1-31-2024

Automatic Construction of Ontology with Public-Domain Datasets for Personalized Tutoring with ECA

Asim Jamal
University of Windsor

Automatic Construction of Ontology with Public-Domain Datasets for Personalized Tutoring with ECA

By

Asim Jamal

A Thesis
Submitted to the Faculty of Graduate Studies
through the School of Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada

2024

Automatic Construction of Ontology with Public-Domain Datasets for Personalized Tutoring with ECA

by

Asim Jamal

APPROVED BY:

L. Oriet
Department of Mechanical, Automotive and Materials Engineering

O. Syrotkina
School of Computer Science

X. Yuan, Advisor
School of Computer Science

January 12, 2024

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

E-tutoring systems have transformed remote learning and personalized education, offering potent tools for tailored instruction. The core of a personalized tutor lies in its robust ontology and knowledge base, working seamlessly to deliver captivating educational experiences. These two integral components collaborate to empower the tutor to discern learners' needs, adapt content accordingly, and provide tailored guidance. This study introduces an automated approach for constructing an ontology utilizing publicly accessible datasets, aiming to enhance personalized tutoring through Embodied Conversational Agents (ECA). The objective is to improve the tutoring encounter by delivering bespoke, domain-specific knowledge to learners. The approach harnesses natural language processing techniques to extract pertinent concepts and relationships from open-domain datasets. This extracted data is then leveraged to generate new ontology classes and enhance existing ones. This innovative methodology employs a comprehensive four-step algorithm to develop a domain ontology from a meta-ontology. This resultant ontology serves as a knowledge repository for ECAs, empowering them to provide individualized and contextually pertinent tutoring. This proposed approach presents an automated and scalable solution for constructing ontologies within educational environments.

DEDICATION

I dedicate this thesis to my loving parents, whose unwavering support and boundless understanding have been the pillars of my academic journey. Their encouragement has been my source of strength, and their belief in my potential has fueled my determination. This accomplishment is as much theirs as it is mine, and I express my deepest gratitude for their indefatigable guidance and love. Additionally, I express deep gratitude to my life partner, Nazia, whose enduring patience, unshakeable belief, and generous support have been indispensable throughout my thesis journey. Words fall short in conveying my appreciation for her constant presence and encouragement. Special thanks also go to my brother and sister for their unwavering support as they have always believed in me. Lastly, I extend heartfelt thanks to all my friends whose laughter and camaraderie kept me grounded and joyous throughout.

ACKNOWLEDGEMENTS

I want to sincerely thank the following individuals and institutions for their essential support and contributions during my thesis research: My supervisor, Dr. Xiaobu Yuan, for his continuous guidance, expertise, and steadfast encouragement throughout my research journey. His invaluable advice and support profoundly influenced the trajectory of this work. To my thesis committee members, Dr. Leo Oriet and Dr. Olenna Syrotkina, thank you for their dedication, expertise, and constructive feedback. Their informed critiques and recommendations have enriched the calibre of this research. To the staff and faculty of the University of Windsor, whose facilities, resources, and academic setting were pivotal in facilitating this research. My time at the University was greatly enriched by the opportunities and resources they availed to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NLP	Natural Language Processing
BSH	Belief State History
COCOM	Contextual Control Model
DWT	Discrete Wavelet Transformation
DM	Dialog Manager
ECA	Embodied Conversational Agent
MDP	Markov Decision Process
NCP	Number of Change Points
POMDP	Partially Observable Markov Decision Process
DRL	Deep Reinforcement Learning
SWRL	Semantic Web Rule Language
NLU	Natural Language Understanding
CAs	Conversational Agents
SW	Semantic Web
KE	Knowledge Extraction
DRT	Discourse Representation Theory
LMS	Learning Management System
VR	Virtual Reality
COQA	Conversational Question Answering
MMLA	Multimodal Learning Analytics

CHAPTER 1

Introduction

E-learning utilizes Information and Communication Technology within learning and teaching [5]. Catalyzed by the global disruption caused by the COVID-19 pandemic, personalized tutoring has taken center stage [72]. The pandemic compelled educational institutions worldwide to pivot towards remote and online learning, necessitating a shift toward tailored educational experiences. Instead of non-personalized methods, personalized tutoring takes a more individualized approach to instruction, aligning teaching strategies with each learner's unique requirements. This dynamic approach has yielded demonstrable improvements in educational outcomes, underscoring personalized e-tutoring's impact [70]. This personalized approach to tutoring goes beyond a one-size-fits-all pedagogy. It tailors instruction, content, and methodology to each student's needs, preferences, and abilities. By customizing the learning experience based on individual attributes, personalized tutoring plays a pivotal role in enhancing comprehension and progress. It empowers learners to traverse tailored educational paths and utilize resources that cater to their unique learning styles and proficiency. In an educational landscape profoundly reshaped by the digital age and global challenges, personalized e-tutoring emerges as a beacon for fostering adaptive, efficient, and effective learning outcomes. Personalized tutoring platforms represent a significant advancement in the realm of education. These platforms harness the power of adaptive learning algorithms to tailor the learning experience to each student's unique needs. This personalization extends to adjusting content and difficulty levels, creating a learning path that aligns with individual learning patterns. This adaptability is a transformative feature, enhancing the efficiency and effectiveness of the educational process [58].

One crucial aspect of personalized tutoring systems is the provision of immediate feedback. This feature dramatically aids students in comprehending their strengths and identifying areas requiring improvement. This feedback mechanism goes a long way in fostering a deeper understanding of concepts, allowing learners to make continuous progress [90]. The backbone of many personalized tutoring systems is the Learning Management System (LMS). This platform enables educators to create and administer quizzes and assignments, enhancing active learning. Additionally, the LMS serves as a repository for a diverse range of learning materials, from documents to videos and links, they are making educational resources readily accessible.

In the dynamic landscape of personalized tutoring, software avatars have emerged as interactive virtual mentors. These digital representations engage students in ways that emulate human interactions, significantly elevating the learning experience. Embodied Conversational Agents (ECAs) drive this engagement, promoting learner motivation and active participation. The adoption of ECAs, often in software avatars, is rising in the education sector. Researchers are increasingly addressing the challenges associated with ECAs, aligning them more closely with human behaviour. These avatars possess the unique ability to adapt their communication approach and content presentation to cater to individual student profiles and learning preferences, ultimately providing customized explanations and guidance [87],[43],[46].

This paper delves into the pivotal role played by a robust ontology and knowledge base in personalized tutoring, emphasizing the synergistic collaboration of these components to elevate the educational experience. Integrating a Robust Domain Ontology with the Existing Partially Observable Markov Decision Process (POMDP) for dialogue management is a critical focus, demonstrating how these sophisticated frameworks work in tandem. An innovative aspect of the document is the introduction of an automated method for constructing ontologies, a departure from conventional manual processes. This method, relying on publicly accessible datasets, constitutes a novel contribution to the field. Leveraging advanced Natural Language Processing (NLP) techniques, the paper adeptly extracts pertinent concepts and relationships from open-domain datasets, showcasing a contemporary, data-driven approach to ontology construction and semantic matching. The methodology further distinguishes itself by employing a comprehensive four-step algorithm, facilitating the creation of a domain ontology from a meta-ontology. This algorithmic depth enriches the ontology and ensures ease of implementation across diverse question-answering datasets. Moreover, the paper's emphasis on automation and scalability addresses a crucial need within educational environments, presenting an innovative and practical solution, thereby amplifying its contribution's impact.

In summary, personalized tutoring platforms, supported by adaptive learning algorithms, immediate feedback mechanisms, and innovative tools like software avatars and ECAs, are at the forefront of educational innovation. They enhance education quality and reshape how students learn and interact with educational content.

The subsequent chapters of this thesis are meticulously organized to explore the research comprehensively. Chapter 2 delves into the foundational aspects of Ontology and its application in Dialog Management. Moving forward, chapter 3 meticulously reviews prior research that forms the bedrock of this thesis, contributing significantly to its subject matter. Chapter 4 intricately details this thesis's specific problem, elucidating its noteworthy contributions to its resolution. It further discusses the architectural framework and novel algorithm proposed within this research. Following this, Chapter 5 systematically unfolds the experiments, offering a detailed exposition of implementation nuances through illustrative examples. The outcomes of these implementations are then dissected and presented in Chapter 6, providing a comprehensive understanding of the suggested system's functionality. Lastly, Chapter 7 engages in a thoughtful discussion, critically analyzing the results. It concludes with final remarks and extends the discourse to explore potential avenues for future research.

CHAPTER 2

Literature Review

This chapter centers on prior endeavours in crafting a resilient educational domain ontology capable of engaging in meaningful conversations with users, gauging their knowledge levels, and furnishing relevant hints accordingly. Moreover, it conducts an in-depth examination of contemporary research endeavours about the automated construction of ontologies.

2.1 Personalized Tutoring in E-Learning

In this section, we will delve into the multifaceted realm of E-Learning, exploring its benefits, current practices, and inherent challenges. Additionally, we will scrutinize the pivotal aspect of Personalized Tutoring, addressing the imperative need for personalized educational experiences in the digital landscape. This subsection will further expound on the advantages of personalized tutoring within the e-learning context, examining various approaches and methodologies to tailor educational content to individual learners' needs. Through this comprehensive exploration, we aim to gain a nuanced understanding of the evolving landscape of E-Learning, with a specific focus on the transformative potential and intricacies associated with personalized tutoring methodologies.

2.1.1 E-Learning

E-learning, also known as online learning, represents a specific form of education delivered through the internet and is facilitated by individual instructors or educational institutions, such as universities [23], [36]. The surge in e-learning in recent years can be attributed to various factors, including the widespread availability of the internet [8], advancements in training and learning technologies [21], and the accessibility of social media networks [18]. The global impact of unprecedented events, such as the Coronavirus pandemic, has further accelerated the adoption of e-learning. Measures implemented to curb the pandemic, such as restricting physical gatherings and promoting social distancing, favoured the shift towards online education.

