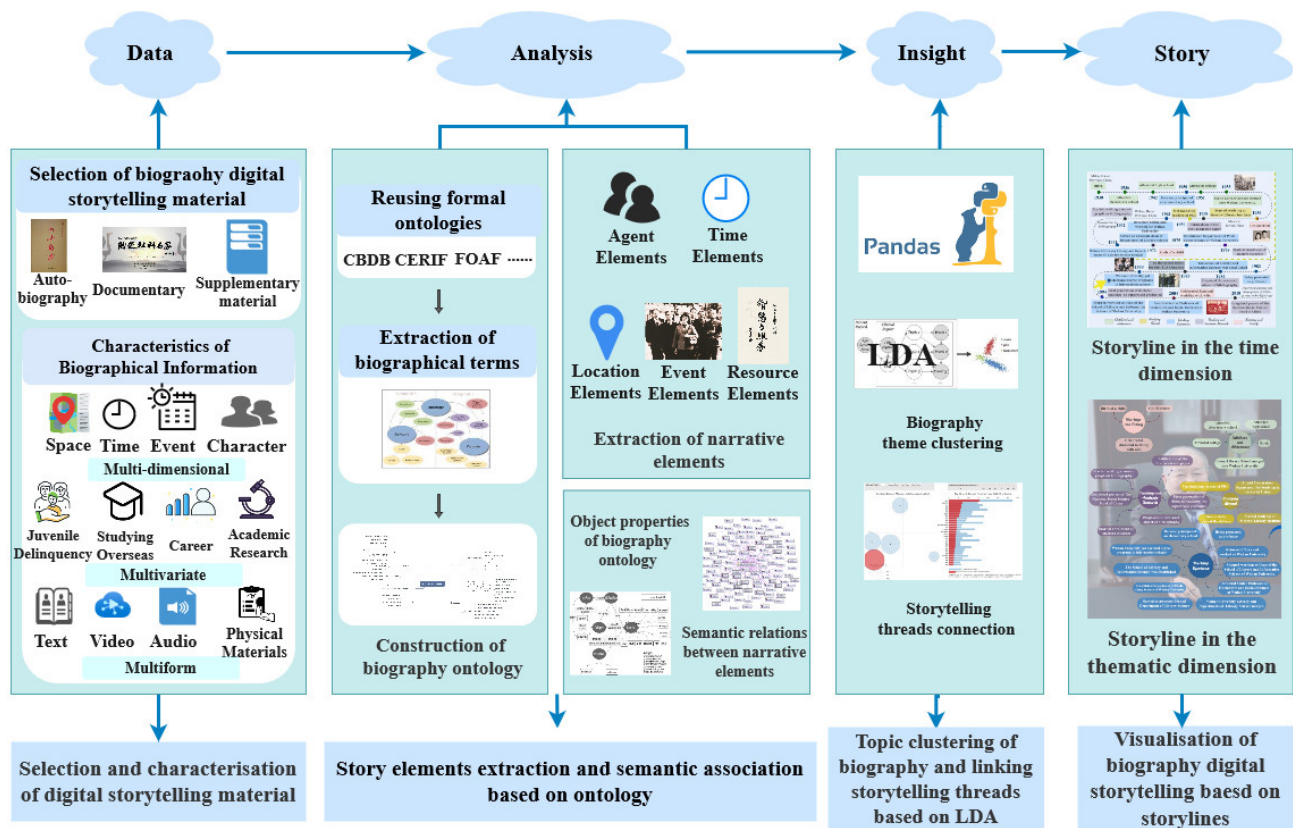
### 4.2 Analysis of the Characteristics of Storytelling Materials

This study characterizes the storytelling materials along three dimensions: content, type, and utilization.

#### 4.2.1 Content Characteristics of the Storytelling Materials

The *Autobiography* contains a substantial amount of information regarding individuals, events, locations, and





**Fig. 1. A research framework for biography digital storytelling.** CBDB, China Biographical Database Project; CERIF, Common European Research Information Format; FOAF, Friend of a Friend; LDA, Latent Dirichlet Allocation.

physical materials. Furthermore, the *Autobiography* provides a detailed account of significant events and periods in the biographical subject's life, offering readers a deeper insight into his professional and personal trajectory. Nevertheless, the *Autobiography*'s linear narrative structure and textual format present a challenge in elucidating the interconnections between biographical knowledge entities, such as characters, time, and events. It is, therefore, necessary to employ ontology to associate knowledge entities, use theme clustering technologies to facilitate readers' comprehension and utilize visualization techniques to present the biographical subject's deeds intuitively.

#### 4.2.2 Typological Characteristics of the Storytelling Materials

The *Autobiography* comprises unstructured textual data, and other types of materials presented as images such as newspapers, photographs, archives, and maps—each type of material presents information in a distinct manner. Texts, newspapers, and archives are presented in a two-dimensional format, while photographs and maps provide spatial context by expanding into three-dimensional representations (Tie et al, 2023). While the diversity of material types broadens the scope of biographical knowledge representation, they also increase the diversity and complexity of storytelling. The effective sharing and dissemination of sto-

rytelling materials could be facilitated by a unified knowledge framework that systematically integrates various material types, ensuring an accurate restoration of historical context and a faithful representation of the subject's experiences.

#### 4.2.3 Utilization of Storytelling Materials

Biography storytelling materials have mainly been presented in print format. The accessibility of the *Autobiography* is still limited to basic reading and retrieval functions, with insufficient capacity for more nuanced knowledge extraction from a humanistic perspective. Biography storytelling materials are rich in modality, correlation, mutual evidence, and connotation. By organizing them through a fine-grained knowledge structure, applying thematic clustering to clarify event sequences, and visualization technologies for intuitive data presentation, it is possible to enable in-depth analysis and revitalization of biographies.