Beyond the pandemic, the cost-effectiveness of e-learning compared to traditional face-to-face learning has contributed significantly to its popularity across the globe [78]. Major e-learning platforms like Coursera and Udemy have emerged as influential providers, offering diverse courses spanning various fields, including information technology, artificial intelligence, computing, management, and health. Like traditional learning systems, e-learners receive certifications across different education levels, ranging from certificates to diplomas and degrees, particularly in practical programs and disciplines. The projected market potential of online learning is expected to reach $325 billion by 2025, signifying a threefold increase from its 2015 levels [40]. This growth underscores the transformative impact of e-learning on the education landscape, emphasizing its role as a dynamic and evolving paradigm with substantial market potential. In the context of my thesis document, this detailed exploration of e-learning establishes its multifaceted nature, touching upon its drivers, impact factors, and future market projections. Contemporary research underscores the pervasive influence of ontology and knowledge-based systems in shaping decision-making processes and providing recommendations. An ontology, an explicit and formal specification of a shared conceptualization, emerges as a pivotal element in this landscape. Within e-learning, ontologies play a fundamental role in modelling and representing domain knowledge, contributing significantly to the structuring and organization of information in educational contexts.

Integrating ontology into recommendation systems, particularly in the context of e-learning, reflects a paradigm shift in addressing critical challenges. Notably, hybrid recommendation systems that amalgamate different recommendation strategies effectively mitigate issues like the cold-start problem, diversity rating concerns, and challenges posed by overspecialized recommendations [28]. The dynamic nature of e-learning environments necessitates sophisticated approaches to recommendation, and the infusion of ontology-based systems introduces precision and adaptability that align with the evolving landscape of educational technologies.

2.1.2 Personalized Tutoring

The historical trajectory of personalized tutoring has established its efficacy in providing invaluable support to students and educators. Noteworthy implementations like Cognitive Tutor in mathematics [10] and AutoTutor for computer literacy [32] have demonstrated successful models. Despite these achievements, developing an effective personalized tutoring system remains a formidable challenge, particularly in addressing students' diverse learning needs and fostering a deeper understanding of complex concepts. Handling this challenge calls for a reevaluation of existing paradigms and a strategic integration of innovative technologies. Recent advancements in natural language processing, notably the emergence of chat-based Large Language Models (LLMs) like ChatGPT [14], present a promising avenue for advancing personalized tutoring systems. These LLMs offer the potential to build upon the foundation laid by Intelligent Tutoring Systems (ITS) and enhance their capabilities by integrating with learning science principles [50], [84].

The need for personalized tutoring is underscored by recognizing that learners possess unique cognitive processes and learning styles. This diversity necessitates a tailored approach to education that goes beyond a one-size-fits-all model. By leveraging LLMs and learning science principles, personalized tutoring systems can offer targeted and adaptive assistance to learners, meeting them at their points of need and promoting a more engaging and practical learning experience. Integrating learning science principles is pivotal in developing ITS, it supports learners' cognitive processes and provides personalized assistance. Learning science principles, encompassing a deep understanding of how individuals acquire knowledge and skills, guide the design of tutoring systems that align with the natural learning trajectories of learners [95], [80]. This holistic approach ensures that personalized tutoring is not merely about customization but is rooted in evidence-based strategies that enhance comprehension, retention, and application of knowledge.

The application of ontologies in e-learning has proven highly effective, serving as a valuable tool for constructing a comprehensive representation of both the learning domain and the student model [57]. Particularly suited for creating student models, ontologies provide a structured framework for representing abstract concepts and attributes, as noted by [93]. To enhance personalized learning experiences, ontologies employ reasoners for knowledge extraction. However, extending inferences and expressing relationships beyond ontological reasoning necessitates incorporating rules, such as Semantic Web Rule Language (SWRL).

In the educational context, ontologies have gained significant importance, representing domain knowledge and tailoring content generation to user preferences [19]. The ongoing research in this field underscores educational ontologies' continuous growth and evolution.

Within this framework, "Dialog" denotes a dynamic, multi-turn interaction between a User and an Embodied Conversational Agent (ECA), facilitating the seamless exchange of information and instructions. Each commencement of a dialogue session triggers the creation of a fresh instance of the Dialog concept. This instance is a dedicated container, encapsulating the ongoing conversation's evolution. The ontology, an integral knowledge base for the dialogue execution algorithm, is collaboratively utilized across the various components of the Dialogue Management system.

2.2 Dialog-Based Tutoring Systems

In Embodied Conversational Agents (ECAs), dialogue managers are pivotal in facilitating natural and interactive communication between users and virtual agents. Functioning as the central component, the dialogue manager is responsible for comprehending user inputs, formulating suitable responses, and orchestrating the flow of conversation. Specifically, the dialogue-manager part is tasked with two primary functions: dialogue modelling, involving the tracking of dialogue states, and dialogue control, entailing decisions on the subsequent system action [9]. The overarching objective of a dialogue manager in ECAs is to enhance interactions by overseeing tasks related to Natural Language Understanding (NLU), Dialogue State Tracking, Dialogue Policy, Response Generation, and Context Management. In essence, dialogue managers in ECAs interpret user inputs, generate contextually relevant responses, manage dialogue states, and uphold context, thereby enabling ECAs to partake in dynamic and interactive conversations, fostering a more natural and human-like interaction experience [34].

Management approaches for dialogue managers are categorized into handcrafted-rule-based approaches and probabilistic (data-driven) approaches. Handcrafted dialogue managers define system states and control through rules set by developers and experts. In contrast, probabilistic dialogue managers learn rules from honest conversations. This thesis delves into probabilistic-based dialogue management approaches, building on the proven success of these methods within our research group [87]. Notably, the Google DialogFlow [42] framework is employed for creating Conversational Agents (CAs). The Google dialogue manager uses Reinforcement Learning (RL) approaches to select the best action, considering user intents derived from the conversation. This involves conceptualizing the conversation as a Markov decision process (MDP) for long-term implications. Previous studies recommend the application of Reinforcement Learning (RL) for goal-oriented dialogue management [81]. Moreover, the utilization of q-learning reinforcement learning (RL) to improve policies [76] was proposed, although the comprehensive methodology or feasibility of this approach is yet to be fully explained. Application of Deep Reinforcement Learning (DRL) to model future rewards in CAs is recommended by [25], wherein informativity, coherence, and ease of answering are integral considerations for calculating agent rewards. A noteworthy example from the Amazon Alexa Prize competition [27] involves an ensemble-based CA utilizing various models, with the dialogue manager employing RL to determine appropriate responses. Training data is sourced from both actual user interactions and crowd-sourcing channels.

2.2.1 Dialog Management

In the landscape of dialogue management for conversational agents, various approaches have been employed [77], each with its distinctive characteristics and considerations. The Finite State Approach relies on finite state machines to model dialogues, leading to meticulously structured conversations that are entirely predefined. However, its lack of flexibility, stemming from predetermined states and transitions, can constrain adaptability to diverse user inputs. In contrast, the Neural Network Approach leverages deep learning models like transformers, enabling more natural responses. Despite its capability for nuanced interactions, this approach demands substantial data for practical training and may exhibit a data-hungry behaviour [25]. The Information Retrieval Approach shares similarities with frame-based methods but incorporates mental states and intentions. It offers deterministic data points, distinguishing them from probabilistic models like POMDP, providing enhanced context at the expense of being limited by deterministic data. Using language models like GPT(Generative Pre-trained Transformer), Generative Approaches excel in response generation but may grapple with challenges such as overgeneration or undergeneration, often requiring fine-tuning for domain-specific applications. Lastly, the Reinforcement Learning approach employs reinforcement learning techniques to optimize conversation actions. Although agents learn through interaction, they may face the challenge of infrequent rewards, requiring a delicate balance between user engagement and the receipt of delayed rewards. These diverse approaches cater to different aspects of dialogue management, reflecting the ongoing pursuit of effective and adaptive conversational agents in various domains.

2.2.2 POMDP and History of Belief States

In the context of this thesis, the Dialogue Manager (DM) adopts the framework of Partially Observable Markov Decision Processes (POMDP), as employed in [76]. To delve deeper into POMDP, this section briefly introduces its fundamental principles and application in handling dialogues. Derived primarily from [35], the outlined concepts elucidate the workings of POMDP. Unlike conventional Markov Decision Processes (MDP), POMDP introduces the concept of observations to determine the system's current state. States represent the system's condition, and actions denote potential maneuvers, such as posing a question. Each action yields outcomes, and immediate benefits are assessed based on the resultant activity. POMDP distinguishes itself by relying on observations to infer the current state probabilistically, creating a belief state that encapsulates this probability distribution. The Belief State History (BSH) is a chronological record of these belief states, crucial for decision-making.

In the integration of POMDP with additional techniques and modifications, such as BSH maintenance and Discrete Wavelet Transformation (DWT) application [94], a nuanced decision-making process is facilitated. BSH is upheld at each step, aiding in tracking the evolution of belief states. DWT converts the history of belief states into a wave, utilizing zero-crossing lines to identify change points, informing the system about the user's knowledge level.

The task involves determining the current dialogue state since belief states comprise probability-state pairings. For example, the system might assess an 85 percent probability that the user is in state X and a 15 percent probability for state Y based on user observations. The belief state, therefore, combines these probabilities with respective states. When receiving new user observations, typically input, the system adjusts the belief state probabilities accordingly. For instance, if the information increases the likelihood of both states, the belief state might be changed to A:80 percent and B:40 percent [35]. This dynamic adaptation of belief states is pivotal for effective dialogue management and user interaction within the system. The research discussed by the author in [87] incorporates a comparable concept, enhancing it with the integration of reinforcement learning (RL) to refine the model. When the system receives a reward for transitioning to the next state, this reward is conveyed to the q-learning model, which employs reinforcement learning techniques to update the system's latest mode. The dialogue management system then considers the updated method to inform the decision-making process for the subsequent set of actions. The adjusted q-learning equation for each way can be expressed as follows:

$$Q(x,a) = (1-\alpha)Q(x,a) + \alpha\left(r + \gamma \max\left(Q'(y,b)\right)\right) \tag{1}$$

In the given equation, Q' represents the updated mode. Notably, various studies exploring POMDP employ different models than those used in this thesis. However, these alternative models serve as valuable resources for comprehending the operational principles and diverse applications of the POMDP method. For instance, the research in [56] introduces a Gaussian process coupled with POMDP dialogue management, offering an automated mechanism for adjusting POMDP policies without manual intervention. While these alternative models differ from the one adopted in this thesis, they provide insightful perspectives on the functioning and potential applications of the POMDP method.