### 5. Extraction and Semantic Association of Storytelling Elements

The extraction and semantic association of storytelling elements are processes of revealing valuable in-depth information in biographies. The initial step involves designing the biography ontology using well-established construction methods, followed by the systematic extraction and associ-

**Table 1. Object properties of the biography ontology.**

Type	Object property/Relation	Definition
Spatio- temporalrelation	time:hasTime, time:hasTemporalDuration	Temporal characteristics of agents, resources, and events.
	shl:nativePlace, shl:burialPlace, shl:region, crm:tookPlaceAt, createdAt/publishedAt, hasBeenTo	Geographical characteristics of agents, resources, and events.
Person- Organisation relation	rel:parentOf, rel:siblingOf, rel:childOf, rel:friendOf, rel:employedBy/rel:employerOf, rel:mentorOf, rel:classmateOf, rel:grandchildOf, rel:worksWith, rel:mentorOf, rel:spouseOf, rel:acquaintanceOf	Relations among relatives, colleagues, and partners.
	org:hasMember, org:headOf, org:subOrganizationOf	Organizational structures: members, leaders, sub-institutions.
Action relation	studyAt	Educational affiliations between individuals and organizations.
	workAt	Employment affiliations between individuals and organizations.
	rel:collaboratesWith	Cooperative interactions among agents.
	org:originalOrganization, org:resultingOrgani- zation	Organizational transformations: mergers, splits, renamings, relocations.
	crm:wasInfluencedBy	Interactions among events, agents, and resources.
	shl:creatorOf	Agents' activities: writing, photographing, designing websites and audiovisual content.
	bibo:editor	Agents publishing and disseminating scholarly and digital media content.
	publisherOf	Agents publishing and disseminate monographs, essays, websites, audiovisual documents.
	bibo:translator	Agents translating scholarly texts.
	bibo:director	Agents organizing and directing audiovisual resources.
	sign	Contracts executed among agents.
	leaderOf	Collaborations between agents and academic projects.
	crm:hadParticipant	Interactions between agents and events.
	isDescribedBy	Descriptions linking events and resources.
	crm:occurredInThePresenceOf	Materials featured in events.
	concurWith	Concurrent events.
	followedBy	Sequential events: before and after.
	causedBy	Causal relationships among events.
	postedIn	Associations between academic papers and journals.
	outcomeOf	Collaborations on academic papers and monographs.

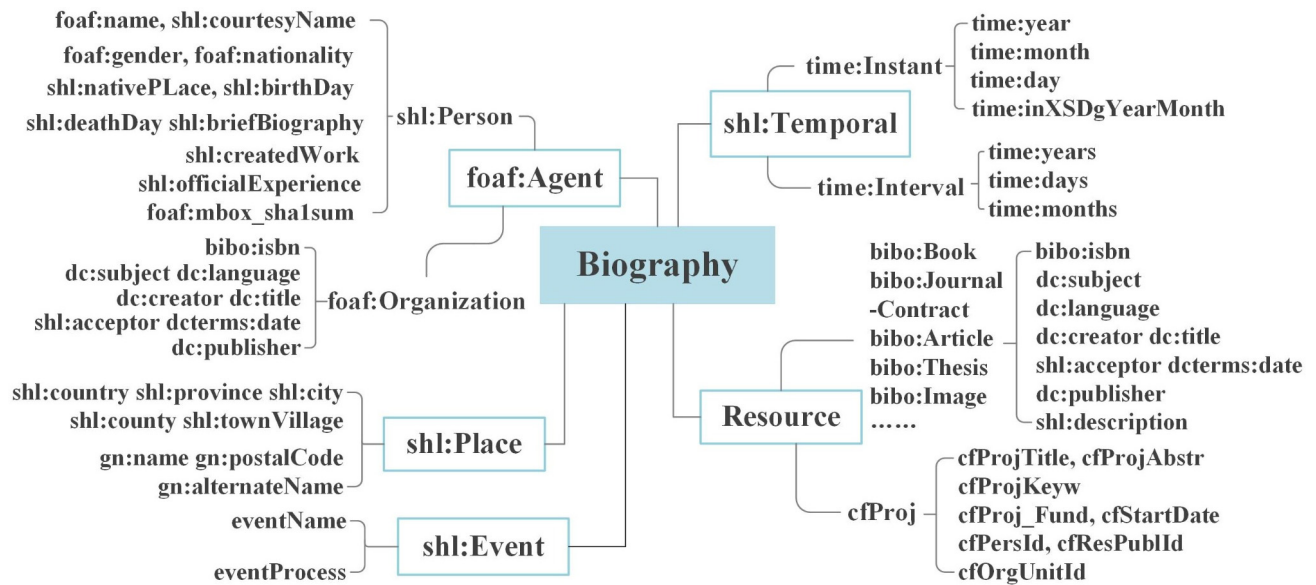


Fig. 2. Classes and data properties of biography ontology.

ation of storytelling elements, such as persons, time, location, event, and resource.

### 5.1 Construction of Biography Ontology

The construction of the biography ontology is informed by mature methodologies, notably the seven-step approach (Noy and McGuinness, 2004) and the IDEF5 method (Integrated Definition for Ontology Description Capture) (Koji et al, 2004). The construction of biography ontology encompasses five key steps:

#### 5.1.1 Defining the Domain and Scope of Biography Ontology

Developed from the *Autobiography*, the biography ontology serves as a structured framework for organizing and revealing the biographical subject's essential details, life experiences, thoughts, emotions, and associated resources. It provides a reusable, scalable, and shareable model for biographical knowledge organization, supporting multi-dimensional narrative expression and serving as a foundation for the systematic extraction and semantic association of storytelling elements.

#### 5.1.2 Reusing the Existing Ontologies

Existing ontologies offer a standard and extensible semantic framework for multi-domain information integration (Tibaut and Guerra de Oliveira, 2022), enhancing interoperability between biography and other domains. By analyzing related biographical articles and the *Autobiography*, as well as investigating existing ontologies, this study takes CBDB, a linked data system ontology for the China Biographical Database Project (CBDB) as the main ontology to be reused.

First, by investigating the biographical articles for librarians and library scientists (for example, the Encyclopedia of the International Society for Knowledge Organization (ISKO)), it was found that the biographer's birth, education, family, career, style of writing and impact are the focus of attention.

Second, the *Autobiography* chronicles significant events and relationships in Mr. Peng's life, such as early experiences, work, studies in the Soviet Union, teaching and academic research, and family life, supplemented by manuscripts, photographs, and certificates. This suggests that people, organizations, time periods, events, and related scholarly materials should be included as primary knowledge entities.

Third, investigate and search for existing ontologies that can describe biographical knowledge entities. CBDB establishes various relationships among knowledge entities in biographies, covering core concepts such as shl:Person, shl:OfficialEvent, shl:Temporal, and shl:Place. Consequently, we reused these core concepts in CBDB and identified people, organizations, time, space, events, and resources as core knowledge entities in the biography ontology.

#### 5.1.3 Extracting Conceptual Terms

Based on the core classes identified in 5.1.2, the *Autobiography* was processed through data cleaning, segmentation, lexical annotation, and named entity recognition using the NLP tools *jieba* and *NLPIR-Paaser* to extract conceptual terms. The process comprises the following steps:

- Data cleaning. After removing the preface and postscript, a list of Chinese stopwords provided by the Intelligent Technology & Natural Language Processing Lab of Harbin Institute of Technology, the Machine Intelligence

Laboratory of Sichuan University, and Baidu is imported. Auxiliaries, conjunctions, prepositions, and quantifiers are subsequently removed.

- Segmentation. Initial segmentation was performed using *jieba*. However, this method proved insufficiently accurate for identifying and segmenting proper nouns. Consequently, *Natural Language Toolkit (NLTK)* was used to calculate bigrams of the segmentation results. The phrases with the ten highest co-occurrence frequencies were then merged with the original segmentation results, and these were subsequently refined and optimized.

- Lexical annotation and named entity recognition. The entity recognition function of *NLPIR-Paesar* and the lexical annotation function of *jieba* were used to annotate entities, including individuals, locations, temporal references, verbs, nouns, and organizations.

#### 5.1.4 Designing Classes and Properties

- Definition of core and extended classes. Drawing on the analysis in Sections 5.1.2 and 5.1.3, the core classes of the ontology are identified as foaf:Agent, shl:OfficialEvent, shl:Temporal, shl:Place, and Resource. Additionally, 19 extended classes are established, referencing Friend of a Friend (FOAF), shlgcn, Time, The Bibliographic Ontology (BIBO), and The Common European Research Information Format (CERIF), to further extend and complement the core classes.

- Design of data properties. Data properties are used to characterize data, providing detailed descriptions of the core and extended classes. The data properties are defined by referencing established ontologies pertaining to people, organizations, time, places, and bibliographic data. For entities lacking appropriate existing ontologies or vocabularies, a custom approach is undertaken to address specific application requirements. Ultimately, 53 data properties are defined to comprehensively describe the ontology's classes, as shown in Fig. 2.

- Definition of object properties. Object properties capture the relationships between various classes, reflecting the semantic connections present among biographical knowledge entities. Drawing on established ontologies such as FOAF, CBDB, BIBO, shlgcn, shlmrc, ORG, REL, and CIDOC-CRM, we incorporate these ontologies alongside custom extensions designed to meet the unique requirements of biographical knowledge. Ultimately, 46 object properties are defined, as shown in Table 1.

#### 5.1.5 Creating Instances

Following the specified classes and data properties in the biography ontology, the corresponding instances in the *Autobiography* are labeled and encoded to facilitate the extraction of digital storytelling elements (see Section 5.2 for details). The object properties subsequently enable the establishment of semantic associations between storytelling elements, thus enhancing the narrative coherence within the

biographical knowledge representation (see Section 5.3 for further explanation).

### 5.2 Extraction of Storytelling Elements

This phase entails extracting storytelling elements—agents, time, locations, events, and resources—utilizing NLP technology, grounded in the core, extended classes, and data properties defined within the biography ontology.

#### 5.2.1 Agent Elements Extraction

The agent elements delineate persons and organizations within the *Autobiography*. Persons (Mr. Peng's relatives, mentors, classmates, colleagues, friends, and students) and their engagements during Mr. Peng's adolescence, academic, and professional phases are documented, with details on names, genders, origins, and educational qualifications. Organizations, such as educational institutions and government bodies associated with his career and education, are characterized by their names and locations. Using the entity recognition function of *NLPIR-Parser* and the lexical annotation function of *jieba*, we extracted 322 personal and 95 organizational elements.

#### 5.2.2 Time Elements Extraction

The time elements reveal temporal characteristics of agents, events, and resources, segmented into time instants and intervals. Time intervals represent the durations of events, tenures, and organizational lifespans, measured in years and months. At the same time, time instants pinpoint times like births, deaths, publications, and project completions (measured in years, months, and days). Using the lexical annotation function of *jieba*, we extracted 178 time elements.

#### 5.2.3 Location Elements Extraction

The location elements detail the geographic characteristics of agents, events, and artifacts in the *Autobiography*, such as birthplaces, venues of significant events, and photography locations. Through the entity recognition function of *NLPIR-Parser*, we identified 45 location elements.

#### 5.2.4 Event Elements Extraction

The event elements chronicle activities within the *Autobiography*, including educational experience, university working experience, academic research, international collaborations, and family. Since most of the titles in the *Autobiography* are named after important events, this study first extracts the important events in the titles, and then identifies other related events by manual labelling, resulting in 131 event elements.

#### 5.2.5 Resource Elements Extraction

The resource elements enumerate physical resource, scholarly accomplishments, and research projects in the *Autobiography*, comprising 14 categories such as manuscripts,



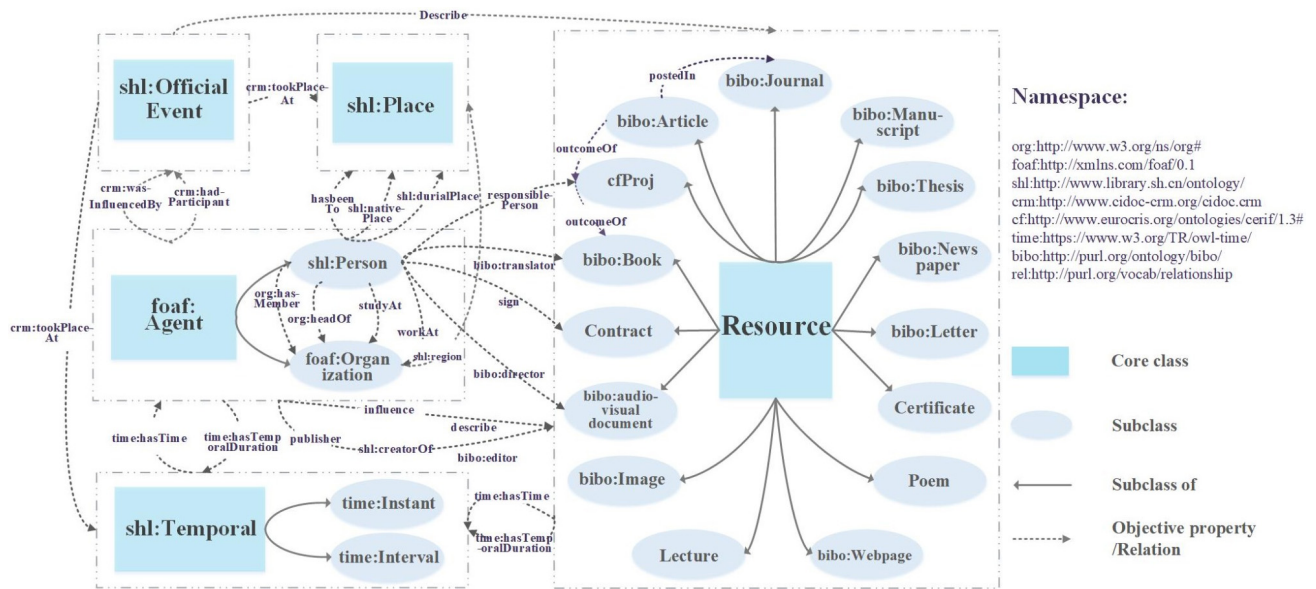


Fig. 3. The classes and their relationships of the biography ontology.

correspondence, contracts, and scholarly articles, which are described by identifiers, keywords, and titles. Through the entity recognition function of *NLPIR-Parser*, we extracted 325 distinct resource elements.