The POMDP dialogue manager utilized is an evolution of the one developed in previous works such as [87], [94], [43], and [46]. Notably, it integrates the Discrete Wavelet Transform to assess the user's knowledge level, a topic elaborated upon in the subsequent section. This amalgamation of existing POMDP dialogue management frameworks with innovative enhancements ensures a robust and practical approach, building upon the foundational research presented in the referenced works.

In conventional POMDP systems, the practice of maintaining a BSH is not standard, as the entire belief state, comprised of transitions and conditional probabilities, is deemed sufficient for transitioning between states without explicit reliance on the BSH [17]. However, insights from Embodied Conversational Agent (ECA) research, as demonstrated in prior work [76], suggest that including a BSH can offer additional advantages beyond the immediate transition functions of the entire belief state. A belief state, denoted as a probabilistic estimate of the system's state, leverages available transitions and observations to construct a probability distribution across all states, as depicted in figure 2.2.1 [43]. The importance of studying historical data is evident in the concept of BSH [98], providing insights into user behaviour and aiding the system in learning more about the user. The system must recognize the

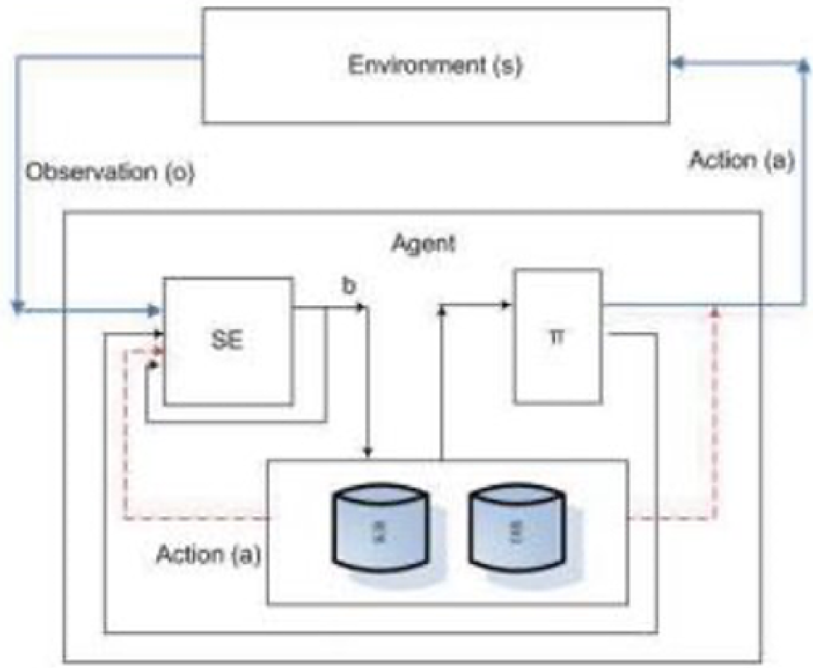

Fig. 2.2.1: Belief State History

probability-state pairings within the belief state to estimate the state. The POMDP interaction between the user and the agent is visually represented in Figure 2.2.1 [98]. In the context of a POMDP, the model is augmented with various observations. Unlike merely observing the current state, the state provides a statement suggesting its probable condition. Consequently, given the probabilistic nature of words, an observation function must be defined. This function articulates the probability of each term for every state within the model, offering a nuanced understanding of the user-agent interaction dynamics.

2.2.3 Wavelet Transform for User Clustering

The Wavelet Transform (WT) is a method employed to analyze diverse signal types. Specifically, the Discrete Wavelet Transform (DWT), a specialized form of the WT, offers an efficient calculation of a signal's compact representation in both time and frequency domains [94]. Despite numerous data analysis techniques, the utilization of wavelet transformation stands out by enabling the extraction of frequency and location information within the dataset [59]. Notably, in the context of discrete datasets, the sampling process of the Wavelet Transform occurs in discrete time. The equation for DWT has the below format:

$$W_q(a,t) = \frac{1}{\sqrt{a}} \sum_{n=1}^{N} q(n)\varphi\left(\frac{n-t}{a}\right) \tag{2}$$

In the given equation, 'n' pertains to the current dataset under analysis, which possesses a size denoted by 'N.' The symbol 'φ' represents the mother wavelet, where 't' signifies discrete time, and 'a' denotes the scale, akin to other equations in wavelet transformations. The DWT is extensively utilized for analyzing non-stationary signals, such as audio. When applied to a wave, the DWT yields distinct variation points within the wave [87].This unique attribute of the DWT proves valuable in detecting the consistency of a student's performance in a tutorial model. Given that these variation points mark locations where the belief state has undergone a change, employing DWT for user clustering is based on the concept that, for an adept user, these variations remain relatively constant. Consequently, users can be classified into different knowledge levels based on the quantity of these change points [31]. DWT and various other wavelet transformation techniques employ "short windows for high frequencies and long windows at low frequencies" during the wavelet transformation process [87]. Wavelet analysis is a methodology that involves the correlation of a given signal, typically a time series, with a set of mathematical functions known as wavelets, which are both scale and time-limited [1]. In a notable study presented in [48], the DWT technique is employed to partition traffic flows into distinct components. This segmentation process changes trends, discrete quantities, and a discrete baseline. The changing trends signify the evolving patterns in the traffic data, while the discrete portions capture abrupt changes in traffic volume. Broadly, trend analysis is conducted on historical and time series data to forecast future subjects of interest [88]. In Figure 2.2.2, the asterisk symbol denotes the "zero-crossing" points, representing shifts in the user's understanding or intentions. The left window, characterized by a larger size, identifies fewer change points, while the right window, with a smaller

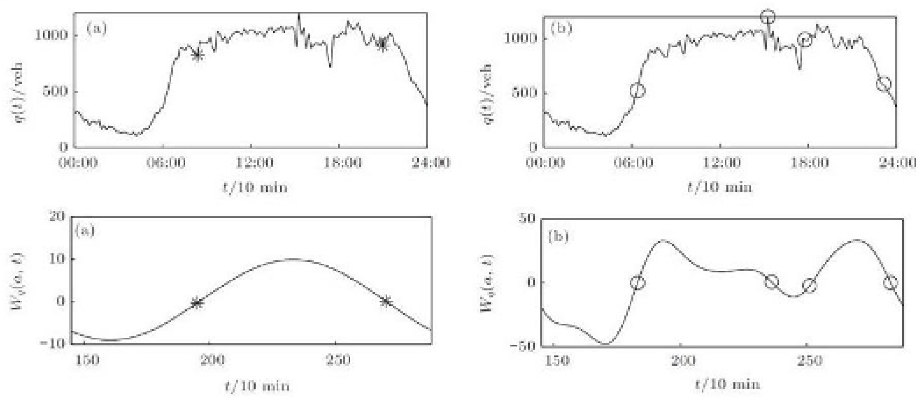

Fig. 2.2.2: Discrete Wavelet Transform (DWT) Example

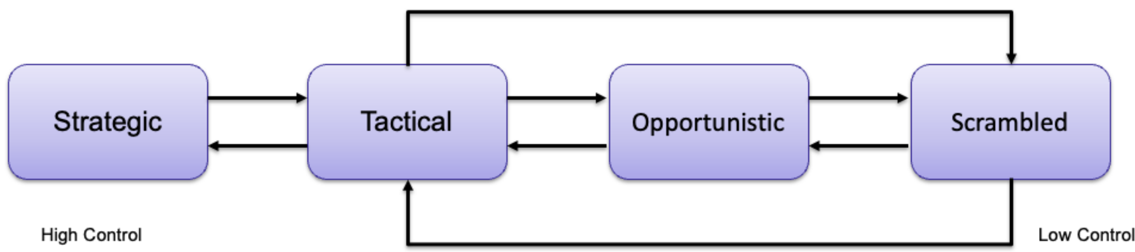

Fig. 2.2.3: Available COCOM mode transition

size, detects a greater number of change points. Figure 2.2.2 above shows how the DWT transformation works on a wave. In the context of this thesis, the methodology outlined in [59], leveraging Discrete Wavelet Transformation, is adopted to categorize system users based on their knowledge levels.

2.2.4 Cognitive Support with COCOM

The COCOM (Contextual Control Model) is crucial to POMDP dialogue management system actions. It operates across four distinct modes, each delineating the system's level of information and its capacity to strategize for optimal information acquisition. The "Strategic Mode" represents a pivotal aspect of dialogue-based tutoring systems, emphasizing establishing effective communication channels between users and the system. In this mode, the system goes beyond immediate responses and engages in long-term conversation planning to ensure sustained user engagement and coherent dialogue. The strategic method aims to enhance the overall conversational experience by considering the broader context of the interaction, maintaining continuity across multiple turns, and strategically planning responses to align with the user's learning objectives. This involves addressing current queries, foreseeing potential future topics, and structuring the dialogue accordingly [87]. By adopting the strategic mode, dialogue-based tutoring systems can foster a more natural, dynamic, and goal-oriented interaction, ultimately contributing to the effectiveness of the learning experience. "Tactical Mode" constitutes a distinctive facet of the tutoring system, characterized by its nuanced approach to information delivery within the ongoing discourse. Unlike strategic modes that involve extensive planning for future interactions, tactical mode prioritizes immediacy and relevance to the current topic under discussion. This mode enables the system to provide supplementary information without diverging into broader subject matter, maintaining a focused and contextually precise engagement with the learner. The tactical mode's strength lies in its ability to cater to the specific needs of the moment, ensuring that the additional information supplied aligns seamlessly with the ongoing conversation. This approach is particularly advantageous in scenarios where a more concentrated exploration of the current subject is beneficial, contributing to a streamlined and efficient learning experience for the user[87]. In the Opportunistic Mode, the tutoring system functions by predicting the subsequent action within the ongoing conversation, primarily relying on the information available at the present moment. This operational mode is characterized by its real-time responsiveness, where the system dynamically evaluates the current state of the dialogue without the capability to plan extensively for future interactions. Unlike modes that involve comprehensive planning or extensive foresight, the Opportunistic Mode prioritizes immediate context, adapting its responses based on the most recent user input and system knowledge. This mode is particularly suited for scenarios where the system needs to respond promptly and contextually, addressing the user's queries or concerns in the flow of the ongoing conversation. While it may lack the capacity for long-term planning, the Opportunistic Mode excels in providing agile and relevant responses within the confines of the current dialogue context, contributing to a more interactive and adaptive learning experience for the user [87]. In the

Scrambled Mode, the tutoring system encounters a scenario characterized by inadequate information, hindering its capacity to make accurate predictions regarding the following action. This mode reflects a state of uncertainty where the system grapples with incomplete or ambiguous data, making it challenging to formulate precise predictions based on the available information. Consequently, actions undertaken in this mode are randomized, introducing an element of unpredictability to the system's responses. Adopting Scrambled Mode is a strategic response to navigate situations where the system faces difficulties discerning the user's intent or context due to information gaps. This mode acknowledges the inherent limitations in the available data and aims to respond, albeit randomized, ensuring continuous engagement and adaptability without clear predictive cues [87]. Incorporating Scrambled Mode adds a layer of flexibility to the tutoring system, allowing it to navigate uncertainties in user input and maintain an interactive and responsive learning environment.

The versatility of COCOM, encapsulated in these four modes as shown in Figure 2.2.2, enables the POMDP dialogue management system to dynamically adapt its approach based on available information, ensuring an optimal and contextually appropriate response within the dialogue.

2.2.5 Multimodal Output Generation

Many text-to-speech engines are currently accessible, enabling the conversion of textual inputs into audio formats. Many text-to-speech machines are commercially available, offering seamless integration into larger applications or software systems. Notable examples include Amazon Polly, Google TTS, and Voicery. In prior research efforts, our team utilized Voicery [87] and Google TTS [46].

The system employs Fuzzy Logic to determine the appropriate emotional expression for the Embodied Conversational Agent (ECA) during interactions. This process involves leveraging the Number of Change Points (NCP), rewards, and established rules. The selection of the specific emotion is guided by applying fuzzy rules, contributing to a nuanced and contextually appropriate ECA response. Sentiment and emotion analysis are pivotal in education, influencing teachers and students. A teacher's effectiveness extends beyond academic qualifications to factors such as enthusiasm, talent, and dedication [62]. The primary aim of sentiment analysis is to discern the underlying sentiments expressed by users, categorizing them as positive, negative, or neutral and potentially identifying specific emotions like happiness, sadness, anger, or surprise. Previous research employs sentiment analysis to evaluate user sentiments [76], yielding three floating-point values between 0 and 1, representing negative, neutral, and positive opinions. This analysis enables the ECA to comprehend user emotions, fostering more empathetic and engaging interactions. By adapting its behaviour, tone, or dialogue strategies based on detected sentiments, the ECA enhances the overall user experience. Sentiment analysis proves valuable in deciphering opinions, as seen in examples like "I hate this" (negative sentiment), "I don't know what that means" (neutral view), and "This is a good experience" (positive emotion) [87].

In addressing real-world problem-solving scenarios, the need arises to navigate a multitude of ambiguous variables, each characterized by various forms of uncertainty such as randomness, fuzziness, indistinguishability, and incompleteness. Different Artificial Intelligence (AI) methodologies, such as fuzzy logic, have been introduced to tackle this inherent ambiguity [79]. Fuzzy logic, a multi-valued logic, assigns logical values to variables as real numbers between 0 and 1. This logic framework enables the implementation of partial correctness, allowing the degree of correctness to range between entirely true and entirely false [83]. In a specific application, the authors of the work presented in [11] leverage a fuzzy logic system (FLS), particularly the Mamdani model, a widely adopted undefined inference method. Their work focuses on identifying jamming attacks in Wireless Body Area Networks (WBANs) by considering three network parameters: Packet Delivery Ratio (PDR), Received Signal Strength Indicator (RSSI), and Energy Consumption Amount (ECA). Similarly, Ruturaj incorporates fuzzy logic to determine the emotion an Embodied Conversational Agent (ECA) should express, utilizing criteria considering recent sentiment analysis and COCOM [87]. A module employing PocketSphinx, a speech recognition engine, is constructed to ensure accurate phonetic alignment. This module takes the text designated for speech output and the corresponding voice file provided by the chosen text-to-speech engine. This integration enhances the overall effectiveness of the speech synthesis process and ensures a coherent and natural delivery of the intended content.

2.3 Ontology in Dialog Management

In computer science and artificial intelligence, ontology functions as a formal representation of knowledge, delineating concepts within a specific domain and elucidating the relationships between these concepts. This structured framework is a pivotal tool for reasoning about entities within the field, enabling a comprehensive and systematic description of the subject matter. The concept of replacing human interlocutors with machines has been explored since the 1960s, coinciding with the design of numerous computer systems aiming to facilitate communication with machines. Regardless of the situation or context, engaging in dialogue with a machine has remained an intriguing area of interest. Initially defined as a "conversation between two or more people on a defined subject", the term "dialogue system" introduces the idea that at least one participant in this conversation is a machine in the broadest sense. A more specific definition proposed in [29] characterizes a dialogue system as enabling interaction between a human and a system within a confined framework. Despite the diverse forms of exchange, such as textual chat, voice servers, or virtual agents, establishing a comprehensive overview of the state-of-the-art remains challenging [91].

The versatility of dialogue systems is reflected in their various forms and characteristics, leading to a need for a unified architectural standard in the literature. Numerous proposals exist, showcasing diverse approaches, including open-domain models like those outlined in [91]. The latter emphasizes leveraging general knowledge sources, such as the semantic web, to enable the system to cover a broad thematic field. The proliferation of different forms and characteristics in dialogue systems underscores the need for further exploration and research to establish more standardized and comprehensive approaches.

Ontologies, fundamental in knowledge representation, are commonly articulated using ontology languages, with the Web Ontology Language (OWL) being a widely adopted choice. OWL encompasses different sub-languages, each offering varying levels of expressiveness tailored to diverse modelling needs. First, OWL Lite is designed for lightweight ontology representation, catering to applications with more straightforward modelling requirements. Despite its limited expressive power compared to other sublanguages, OWL Lite is lauded for its simplicity, enhancing ease of use and computational efficiency. Moving up in expressiveness, OWL DL (Description Logic) is crafted for applications requiring more intricate modelling and advanced reasoning capabilities. OWL DL facilitates efficient automated reasoning by operating on a decidable subset of first-order predicate logic. OWL Full is on the end of expressiveness, offering maximum flexibility in ontology modelling [69]. However, this flexibility comes with the trade-off of decidability challenges and increased computational complexity. While OWL Full allows for rich modelling, it is generally cautioned against in practical applications due to the potential creation of undecidable ontologies. This hierarchy within OWL sublanguages allows ontology developers to choose the appropriate level of expressiveness based on the complexity and requirements of their specific applications.

As a structured representation of knowledge, ontology encompasses several fundamental components essential for comprehensively defining and organizing information within a specific domain. Classes, as foundational building blocks, represent categories or entities inherent to the environment. Instances, however, embody specific occurrences or examples within these classes. Attributes, or Data Properties or Slots, contribute by describing features associated with classes, aiding in their precise definition [66]. Relations, or Object Properties, define how classes and instances are interconnected within the ontology, establishing the relationships between them. Function Terms capture complex structures constructed from other ontology components, providing a mechanism for representing intricate relationships. Axioms, explicit statements within the ontology, serve as foundational truths or assertions, establishing the interpretation of terms. Events become crucial in dynamic ontologies, representing occurrences and changes over time. Restrictions impose constraints on properties, ensuring adherence to specific rules and conditions. Controls enable inferencing, allowing for the deduction of new knowledge based on existing ontology components. Annotations provide additional metadata or descriptive notes about ontology components, offering insights, comments, and context to enhance understanding. Together, these components form the cohesive ontology framework, facilitating the structured representation and organization of knowledge within a defined domain.

The judicious combination of these components facilitates the creation of specific ontologies tailored to diverse purposes, ranging from general knowledge representation to highly specialized domain representations. Ontologies contribute significantly to knowledge sharing, interoperability, and reasoning across various applications and systems, forming a cornerstone in advancing intelligent systems.

2.3.1 Ontology and Its Construction

Ontology building is a fundamental process in knowledge representation, involving the formal creation of an ontology—a structured representation of knowledge specific to a particular domain or subject area. This intricate process encompasses defining and organizing the domain's concepts, entities, relationships, and properties to capture its structure and semantics [19]. The steps involved in ontology building are systematic and iterative, ensuring the accuracy and relevance of the resulting knowledge model.

The initial step in ontology construction is Domain Analysis, where a comprehensive understanding of the domain is developed. This involves identifying key concepts, entities, and relationships within the specified subject area [19]. Following this, the Conceptualization phase takes shape, wherein the defined concepts and their relationships are formalized based on the outcomes of the domain analysis. This step includes determining the hierarchical structure of concepts and specifying their properties. Each key concept identified during domain analysis is precisely defined in the Conceptualization phase. These definitions go beyond natural language descriptions and aim for a formal specification that captures the essence of the concept. This involves clarifying the boundaries of each concept and determining the essential characteristics that distinguish one concept from another. Building on the identified relationships between concepts, the Conceptualization phase involves formalizing these relationships. This includes specifying the nature of connections between different concepts, such as whether they represent part-whole relationships, membership associations, or any other relevant link. The goal is to capture the semantics of these relationships within the ontology accurately [19]. One key aspect of Conceptualization is establishing the hierarchical structure of concepts. This involves organizing concepts into a taxonomy or hierarchy based on their broader and narrower relationships. The hierarchical arrangement helps understand the subsumption relationships between concepts, where one concept is considered more general (parent) than another (child). Concepts in an ontology often possess properties that define additional attributes or characteristics. In the Conceptualization phase, these properties are specified, describing the essential features associated with each concept. This may include data types, constraints, and other relevant information that enriches the understanding of the concept. As the concepts and their relationships are refined and formalized, the Conceptualization phase ensures alignment with the chosen formal language for ontology representation. Whether using OWL, RDF, or another notation, adhering to the syntax and semantics of the selected language is crucial for ensuring interoperability and compatibility with other systems [19].