### 5.3 Semantic Association of Storytelling Elements

Semantic association involves a detailed exploration of biographical knowledge and a comprehensive understanding of biographical content. It aims to sequentially organize the text by linking storytelling elements to form story units, serving as the basis for constructing storytelling plots and threads. Ontology offers a detailed and structured framework for knowledge modeling and representation. This section details the associations among storytelling elements through the object properties defined in the biography ontology. The storytelling elements and their interrelationships are illustrated in Fig. 3.

## 6. Topic Clustering of Biography and Linking Storytelling Threads

Storytelling elements are extracted and linked to form character-centered or event-centered story units, each containing independent and complete storytelling components. These units can be connected through content-focused, logical, and clear storytelling threads to enhance the expressive power of biography digital storytelling and enable users to access relevant information based on thematic categories efficiently. In this study, thematic clustering of the *Autobiography* is conducted using LDA, enabling story units to be linked through specific topics, thus forming coherent storytelling threads. This approach seeks to provide a streamlined and targeted exploration path, thereby enhancing the user experience.

### 6.1 Topic Clustering of Biography

LDA is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Blei et al, 2003). By giving the topic probability distribution of each document in the corpus, LDA enables text mining, topic discovery and clustering. In this study, the data is stored as a corpus, which is preprocessed as described in Section 5.1.2 after data cleansing and segmentation. Each natural paragraph is treated as a separate document, resulting in a total of 1169 documents. The terms in the documents are converted into a word frequency matrix, and LDA is applied to model the topic distribution. The analysis yielded a different number of topic clustering results and their Coherence Score, as shown in Fig. 4.

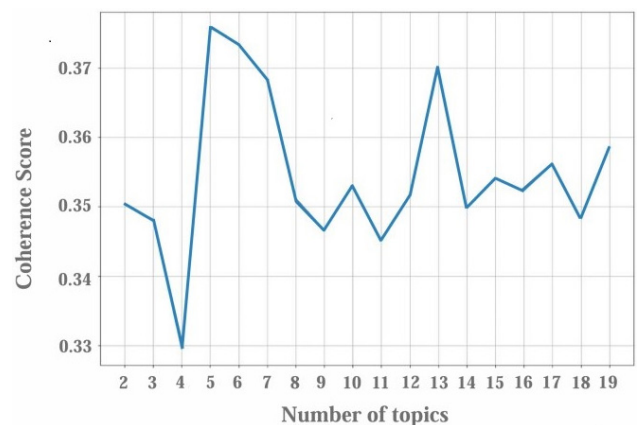


Fig. 4. The Coherence Score corresponding to different number of topics.



The Coherence Score is a key metric for determining the number of topics. The underlying principle is that a high frequency of keyword co-occurrence within each topic corresponds to a high Coherence Score, indicating strong intra-topic correlation, high text interpretability, and effective clustering. The Coherence Score for various clustering results is computed using the U<sub>mass</sub> method, with the findings presented in Fig. 4. The analysis reveals that the highest Coherence Score occurs when the number of topics is set to 5, indicating the optimal clustering configuration. Consequently, the Autobiography is segmented into five topic modules.

## 6.2 Linking Storytelling Threads

As an indispensable component of narrative, the theme represents the central idea of the storytelling text. It can be presented through storytelling threads, linking storytelling units and forming a coherent storyline. Following the thematic clustering analysis, the *Autobiography* is divided into five thematic modules, which form the storytelling threads. The associated story units are then systematically organized and linked across distinct categories. This process encompasses the following two primary steps.

First, the intertopic distance map (Fig. 5, left) and the most relevant terms for topics (Fig. 5, right) are presented based on the clustering results. In the intertopic distance map, higher inter-theme similarity values indicate that the circles representing the topics are closer together, signifying greater similarity in the content of those topics. The underlying principle is that each topic is represented as a vector, and the cosine of the angle between two topic vectors is computed to derive the inter-topic similarity. A higher similarity value reflects more substantial thematic similarity. Fig. 5 shows that the similarities between topics 1, 2, and 3 are relatively high and displays the most relevant terms for each topic, aiding in the determination of topic content. Due to space limitations, only the relevant terms for topic 1 are presented.

Second, determine the content of each topic module, and organize the storytelling threads. In conjunction with the content of the *Autobiography*, five thematic modules are identified: childhood and adolescence, working, teaching, studying abroad, scientific research, and family. Based on this categorization, five storytelling threads are developed to link story units: ① Working experience, including employment at Wuhan University, promotion to professor; ② Studying abroad, including the events such as arrival at the Union of Soviet Socialist Republics (USSR), studying under Prof. Alexander Davidovich Eichengolts, and completing the thesis defense; ③ Teaching and academic research, including starting enrolment of Masters students, proposing the research object of bibliography, publishing the first academic paper, and completing project of the National Social Science Fund of China; ④ Childhood and adolescence, covering events such as birth, attending primary school, high

school and Boone Library School; ⑤ Marriage and family, encompassing marriage, birth of a child, and celebration of a diamond wedding with wife. Story units linked by storytelling threads can be accessed by theme, enhancing the logical flow and plot of biography digital storytelling.

## 7. Visualization of the Biography Digital Storytelling Based on Storylines

Biographies play a crucial role in transmitting social memory and promoting the ethos of role models, serving as an essential bridge between contemporary perspectives and history. However, biographies are typically structured in a linear textual format, which can hinder readers' ability to efficiently identify key themes and extract valuable insights. Therefore, based on the organization of storytelling elements using ontology, it is necessary to apply intuitive visualization to enhance the comprehensibility of biographies. Storylines offer a potentially effective solution. As a visual storytelling technique, the concept of the storyline was first introduced in 2009 by R. Munroe in his hand-drawn illustrations, Movie Narrative Charts (<https://xkcd.com/657/>). These charts used lines to depict characters' experiences and the start and end of events (Dehghani and Asadpour, 2019), thereby clarifying the chronological sequence and semantic relationships among events. Storylines act as joint probability distributions, capturing elements such as events, time, characters, topics, and keywords. They utilize timelines and apply events as memory storage units (Chao and Zhang, 2019, pp. 64), conveying humanistic information intuitively and interactively, thereby enhancing the content's comprehensibility, memorability, and experiential quality. Consequently, storylines can serve as a valuable presentation tool for biography digital storytelling.

In this study, storytelling elements are visualized, integrated, associated, and presented across temporal and thematic dimensions based on extraction and semantic association. In the temporal dimension, storytelling elements are organized chronologically to form a timeline (see Section 6.1 for details). In the thematic dimension, the *Autobiography* is segmented into five thematic modules to form thematic lines (see Section 6.2 for details).

### 7.1 Construction of Biography Storyline from Temporal Dimension

The storyline is designed to integrate complex data in a visually intuitive manner (Chao and Zhang, 2019, pp. 70) and to organize storytelling elements chronologically. For the event elements extracted in Section 4, significant events are initially screened, then sorted and classified based on chronological order and thematic clustering to construct a timeline. The associated storytelling elements are then interconnected to form a cohesive storyline. This approach provides a comprehensive and intuitive representation of the biographical subject's life experiences (see the left side of Fig. 6). Finally, the storytelling texts are composed in the

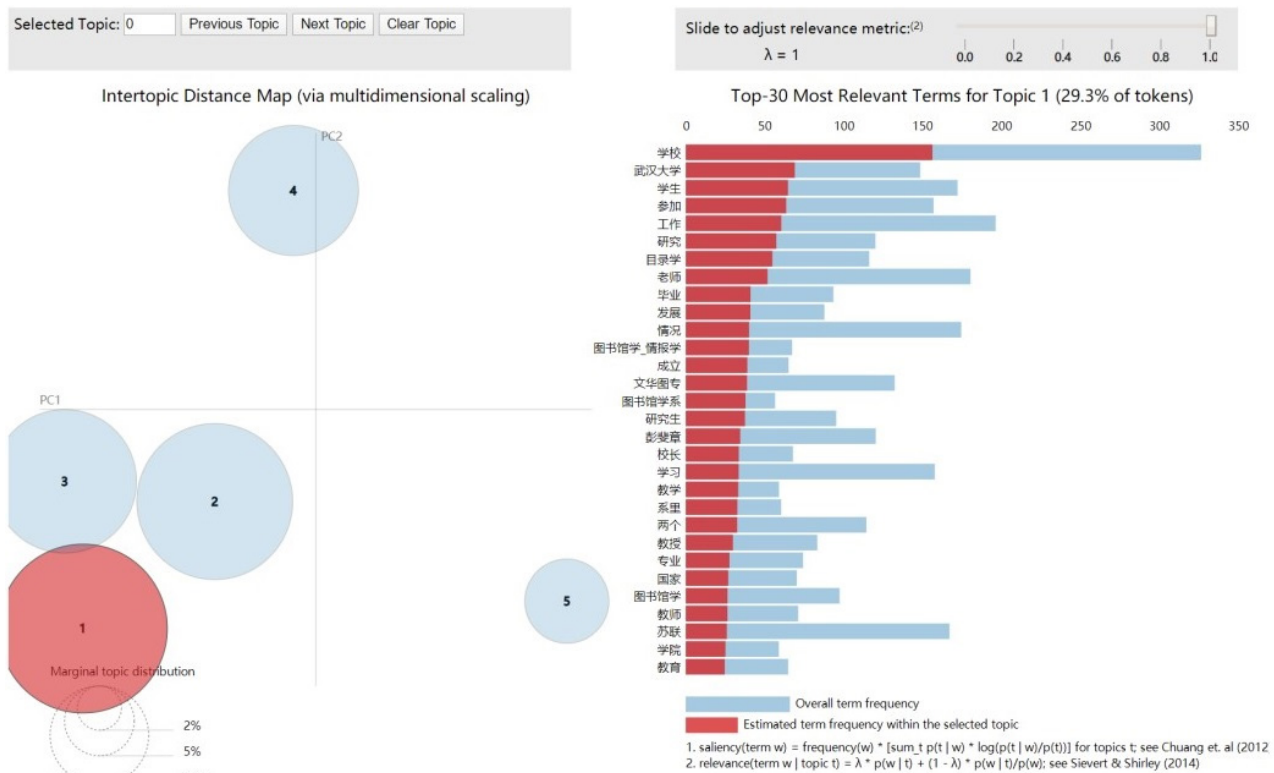


Fig. 5. Topic clustering results for the *Autobiography*.