Formalization is the subsequent stage, wherein the concepts, relationships, and properties are expressed in a formal language or notation, such as OWL or RDF. This crucial step involves detailing the semantics and constraints associated with the ontology elements. Knowledge Acquisition is then undertaken, gathering pertinent information and knowledge from domain experts, existing resources, documents, or databases [33]. This collected information enriches the ontology with accurate and comprehensive data. The process begins with clearly defining the concepts that constitute the domain. These concepts are the building blocks of the ontology and represent the entities or ideas relevant to the subject area. Once concepts are defined, the next step is to identify the relationships between these concepts. Relationships establish connections and dependencies, illustrating how different entities within the domain interact or relate. Conceptualization involves organizing concepts hierarchically to represent their taxonomic relationships [33]. This results in a structured hierarchy, often resembling a tree-like structure, where more specific concepts are linked to broader, more general ones. Each concept is associated with properties that define its characteristics or attributes. Properties describe the various aspects of a concept and help establish distinctions between different instances of the same concept. During Conceptualization, it's crucial to clarify the semantics of each concept and relationship. This ensures a shared understanding of the intended meaning, preventing ambiguity and facilitating consistent interpretation. Constraints on concepts and relationships are identified to capture the inherent rules and restrictions within the domain. These constraints are crucial in guiding the behaviour and interactions of entities in the ontology. Sometimes, it is beneficial to model specific scenarios or use points within the Conceptualization phase. This involves illustrating how concepts and relationships come into play in real-world situations, providing a practical context for the ontology. Conceptualization is an iterative process that may include feedback loops with domain experts. Continuous refinement ensures that the evolving conceptual model accurately represents the nuances of the domain and aligns with the intended objectives of the ontology.

Validation and Evaluation represent pivotal phases, assessing the quality and correctness of the ontology. Consistency, coherence, and alignment with the intended domain are scrutinized, often involving expert reviews and tests or evaluations. Finally, Maintenance and Evolution recognize the dynamic nature of ontologies. Over time, as the domain or knowledge changes, it becomes imperative to update and maintain the ontology to reflect new information, emerging trends, or evolving requirements. The Validation and Evaluation phase in ontology construction is critical in ensuring the developed ontology's robustness, accuracy, and relevance [2]. This phase is characterized by a meticulous examination of several key aspects, each contributing to the overall quality assurance of the ontology. One of the primary focuses during validation is to assess the consistency of the ontology. This ensures no logical contradictions or conflicts within the defined concepts, relationships, or axioms. Any inconsistencies could undermine the reliability of the ontology and its effectiveness in accurately representing the intended domain. The coherence of the ontology is another paramount consideration [74]. This aspect evaluates the interconnectedness and logical flow of the ontology elements. A coherent ontology ensures that the defined concepts and relationships align logically, fostering a seamless representation of knowledge within the specified domain. Incoherence may lead to ambiguity and hinder the ontology's interpretability. The validation phase rigorously checks whether the ontology aligns with the intended domain. Domain experts play a crucial role in this process, providing their insights to verify that the ontology accurately captures the essential concepts and relationships within the targeted subject area [2]. This collaborative review helps bridge the gap between formal ontology representation and real-world domain knowledge. Expert input is invaluable during the validation process. Domain experts, possessing in-depth knowledge of the subject area, conduct thorough reviews to identify potential inaccuracies, omissions, or conceptual misalignments. Their feedback contributes to refining the ontology and ensuring its fidelity to the nuances of the domain. Various tests and evaluations are employed to assess different facets of the ontology [20]. This may include assessing the ontology's performance against predefined criteria, evaluating its efficiency in knowledge retrieval, and validating its adherence to specified constraints. Rigorous testing methodologies help identify and rectify any shortcomings in the ontology construction process.

The ultimate goal of the Validation and Evaluation phase is to establish a high level of confidence in the ontology's quality and fitness for its intended purpose [74]. By addressing issues of consistency, coherence, and alignment and incorporating feedback from domain experts and systematic evaluations, this phase contributes significantly to the overall success of ontology construction. It ensures that the ontology is a reliable and accurate representation of knowledge within the designated domain, laying the foundation for effective utilization in various applications and scenarios.

Furthermore, the ontology construction process unfolds as an iterative and methodical journey encompassing six distinct tasks. The initial task involves Specifying the Domain, where the focus is on creating well-defined terms and concepts that encapsulate the intricacies of the subject area. Following this, the Identification of Key Terms and Concepts ensues, emphasizing the importance of pinpointing the pivotal elements that form the backbone of the ontology.

Establishing Rules and Axioms is a critical step in the ontology construction process, as it involves defining the structural properties of the domain. This task contributes to the precision and integrity of the ontology, ensuring that the relationships and constraints within the knowledge model are accurately articulated. Subsequently, the process moves to Encoding the Ontologies, where the formal representation languages, such as RDF, RDFS, or OWL, come into play [56]. This step involves translating the conceptual framework into a machine-readable format, providing a standardized and structured representation of the ontology elements.

A noteworthy aspect of ontology construction is combining it with existing ontologies. This acknowledges the collaborative and interconnected nature of ontological knowledge. Integrating established ontologies enhances the comprehensiveness and interoperability of the resulting knowledge model, fostering a more holistic understanding of the domain. Importantly, these tasks are not rigidly sequential but facilitate an Iterative and Interactive execution. This flexibility allows for refinement and adjustment as insights emerge during construction. Ontology construction methods can vary, offering Manual Construction, where experts actively contribute to the creation process, Cooperative Construction, involving expert supervision throughout various tasks, and Automatic Construction, where the process is somewhat automated, with minimal user or specialist intervention [24].

Despite technological advancements, achieving Fully Automatic Construction of ontologies remains a formidable challenge [67]. The nuanced nature of ontology development, often requiring domain expertise and nuanced decision-making, poses hurdles to complete automation. As such, the human intellect and expertise continue to play a crucial role in ensuring the accuracy, relevance, and coherence of ontologies crafted for diverse applications and domains.

2.3.2 Challenges in Ontology Construction

The endeavour to express human knowledge within ontology construction confronts a noteworthy challenge, particularly in applying formal logic, which may inadvertently introduce gaps in machine interpretation [68]. This challenge becomes particularly pronounced in domains characterized by intricate and interlinked concepts, where the design of ontologies necessitates meticulous consideration, as articulated by [63]. The complexity inherent in such domains requires a thoughtful balance between formalization and the nuanced nature of human knowledge.

Expanding domains bring Maintenance Challenges, a critical facet of ontology development highlighted by [89]. Maintaining the ontology's relevance and accuracy becomes formidable as data and relationships within a domain proliferate. The sheer volume of evolving information necessitates a dynamic approach to ontology upkeep to ensure that the representation reflects the ever-changing domain landscape.

Furthermore, the accuracy of relationships encoded within ontologies hinges on comprehensive and precise data availability. Only complete or adequate data can introduce uncertainty, potentially leading to inaccurate conclusions about the domain under consideration, as emphasized by [68]. This underscores the pivotal role of data quality in shaping the efficacy and reliability of ontologies. Efforts to address these challenges require a holistic approach, considering both the intricacies of human knowledge representation and the evolving nature of the domains they seek to capture. As ontologies bridge human understanding and machine processing, striking this balance is essential for their successful application in various knowledge-intensive domains.

Meta Ontology Alignment presents another challenge, requiring harmonizing meta-ontology concepts with specific domain requirements [63]. The complex task of Relationship Mapping from meta ontology to domain-specific ones is hindered by semantic variations [68]. Eliminating Redundancy becomes essential in ontology construction, necessitating the removal of redundant concepts and relationships from the meta-ontology while maintaining domain relevance [89].

Upon closer examination of ontology construction studies, a consensus emerges on several challenges that necessitate further attention. Firstly, while achieving Fully Automatic Construction for ontologies may be challenging, there is a pressing need to decrease human intervention in the process, advocating for semi-automatic construction over existing cooperative systems [15], [30], [13], [56], [96], [104]. Secondly, mitigating noise terms, which are irrelevant or overly general, is crucial early in the construction process to minimize unnecessary additional efforts [96], [104].

Discovering relations between concepts remains an unsatisfactory aspect of ontology construction, requiring more dedicated efforts [7], [15], [13], [96], [104]. Learning axiom is still in its initial stages in existing ontology construction systems and demands further refinement [13], [56], [96], [104]. The transformation of data, whether from small static text collections or massive heterogeneous sources on the World Wide Web, should be a focal point in designing ontology construction systems [15], [56], [96], [104].

Building a standard platform to evaluate ontology construction systems remains a challenging task [7], [15], [96], [104]. These challenges underscore the intricacies of ontology construction, requiring continuous exploration and refinement to enhance the effectiveness and reliability of automated ontology development processes.

2.3.3 Automated Ontology

In the context of the burgeoning volume of educational data, the imperatives for consistency, rapid adaptation to evolving domains, and the delivery of personalized and sophisticated educational experiences have become paramount [24]. Automation is a crucial ally in addressing these challenges, particularly in coping with the vast volumes of data generated within educational platforms. As academic domains naturally evolve and new information continually emerges, keeping ontologies up-to-date and maintaining their accuracy necessitates continuous monitoring and adjustment [75]. This challenge is particularly pronounced in intricate domains where the interconnection of concepts can be overwhelming. Automation efficiently manages this complexity, ensuring that educational ontologies reflect the dynamic nature of knowledge domains. The imperative of semantic interoperability, requiring mapping varied concepts to a unified ontology [105], underscores the vital role of automation in seamlessly incorporating new data in fast-changing fields. Automated ontologies become instrumental in achieving semantic coherence and standardized representation in educational content.