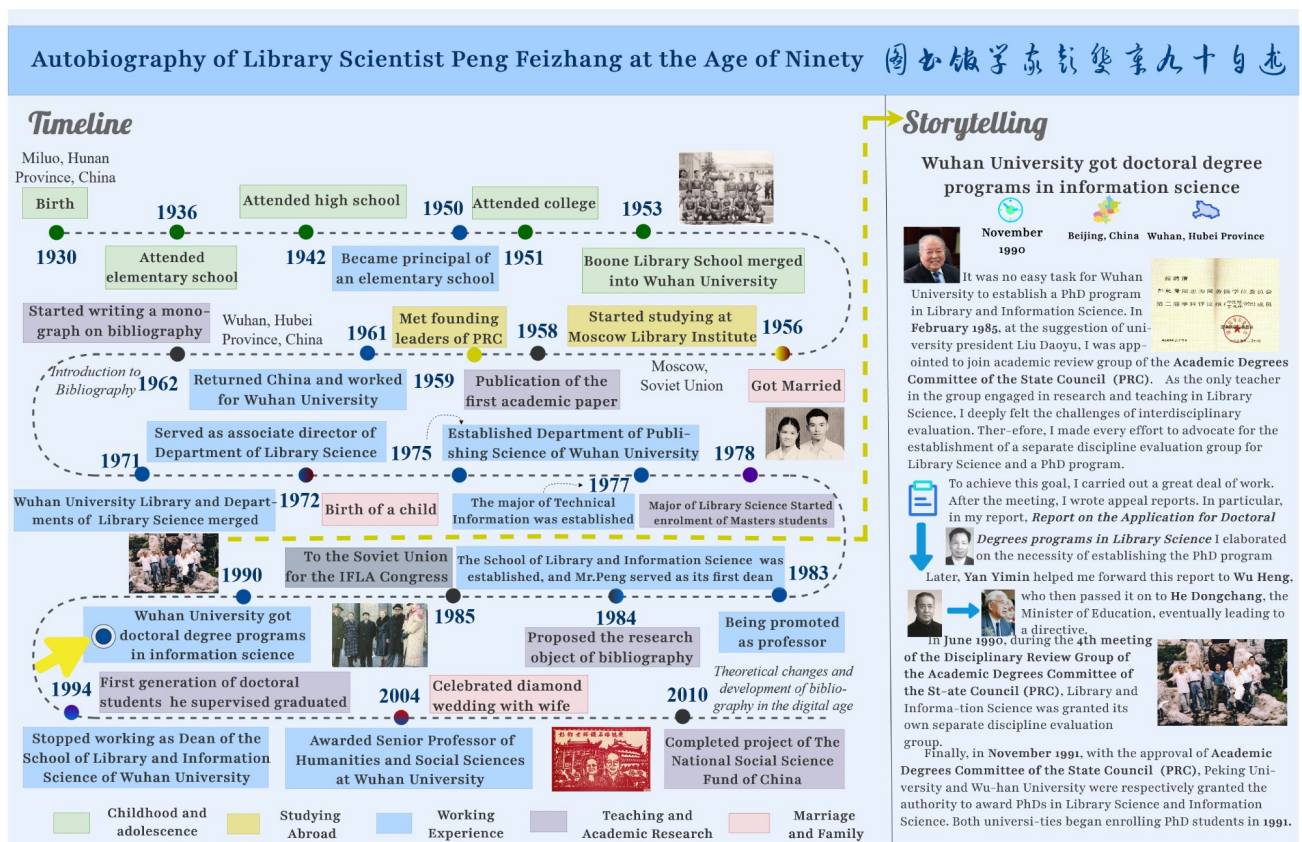


Fig. 6. Biography storyline from temporal dimension.

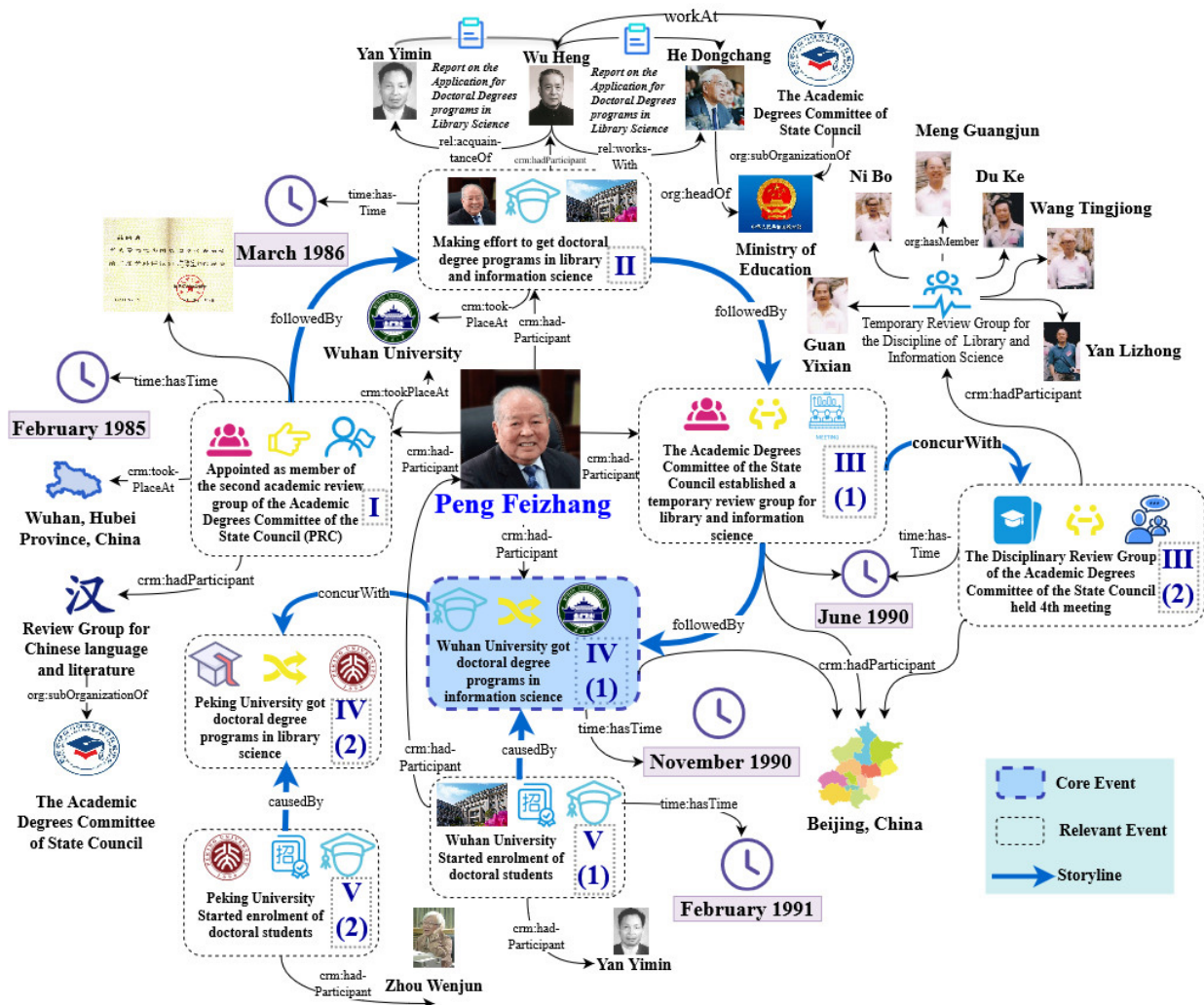


Fig. 7. A node of Storyline.

first person for each node in the storyline, enabling users to access detailed information about the characters, organizations, time, locations, and the causes and consequences of the events by clicking on the nodes. This facilitates users to foster a deep and intuitive understanding and empathy towards Mr. Peng's experience. Due to space constraints, this study uses the node "Wuhan University got doctoral degree programs in information science" (illustrated by the yellow arrow on the left of Fig. 6) as an example, as shown on the right of Fig. 6.

Each node within the storyline is developed using storytelling elements—such as persons, organizations, time, location, and events—thereby forming the storyline based on the associations between these elements. For brevity, we use the node "Wuhan University got doctoral programs in information science" as an example to display storytelling, as shown in Fig. 7. The node revolves around Mr. Peng, with the blue box representing the core event, white boxes symbolizing related events, numbers within the boxes signifying the stages of events, and lines denoting causal, sequential, and co-occurring relationships.