Given these challenges, leveraging automation becomes imperative for educational platforms seeking efficient, accurate, and adaptive ontology creation. This, in turn, leads to improved personalized tutoring and enriched learning outcomes. Addressing these challenges effectively requires a holistic approach, combining domain expertise, sophisticated data processing techniques, ontology engineering skills, and integrating artificial intelligence (AI) and machine learning (ML) approaches. Automated methods, empowered by advanced technologies, streamline the ontology construction process and mitigate potential pitfalls associated with manual systems.

Automation emerges as a critical facilitator in coping with these extensive datasets. The sheer volume of data generated within educational platforms, including student interactions, assessment results, and content consumption patterns, necessitates robust mechanisms for handling, processing, and extracting meaningful insights from this information. This involves the implementation of algorithms and computational processes that can systematically organize, analyze, and interpret large datasets. This streamlines the data processing pipeline and ensures that relevant information is extracted promptly and efficiently [24]. For instance, automated data management systems can handle tasks such as aggregating student performance metrics, tracking learning progress, and identifying patterns in user engagement. Furthermore, the efficient coping with data volume goes beyond mere storage and retrieval. Automation facilitates data-driven decision-making by providing real-time analytics and actionable insights. This is particularly valuable in educational settings, where prompt responses to student needs and timely adjustments to instructional strategies can significantly impact the learning experience. From a technical perspective, automation in data management involves the implementation of data processing pipelines, integrating databases, and deploying machine learning algorithms for predictive analytics [96].

These automated systems can handle tasks such as data cleaning, normalization, and generating reports, freeing human resources from mundane and time-consuming data-related activities. Thus, managing data volume through automation underscores the pivotal role of automated data processing in educational platforms. It highlights how automation ensures that the vast amounts of data generated within these platforms are effectively handled and leveraged to enhance the overall educational experience. The efficiency gains achieved through automated data management contribute to educational systems' agility, adaptability, and responsiveness in the face of evolving user needs and dynamic learning environments.

Ensuring consistency within educational content is a critical facet of educational ontology construction, and automation is pivotal in achieving this objective. The dynamic nature of educational landscapes, characterized by evolving knowledge domains, curriculum updates, and emerging pedagogical trends, poses a substantial challenge to maintaining coherence and uniformity across educational materials [104]. Automation addresses this challenge by providing a systematic and standardized approach to constructing and updating educational ontologies. Through predefined rules, algorithms, and ontological structures, automated processes ensure that the representation of concepts, relationships, and attributes within the academic domain remains consistent over time. This consistency is essential for creating a cohesive and reliable knowledge framework that learners, educators, and educational systems can rely on.

Firstly, through terminology standardization, automation is pivotal in maintaining a unified and unambiguous language within the ontology. Adhering to standardized language conventions minimizes ambiguity, fostering a common understanding of terms across diverse educational contexts. Conceptual alignment is another crucial aspect facilitated by automated processes. These processes systematically align concepts within the ontology to predefined conceptual frameworks and taxonomies. This alignment creates a clear and logical structure, empowering users to navigate and comprehend educational content easily. Semantic harmony is crucial for avoiding conflicting interpretations of concepts, and automation plays a vital role in enforcing semantic coherence [67]. Automation ensures a consistent and accurate representation of knowledge by establishing clear definitions, specifying relationships, and eliminating semantic ambiguities that may arise from manual construction processes. Automation facilitates seamless adaptation in the dynamic landscape of education, where curriculum changes are frequent to align with evolving educational objectives and learning outcomes. Automated processes efficiently update the ontology to reflect modifications in educational content, ensuring alignment with the latest curriculum standards. The cross-platform compatibility of educational materials is paramount, given the diverse platforms and technologies through which content is accessed. Automation guarantees that the ontology remains compatible across different educational systems, media, and delivery modes. This fosters consistency in the educational experience, irrespective of the technology employed. Quality assurance is integral to the reliability of educational content, and automation incorporates robust validation mechanisms. These mechanisms encompass error checking, verification of logical consistency, and adherence to ontology design principles. Through these quality assurance measures, automation contributes significantly to the overall reliability and accuracy of the educational ontology.

The ability to swiftly adapt to domain changes and the continuous emergence of new information is critical to effective educational content management. Automated processes play a pivotal role in this realm, offering a remarkable capacity for agility in responding to the evolving educational landscape [67].

Dynamic Domain Response: Educational domains are subject to constant evolution, reflecting advancements in knowledge, pedagogy, and technological landscapes. Automated processes excel in their dynamic response to these changes, ensuring educational content remains aligned with the latest developments in the respective domains. Whether it involves updates in curriculum standards, introducing new subjects, or shifts in educational methodologies, automated systems can promptly integrate these changes into the content structure.

Real-time Information Integration: The emergence of new information is a constant in the educational domain. Whether it be breakthroughs in research, updated statistical data, or novel insights into pedagogical approaches, automated processes facilitate real-time integration of this information into educational content. This real-time adaptability ensures learners can access the most current and relevant insights, fostering a dynamic and engaging learning experience.

Agile Curriculum Development: Changes in educational objectives, learning outcomes, or the overall curriculum necessitate a rapid response to update educational materials. Automated processes streamline curriculum development by providing agile content creation, modification, and deployment mechanisms. This agility is particularly valuable when curriculum changes are frequent or when a flexible and responsive educational approach is essential.

Customized Learning Paths: Educational content adaptation goes beyond just updating information, it extends to tailoring learning experiences for individual students. Automated systems, leveraging data analytics and machine learning, can dynamically adjust learning paths based on personal progress, preferences, and learning styles. This adaptability ensures that each learner receives a personalized and effective educational journey.

Feedback-driven Iteration: Automated systems are adept at collecting and analyzing feedback from various sources, including students, educators, and educational analytics. This feedback loop enables iterative improvements in educational content, responding to the evolving needs and preferences of the learning community. This continuous improvement cycle ensures that the educational content remains relevant, effective, and aligned with the overarching educational goals.

The aspect of handling complexity is a critical dimension addressed by automation. Educational domains are inherently multifaceted, characterized by intricate interconnections among various concepts, entities, and relationships. The complexity arises from the diverse nature of educational content, which spans multiple subjects, topics, and knowledge interdependencies. Automation plays a pivotal role in efficiently navigating and managing this complexity. As the educational landscape expands, the sheer volume of interconnected concepts can become overwhelming for manual construction processes. With advanced algorithms and data processing capabilities, automation excels in untangling and organizing these intricate webs of relationships [104]. One key challenge in handling complexity is ensuring that the relationships and dependencies among educational concepts are accurately captured and represented within the ontology. Manual construction may need help to maintain the coherence of these interconnections, leading to potential oversights or inconsistencies. On the other hand, automation leverages systematic algorithms to analyze and model intricate relationships, enhancing the overall coherence of the educational ontology. Moreover, educational content often evolves and diversifies over time, introducing dynamic elements that contribute to the complexity of the ontology. Automation exhibits adaptability in managing these growing complexities by incorporating updates seamlessly. It ensures that the ontology remains reflective of the dynamic nature of educational knowledge, accommodating changes in curriculum, emerging topics, or advancements in pedagogical approaches.

Additionally, handling complexity extends to the representation of hierarchical structures and dependencies among educational concepts. Automation aids in constructing ontologies with well-defined taxonomies, ensuring that the relationships between broader categories and more specific ideas are accurately captured. This enhances the organization of educational content and facilitates efficient retrieval and navigation within the ontology.

Semantic interoperability in educational content refers to the seamless exchange and integration of diverse academic concepts through a unified ontology, and automation plays a pivotal role in facilitating this intricate process [105]. The education domain is inherently diverse, with many concepts, terms, and relationships across various disciplines and subjects. Semantic interoperability is crucial for establishing a shared understanding and representation of these diverse concepts, fostering a cohesive and standardized educational landscape. Automation streamlines the complex task of mapping and aligning various academic concepts into a unified ontology. This involves systematically encoding educational entities' relationships, hierarchies, and properties, ensuring a structured and standardized representation [74]. The challenges associated with manual mapping, which can be error-prone and time-consuming, are mitigated through automated processes. Automation leverages advanced technologies, such as natural language processing and machine learning, to analyze and interpret educational content, identifying patterns and connections that contribute to creating a unified ontology.

A cohesive and standardized representation of educational content is paramount for enhancing communication and knowledge sharing across diverse educational systems, platforms, and stakeholders [74]. Semantic interoperability ensures academic concepts are accurately represented and interpreted consistently across different contexts. This consistency is vital for promoting collaboration, developing interoperable educational tools and resources, and fostering a more coherent educational ecosystem. Besides, semantic interoperability facilitated by automation contributes to the adaptability of educational systems. As educational content evolves and new concepts emerge, automated processes can efficiently update and align the unified ontology, ensuring that it remains reflective of the dynamic nature of knowledge domains. This adaptability enhances the longevity and relevance of educational content, supporting educators, learners, and educational technology developers in navigating the ever-changing landscape of knowledge.

The facet of personalization within automated ontologies represents a pivotal dimension in the realm of educational technology. By harnessing the power of automation, educational platforms can tailor learning experiences to meet individual users' unique preferences and needs. This process involves the dynamic adaptation of educational content based on various factors such as learning styles, proficiency levels, and individual preferences, ensuring a more customized and engaging learning journey for each user. In the context of personalization, it can swiftly analyze user interactions, learning patterns, and historical data. By discerning individual preferences and understanding how users engage with educational content, these automated systems can intelligently curate and present material that resonates most effectively with each learner.