The storyline illustrated in Fig. 7 contains the following key events: In February 1985 (Phase I), Mr. Peng was appointed as a member of the Second Discipline Review Group of the Academic Degrees Committee of the State Council, and was assigned to the Chinese Language and Literature Group due to the lack of a separate review group for the discipline of library science, and thus he resolved to strive for a separate disciplinary review group and doctoral degree programs; In March 1986 (Phase II), he wrote *The Report on the Application for Doctoral Degrees in Library Science*, which was forwarded to Wu Heng at the Academic Degrees Committee via Yan Yimin, and then to the Minister of Education, He Dongchang, who provided a response. In June 1990 (Phase III), the Academic Degrees Committee of the State Council formed a distinct academic review group for library and information science (Phase III(1)) and held its fourth meeting in Beijing (Phase III(2)). During this meeting, Mr. Peng, along with six other experts—Ni Bo, Du Ke, Meng Guangjun, Guan Yixian, Yan Lizhong, and Wang Tingjiong—discussed key issues regarding master's and doctoral programs in library and information sci-





Fig. 8. Biography storyline from thematic dimension.

ence. In November 1990 (Phase IV(1)), with the approval of the Academic Degrees Committee of the State Council, Wuhan University was granted the right to offer doctoral programs in information science (Phase IV(1)), and Peking University was granted the right to offer doctoral programs in library science (Phase IV(2)). In 1991, Wuhan University commenced the enrollment of doctoral students in Information Science (Phase V). Fig. 7 links events chronologically and associates relevant storytelling elements through the classes and properties defined within the biography ontology.

## 7.2 Construction of Biography Storyline from Thematic Dimension

Linking story units and associating storytelling elements with threads facilitate the extension of these units towards the core of the *Autobiography*. This approach supports the deep mining of storytelling materials, the identification of relevant historical resources, and the formation of a storytelling perspective on the development of library science education. To construct the storyline from multiple thematic dimensions, we divide the *Autobiography* into five thematic modules: adolescence and childhood, studying abroad, working experience, teaching and academic re-

search, and marriage and family. This segmentation was derived from storytelling threads generated through thematic clustering in Section 5, with the modules arranged based on the similarity between thematic clusters (as depicted on the left side of Fig. 8).

On the left side of Fig. 8, users can select a module to view the corresponding storyline. To illustrate, upon selecting “working experience”, the upper right section of the interface displays a first-person storytelling text about Mr. Peng’s career, incorporating relevant characters, institutions, and temporal and spatial elements. Deeper insights into a specific event can be further examined in the lower right corner of the interface, where the specific course of action is outlined, while images, manuscripts, and other relevant resources can be accessed. Due to space constraints, the author uses the example of “Wuhan University got doctoral degree programs in Information Science” (as shown on the right side of Fig. 8). Storylines from the thematic dimension are integrated into the storytelling threads, representing biographies across a variety of subjects with granularity and precision, allowing users to quickly identify and gain profound insights into the biographical subject’s life experiences.



## 8. Conclusion

Biography digital storytelling involves reconstructing and interpreting historical and cultural materials through digital technology and media, which primarily entails the analysis, extraction, correlation, organization, and visualization of storytelling elements. Using narratology theory as a foundation, this study explores the realization of biography digital storytelling with the *Autobiography* as a case study, focusing on three primary aspects: ① extracting and associating storytelling elements using ontologies, and constructing semantically rich storytelling units that integrate people, events, places, times, and objects; ② clustering themes using LDA, connecting storytelling units through distinct thematic threads, and promoting the expression of biographies within specific scenarios; ③ utilizing storylines as a visualization tool and designing an interactive interface based on temporal and thematic dimensions, facilitating users' transition from passive reception to active exploration. The contribution of this study lies in: attempting to apply ontology in the implementation of digital storytelling, making it a valuable effort to integrate knowledge organization with digital storytelling; using storylines to present stories, along with the design of timelines and thematic lines, to deconstructs and reconstructs traditional biographical texts, making them more accessible and applicable, thereby enhancing users' understanding of historical figures, events, and culture.

However, this study has limitations. First, it relies on the *Autobiography* as the primary data source and has not yet included relevant memoirs, documentaries, or archival memory works as supplementary sources. Second, the study serves as an exploratory experiment in biography digital storytelling and has yet to establish a fully developed implementation framework or user interface. Future work could combine multi-source and multi-modal resources to validate and refine the implementation path of biography digital storytelling on a larger scale. Additionally, Applications (APPs) and user interfaces could be developed to enhance interactivity and immersion of digital storytelling by leveraging Augmented Reality (AR) and Virtual Reality (VR) technologies.

## Availability of Data and Materials

The data sets generated and/or analyzed during the current study are not publicly available due to copyright restrictions, but are available from the corresponding author on reasonable request.

## Author Contributions

LS and YL designed the research study. LS performed the research. YL analyzed the data. Both authors contributed to editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

## Acknowledgment

Not applicable.

## Funding

The National Social Science Foundation of P.R. China (Project Name: Research on Theories and Practices of Digital Literacy Improvement for All in the Intelligence Age, Project No. 24&ZD179).

## Conflict of Interest

The authors declare no conflict of interest.

## References

- Adhi BP, Widodo W. Development of biography information system based on semantic web using biography ontology: Requirement phase. In Journal of Physics: Conference Series (Vol. 1402, No. 6, p. 066063). 2019. <https://doi.org/10.1088/1742-6596/1402/6/066063>
- Araújo C, Martini RG, Henriques PR, Almeida JJ. Annotated documents and expanded CIDOC-CRM ontology in the automatic construction of a virtual museum. In Developments and Advances in Intelligent Systems and Applications 11th Iberian Conference on Information Systems and Technologies (pp. 91–110). 15–18 June 2016 Gran Canaria Island, Spain. Springer: Cham. 2018. [https://doi.org/10.1007/978-3-319-58965-7\\_7](https://doi.org/10.1007/978-3-319-58965-7_7)
- Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research. 2003; 3: 993–1022.
- Chao L, Zhang C. Data Storytelling: From Data Perception to Data Cognition. Journal of Library Science in China. 2019; 45: 61–78. <https://link.cnki.net/doi/10.13530/j.cnki.jlis.190042> (In Chinese)
- Dehghani N, Asadpour M. SGSG: Semantic graph-based storyline generation in Twitter. Journal of Information Science. 2019; 45: 304–321. <https://doi.org/10.1177/0165551518775304>
- Dupagne M. Story Circle: Digital Storytelling Around the World. Journal of Broadcasting & Electronic Media. 2010; 54: 532–533. <https://doi.org/10.1080/08838151.2010.498704>
- Dykes B. Data Storytelling: What It is and How It Can be Used to Effectively Communicate Analysis Results. Applied Marketing Analytics. 2015; 1: 299–313. <https://doi.org/10.69554/AOPX4542>
- Fu Y, Mahony S, Liu W. Reconstruction of cultural memory through digital storytelling: A case study of Shanghai Memory project. Digital Scholarship in the Humanities. 2023; 38: 1522–1535. <https://doi.org/10.1093/lhc/fqad044>
- Guarino N, Pierdaniele G. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In Mars NJI (ed.) Towards Very Large Knowledge Bases (pp. 25–32). IOS Press: Amsterdam. 1995.
- Gruber T. Ontology. In Liu L, Özsu MT (eds.) Encyclopedia of

- Database Systems (pp. 1963–1965). 1st edn. Springer: Berlin. 2009.
- Koji Y, Kitamura Y, Mizoguchi R. Towards Modeling Design Rationale of Supplementary Functions in Conceptual Design. In TMCE 2004 Conference Proceedings TMCE 2004 Conference (pp. 117–130). 13–17 April 2004 Lausanne, Switzerland. Millpress: Rotterdam. 2004.
- Krieger HU, Declerck T. An Owl Ontology for Biographical Knowledge. Representing Time-Dependent Factual Knowledge. In Proceedings of the First Conference on Biographical Data in a Digital World 2015 Biographical Data in a Digital World (pp. 101–110). 9 April 2015 Amsterdam, The Netherlands. CEUR-WS.org: Aachen. 2015.
- Lambert J. Digital Storytelling: Capturing Lives, Creating Community. Routledge: New York. 2013.
- Leskinen P, Hyvönen E. Linked Open Data Service About Historical Finnish Academic People in 1640-1899. In CEUR Workshop Proceedings CEUR Digital Humanities in the Nordic Countries (pp. 284–292). 21–23 Oct 2020 Riga, Latvia. RWTH Aachen University: Aachen. 2020.
- Malkawi R, Alzaqebah M, Al-Yousef A, Abul-Huda B. The Impact of the Digital Storytelling Rubrics On the Social Media Engagements. International Journal of Computer Applications in Technology. 2019; 59: 269–275. <https://doi.org/10.1504/IJCAT.2019.098605>
- Meier S. Digital Storytelling: A Didactic Approach to Multimodal Coherence. Frontiers in Communication. 2022; 7: 906268. <https://doi.org/10.3389/fcomm.2022.906268>
- Miller CH. Digital Storytelling 4E: A Creator's Guide to Interactive Entertainment. CRC Press: Boca Raton. 2019.
- Motta E, Siqueira S, Andreatta A. Populating a Domain Ontology From a Web Biographical Dictionary of Music-an Unsupervised Rule-Based Method to Handle Brazilian Portuguese Texts. In Proceedings of the Fifth International Conference on Web Information Systems and Technologies - Volume 1: WEBIST the 5th International Conference on Web Information Systems and Technologies (pp. 192–199). 23–26 March 2009 Lisboa, Portugal. SciTePress. 2009. <https://doi.org/10.5220/0001842301920199>
- Nosrati F, Detlor B. The Power of Stories for Impression Management: Evidence From a City Cultural Digital Storytelling Initiative. Information Technology & People. 2022; 35: 1410–1427. <https://doi.org/10.1108/ITP-06-2020-0400>
- Noy NF, McGuinness DL. Ontology Development 101: A Guide to Creating Your First Ontology. 2004. Available at: [https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf) (Accessed: 2 March 2025).
- Ockeloen N, Fokkens A, Ter Braake S, Vossen P, De Boer V, Schreiber G, et al. Biographynet: Managing Provenance at Multiple Levels and From Different Perspectives. In LISC'13: Proceedings of the 3rd International Conference on Linked Science the 3rd International Conference on Linked Science (pp. 59–71). 21 October 2013 Sydney, Australia. CEUR-WS.org: Aachen. 2013.
- Peng F, Ke P, Liu L. Autobiography of Library Scientist Peng Feizhang at the Age of Ninety. National Library of China Publishing House: Beijing. 2020. (In Chinese)
- Perouli C. Digital Storytelling in the Museum: Bringing Cultural Heritage to Life. Crossing Conceptual Boundaries. 2021; 11: 34–46. <https://doi.org/10.15123/uel.8918q>
- Planer R, Godulla A. Longform Journalism in the Usa and Germany: Patterns in Award-Winning Digital Storytelling Productions. Journalism Practice. 2021; 15: 566–582. <https://doi.org/10.1080/17512786.2020.1742771>
- Ryan L. Visual Data Storytelling with Tableau. Addison-Wesley Professional: Boston. 2018.
- Tibaut A, Guerra de Oliveira S. A framework for the evaluation of the cultural heritage information ontology. Applied Sciences. 2022; 12: 795. <https://doi.org/10.3390/app12020795>
- Tie Z, Xia C, Huang W, Zhu Y, Liu M. Event-Driven Interaction Design for Data Visualization: Exploration and Practice Based On “Data Storytelling”. Journal of Library Science in China. 2023; 49: 72–87. <https://link.cnki.net/doi/10.13530/j.cnki.jlis.2023032> (In Chinese)
- Valtolina S. A storytelling-driven framework for cultural heritage dissemination. Data Science and Engineering. 2016; 1: 114–123. <https://doi.org/10.1007/s41019-016-0007-z>
- Wang X, Zhao K. Trust Model for Digital Storytelling of Cultural Heritage: Conception and Framework. Journal of Library Science in China. 2024; 50: 30–41. <https://link.cnki.net/doi/10.13530/j.cnki.jlis.2024011> (In Chinese)
- Wu J, Chen DTV. A systematic review of educational digital storytelling. Computers & Education. 2020; 147: 103786. <https://doi.org/10.1016/j.compedu.2019.103786>
- Yang L, Xu X, Lan X, Liu Z, Guo S, Shi Y, et al. A Design Space for Applying the Freytag's Pyramid Structure to Data Stories. IEEE Transactions on Visualization and Computer Graphics. 2022; 28: 922–932. <https://doi.org/10.1109/TVCG.2021.3114774>
- Yeh J. Towards a Biographic Knowledge-Based Story Ontology System. In Proceedings of the 2018 International Conference on Intelligent Information Technology (pp. 33–38). 26–28 February 2018 Hanoi, Vietnam, New York. Association for Computing Machinery. 2018. <https://doi.org/10.1145/3193063.3193065>
- Zou Q, Park EG. Linking Historical Collections in an Event-Based Ontology. Digital Library Perspectives. 2018; 34: 262–275. <https://doi.org/10.1108/DLP-02-2018-0005>