The adaptability of automated ontologies enables real-time adjustments to the delivery of educational content, allowing for seamless modifications in response to user interactions and evolving learning needs [47]. For instance, if a user prefers visual learning modalities, the system can dynamically adjust to present information using visual aids or multimedia content. Similarly, if a user exhibits proficiency in specific topics, the system can intelligently progress the learning path to more advanced material. Moreover, the personalization aspect extends beyond content delivery to encompass the style and format of interactions. Automated ontologies can customize the mode of instruction, ranging from the choice of language and tone to the level of detail in explanations, aligning with individual preferences and optimizing user engagement [47]. This adaptability and personalization foster a learning environment where users feel more connected to the educational content, leading to increased motivation, better retention, and higher overall satisfaction. The automated personalization of educational experiences, facilitated by ontologies, aligns with the contemporary trend of moving away from one-size-fits-all educational approaches toward more learner-centric models.

In conclusion, integrating automation in educational platforms is pivotal for overcoming the challenges of the evolving educational landscape. Automation copes with data volumes and ensures consistency, facilitates adaptation to change, manages complexity, achieves semantic interoperability, and enables personalized educational experiences. This comprehensive approach, empowered by advanced technologies, aligns with the overarching goal of enhancing educational outcomes through efficient and adaptive ontology construction.

2.4 Knowledge Extraction from Datasets

Extracting knowledge from textual data has become a crucial facet of semantic technology, gaining prominence in the Semantic Web (SW) [67]. While the interest in ontology learning is not a recent development, with earlier works dating back to 2001 [51], it gained advanced tools like Text2Onto in 2005 [20]. Initially, the Semantic Web community showed limited interest in Knowledge Extraction (KE), preferring manual ontology design as a mark of quality. However, this perspective shifted with the advent of linked data bootstrapping, notably facilitated by DBpedia [12]. This led to a growing need for a substantial knowledge base population, schema induction from data, and natural language access to structured data. The emergence of applications integrating structured and unstructured content further fueled interest in KE.

The intersection of Natural Language Processing (NLP) research with SW resources became evident as SW knowledge was utilized as background knowledge. Notably, graph-based methods have gained traction in semantic technologies, catering to diverse applications. This shift is exemplified by the increasing reliance on SW languages and tools for tasks such as learning basic semantic data structures, including tagged named entities, factual relations, and topics for document classification. The integration of these tools with SW languages has witnessed rapid growth, giving rise to a set of devices considered in this study. These tools are pivotal in providing scalable, application-ready, and precise learning of essential semantic data structures, contributing significantly to the evolving landscape of knowledge extraction and semantic technology.

2.4.1 Knowledge Extraction

In NLP, tasks are traditionally categorized into essential functions, such as named entity recognition and applied procedures, exemplified by question answering. When transitioning NLP algorithms for application in the Semantic Web (SW), a distinction can be made between basic tasks, like class induction, and applied studies, such as Natural Language querying of linked data. This landscape analysis aims to align NLP essential functions with their SW counterparts and compares various tools concerning functionalities that fulfill these tasks. Notably, the semantics provided by NLP resources differ significantly from the assumptions made for ontologies in knowledge representation, especially within the context of the Semantic Web. Additionally, aside from formal deep parsing methodologies like those based on Discourse Representation Theory (DRT) [44], or Markov Logic [22], the formal semantics of NLP data tends to be relatively shallow. This shallowness is evident in its limitation to intensional relations between (multi-)words, senses, or synsets, as well as informal identity relations in entity resolution techniques, sense tagging using typically small sets of tags (e.g., WordNet's "super senses"), and lightweight concept taxonomies. As outlined in this analysis, the distinct characteristics of NLP semantics underscore the challenges and nuances of integrating NLP techniques into the Semantic Web paradigm, calling for careful consideration and adaptation in the pursuit of seamless interoperability and knowledge integration.

The effective utilization and enhancement of ontologies are contingent on the capability to repurpose NLP outcomes through appropriate conversion, a facet often categorized as "semantic technology" in various academic and industrial applications. Linguistic knowledge draws upon legal background knowledge and can reciprocally enable access to and enrichment of traditional knowledge. The amalgamation of formal and formalized linguistic knowledge further extends through automated inferences.

Despite the recent surge in integrating NLP techniques with the Semantic Web (SW) and vice versa, a substantial disparity persists between lexical and NLP data structures and the formal semantics predominantly adopted for ontologies in the Semantic Web. While existing initiatives such as LMF [26], SKOS-XL [82], LIR [69], Lemon [53], FISE, NIF [38], and implementations like Apache Stanbol and NERD aim to bridge the gap by facilitating the "porting" or "lifting" of NLP results or lexical resources to the SW, the formal reuse of NLP results in the SW remains largely contingent on specific applications or user choices. Consequently, exploring the current tool-level landscape is pertinent to discerning emerging best practices and shedding light on practical considerations, especially in scenarios where direct bridging between the two realms is absent.

2.4.2 Tutoring Datasets

Addressing the critical societal need for effective learning, especially in the context of diverse student populations, underscores the importance of adopting educational strategies that yield substantial learning gains. One-on-one tutoring has emerged as a particularly potent technique, supported by numerous studies attesting to its efficacy as an educational aid [92], [64], [31], [77]. The effectiveness of conversational tutors in facilitating learning has been a focal point, emphasizing the need for personalized and interactive educational experiences.

Specific indispensable properties should characterize the interactions in tutoring dialogues, differentiating them from existing open datasets. Firstly, the conversation must be grounded in common concepts recognized by the student and the tutor as the core topics for learning [31]. This grounding ensures a shared understanding of the educational objectives. Furthermore, the conversation needs to be extended, allowing sufficient duration for students to be exposed to new concepts, creating opportunities for retention and recall in subsequent interactions. Variability in responses is another crucial aspect, acknowledging that there can be multiple valid ways for a tutor to respond to a student at any given point in the conversation. This diversity enriches the learning experience and accommodates different learning styles. Lastly, to facilitate open access and ethical considerations, the dialogue collection should exclude personally identifiable information, ensuring that it can be shared as open-access data, and one such dataset is the CIMA [85] dataset that we have practised in this research to mod up the dialogue delivery.

2.4.3 Ontology Construction with Word2Vec

The word embedding stage is crucial in transforming words into numerical vectors, aiming to capture semantic relationships based on contextual similarity. This process is integral in creating an ontology for a collection of texts, as it enables grouping words with comparable contexts into the same class. Traditionally, word vectors had limitations in effectively representing words until the introduction of Word2Vec by Thomas Mikolov under Google's auspices [55]. Word2Vec has emerged as the predominant technique in term embedding, surpassing previous methods in capturing nuanced semantic associations. This approach enhances the representation of words in vector space, allowing for more accurate modelling of semantic relationships within a given text corpus. The utilization of Word2Vec in the word embedding stage signifies a significant advancement, providing a robust foundation for ontology construction and semantic understanding within the realm of natural language processing. Word2Vec, a neural network-based approach, stands out as a powerful method providing probabilities for words in multidimensional spaces and has demonstrated excellence in tasks such as word analogies and similarity assessment. Comprising two models, Continuous Bag of Words (CBOW) and Skip-Gram (SG), Word2Vec employs a neural network to learn weights that serve as word representations. Notably, the Skip-Gram model is an effective technique for acquiring high-quality vector representations of words from extensive unstructured text data. In the Skip-Gram approach, the model predicts context based on input words, offering a versatile framework for understanding the relationships between terms. This contrasts CBOW, where the prediction is reversed, focusing on predicting the input word based on the provided context. The Skip-Gram model, integral to this study, operates by predicting the surrounding words within the associated context, utilizing word vector representations. This methodology, grounded in predictive modelling, enhances the understanding of word semantics and relationships, contributing to the effective model of knowledge in ontology construction.

Our research employs the Skip-Gram model from Word2Vec as it offers superior performance in capturing the contextual information of text data compared to CBOW. The Word2Vec model is developed with specific parameter configurations for optimal results. A crucial parameter, 'window = 5' is set to consider the five words surrounding the word under analysis, enabling the determination of the vector. Additionally, 'min_count = 10' is implemented to filter out tokens with frequencies below 10. This step is essential for eliminating tokens dominated by typos and rare words, preventing computational burdens associated with calculating typos. The parameter 'sg = 1' is selected to indicate the use of skip-gram, allowing the model to derive context based on the given word. The 'iterate = 10' parameter determines the number of epochs during model training, ensuring the production of a well-tailored model. Following the construction of the Word2Vec model, a dictionary is created with vocabularies as keys and corresponding vectors as values. This dictionary facilitates the conversion of vectors to language, streamlining the process before ontology construction. The ontology is a structured representation of concepts or classes and their taxonomy through the “subclass-of” relationship. During the design phase of the meta Ontology, approximately 24 types were utilized in alignment with our Meta Ontology model. Subsequently, each list generated from the clustering stage undergoes a thorough analysis, and the predominant tags within each list determine the value stored in a corresponding ontology class. The ontology construction process leverages Protégé and the OWL ontology language, facilitated by a Python library that supports both Protégé and OWL. Specifically, the Owlready2 library in Python is employed for ontology construction, providing a cost-effective and accessible solution for this study. This choice aligns with the project’s requirements and objectives, seamlessly integrating ontology construction processes within the Python programming environment.

CHAPTER 3

Prior Work

This chapter of research predominantly relies on prior investigations conducted by our research group at the University of Windsor, as well as studies conducted by other research institutions. It delves into earlier research endeavours to assess their contributions and explore their sources of inspiration.

3.1 Use of Ontology in E-Learning

E-learning recommender systems often leverage ontology and knowledge-based systems as their fundamental infrastructure. Despite their prevalence, a comprehensive exploration of ontology use within these systems needs more systematic investigation. This research aims to fill this gap by examining the development, evaluation, and technical aspects of ontology-based recommender systems in e-learning. Within these systems, ontology plays a pivotal role in multiple dimensions. Primarily, it models elements such as student profiles, learning objects, feedback mechanisms, assessments, and contextual data. However, the prevalent use largely centers around student models and teaching things, with potential future considerations for incorporating learning paths, feedback mechanisms, and learning device recommendations.

The recommendation process within ontology-based systems exhibits a reciprocal nature, allowing initiation either by the system or students. Despite the utilization of standard ontology languages, the adoption of standards for student profiles and learning object metadata still needs to be improved, reflecting a gap in standardized approaches within the field. Moreover, developing these systems often requires robust ontology-building methodologies and other associated ontology methodologies. This gap presents an area ripe for further exploration and refinement in future research endeavours. The evaluation of ontology-based recommender systems, while encompassing diverse methods such as algorithmic performance tests, statistical analyses, questionnaires, and qualitative observations, notably needs explicit ontology evaluation methodologies within primary studies [74].

In conclusion, the findings advocate for implementing ontology methodologies and integrating ontology-based recommendations into existing learning technologies. Additionally, this study encourages recommender systems in social science and humanities courses, non-higher education settings, and open learning environments, reflecting a broader potential application beyond traditional educational domains.

In their recent work [33], the researchers delve into the intricate challenges associated with ontology development. They introduce an advanced methodology termed Ontology Development Methodology (ON-ODM), designed to elevate the conceptualization process and address additional aspects, such as enrichment throughout the developmental stages. The approach, grounded in Ontology-Driven Conceptual Modelling (ODCM) and harnessing the capabilities of Natural Language Processing (NLP), is employed to craft a comprehensive tourism ontology with versatile applications, including e-tourism. The study demonstrates the ontology's effectiveness in handling specific queries through rigorous evaluation, engaging competency questions and quality metrics. It highlights its superiority in terms of conciseness when compared to existing tourism ontologies. The findings underscore the significant facilitation and enhancement of integrating ODCM and NLP in ontology development, resulting in more refined and efficacious domain representations.

The study presented in [73] focuses on harnessing the potential of ontology modelling in facilitating personalized feedback for students through formative assessments within a pedagogical framework. Despite needing a comprehensive ontology model encompassing the necessary data, this research endeavours to construct an ontology model tailored explicitly for the formative assessment using the pedagogy approach. Beginning with a review of nine articles, the study transitions into design research. It delineates two forms of formative assessments under the pedagogical approach: traditional (such as objective and essay tests) and outcome-based (including developmental portfolios, analytic rubrics, and perception rubrics). While both assessment types offer avenues for personalized feedback, they differ in several aspects: response initiation, learning domain, assessment result speed, focus, and execution duration. Employing the pedagogy approach, the design research culminates in an ontology schema comprising learning resources and formative assessments. This ontology model is the foundational structure for personalized feedback within an ontology-based recommender system [73]. Additionally, it has the potential for integration with other recommendation technologies, offering broad applicability in enhancing personalized learning experiences.

Table 3.1.1: Related Work on the Use of Ontology in E-Learning

Title	**Contribution**	**Drawbacks**
A systematic review of ontology use in E-Learning recommender system [74]	The paper explores how ontology-based recommender systems impact e-learning by involving collaborations across artificial intelligence, computing technology, education, psychology, and social sciences.	The paper centers on English-published articles in particular databases, potentially excluding relevant papers in other languages or databases. However, it doesn't delve deeply into obstacles to applying recommender systems in real learning scenarios, like technology readiness, educator adoption, and students' digital skills.
An Ontology Development Methodology Based on Ontology-Driven Conceptual Modeling and Natural Language Processing: Tourism Case Study [33]	The authors introduce ON-ODM, an advanced ontology development method. It fills gaps in existing methods and improves certain activities. This method merges ontology-driven conceptual modeling (ODCM) with NLP. ODCM enriches conceptual modeling with ontological theories, while NLP extracts potential class relations from text.	The case study in the field of tourism uses only a portion of the available data. The challenges and limitations encountered during the complete application of the case study are not fully discussed. The paper briefly mentions the process of populating the ontology with manually extracted individuals as a potential challenge.
An Ontology Model for Formative Assessments [73]	The authors propose an ontology model of formative assessment employing a heutagogical approach, while the study scrutinizes the differences between heutagogy-based assessments and traditional ones, encompassing response initiation, learning domain, result delivery speed, focus, and execution time.	The paper relies mostly on review papers, potentially missing relevant research findings. The ontology model prioritizes evaluating question components but lacks a feedback generator.

3.2 Personalized Tutoring with Learner Modeling

Information and Communication Technology (ICT) has significantly revolutionized the education landscape. Its integration into educational systems has paved the way for personalized learning experiences. Using various technological tools and platforms, ICT enables the customization of learning paths, content, and pace based on individual student needs and preferences. This adaptability enhances student engagement and comprehension, fostering a more effective learning environment. The traditional "teacher-centred" approach in education is gradually transitioning towards a "learner-centred" paradigm, facilitated by advancements in ICT. Learner modelling lies at the heart of this transformation, emphasizing the importance of understanding individual learners. Educators gain deeper insights into students' strengths and challenges by capturing and interpreting data on cognitive levels, preferences, and learning behaviours. Consequently, educational interventions can be customized to cater to diverse learner needs.

The applicability of learner modelling extends widely in the realm of online education. It stands poised to enhance teaching support methodologies, streamline educational management practices, and establish a foundation for evaluating, intervening, and predicting learner performance. Online education platforms heavily rely on learner modelling to enhance the learning experience. These models drive the development of adaptive learning systems that dynamically adjust content, pace, and assessment methods. By leveraging learner data, these platforms offer personalized recommendations, targeted interventions, and predictive analytics to optimize learning outcomes. Furthermore, learner modelling facilitates adequate teaching support, enables efficient teaching management, and is a foundation for evaluating and predicting learner performance.

Despite the promising potential of ICT and learner modelling in education, challenges exist, including data privacy concerns, interoperability among educational systems, and ensuring equitable access to technology for all learners.

The learner model's significance in education is widely acknowledged among researchers, underscoring its pivotal role in shaping the learning process. This model is a comprehensive repository, capturing and delineating learners' personal data and intricate learning behaviours crucial for informing adaptive learning platforms. As articulated by [45], an effective student model necessitates a robust data structure that encapsulates multifaceted aspects such as cognitive characteristics, proficiency levels, and learning patterns. By amalgamating these diverse elements, the learner model essentially formulates a detailed profile, essential for tailoring personalized learning experiences that cater to the individualized needs and preferences of each learner [86].

Moreover, the learner model's essence lies in its ability to efficiently organize and encapsulate individual learners' intricate and diverse characteristics. It is a structured framework that systematically categorizes and represents learners' traits, behaviours, and cognitive attributes. Through this organized framework, educational platforms and systems can better comprehend the nuances of each learner's strengths, weaknesses, and learning preferences. By leveraging this detailed learner model, educational media can dynamically adjust content, pacing, and instructional strategies, thereby optimizing the learning journey for each student, fostering engagement, and facilitating more effective knowledge assimilation.

As proposed by [86], the learner model is conceived as a structured compilation of individual learner data. This data is abstracted and processed algorithmically to encapsulate essential learner attributes. In practical applications, the modelling process involves a systematic approach that requires meticulous consideration of various elements. This includes defining objectives, collecting pertinent data, conducting thorough data analysis, and integrating feedback obtained from practical applications.

Identifying effective assessment indicators conducive to personalized learning has remained a persistent and multifaceted subject of discourse. Numerous studies underscore the pivotal role of Information and Communication Technology (ICT) solutions in bridging this gap [71]. The recent global pandemic has further highlighted the urgency to introduce educational methodologies supporting Blended Learning scenarios. In response, educators have increasingly turned to diverse solutions, including implementing Learning Management Systems (LMS) and Virtual Reality (VR) educational platforms. However, recent scholarly reviews have illuminated a crucial limitation: examining students' learning artifacts solely within a single solution may yield incomplete insights. Consequently, this approach might fail to provide comprehensive support for academic knowledge augmentation and skillset enhancement [71].

To address this challenge, the research authors introduced in [71] adopted multimodal assessments as a promising solution, offering a potential remedy to this limitation. Multimodal assessments possess the ability to transcend the confines of singular solution-based analysis. They can provide a holistic perspective, empowering educators with a more comprehensive understanding of students' learning trajectories. This research introduces an integrated Multimodal Learning Analytics (MMLA) framework. This innovative framework aims to concurrently orchestrate and classify elements encompassing students' personality traits, behavioural impacts, academic performance metrics, and practical skill development. This proposal represents a pivotal step within a broader initiative focused on Higher Education enhancement [71].

Table 3.2.1: Related Work on Personalized Tutoring with Learner Modeling

Title	Contribution	Drawbacks
Learner Modeling Framework Based on Learning Analytics [86]	The paper conducts empirical research to validate the learner modeling framework. It presents a detailed explanation of how a learner model can be formed using the proposed framework. The authors proposes a learner modeling framework from the perspective of learning analytics. The framework consists of four layers: logic layer, data layer, analysis layer, and application layer. This framework provides a systematic approach to constructing and optimizing learner models based on learning analytics.	The paper provides a brief overview of the learner modeling framework and its empirical application. It does not delve into detailed discussions of specific algorithms, techniques, or case studies. The paper does not extensively address the ethical and privacy implications of learner modeling.
The Value Proposition of An Integrated Multimodal Learning Analytics Framework [71]	The authors proposes an integrated MMLA framework that aims to orchestrate and classify students' personality traits, behavioral effects, academic performance, and practical skills simultaneously. The paper highlights the research gap in the field of multimodal assessment and its potential for providing actionable insights.	The paper mainly presents a conceptual framework and discusses the potential of multimodal learning analytics. It does not provide empirical evidence or case studies to demonstrate the effectiveness and practical implementation of the proposed framework.
Effectiveness of ontology-based learning content generation for preschool cognitive skills learning [61]	The authors proposes a child-friendly tutoring application, called CogSkills2, that dynamically generates cognitive skills learning content using ontologies as domain knowledge. The CogSkills2 application is designed for mobile devices, leveraging the benefits of mobile learning (M-learning). It takes advantage of multimedia content	The evaluation study conducted in the preschool environment involved three groups of children, which may not represent a large and diverse sample The correctness of the dynamically generated learning content is assessed through expert-based evaluation.

3.3 Ontology for Personalized Tutoring

The study presented in [2] delves into the challenges associated with creating ontology models, highlighting a common oversight among developers: the need for more emphasis on properties alongside classes. It aims to address this issue by examining the significance of a well-defined set of properties in crafting versatile ontology models that can be applied across diverse domains. To achieve this, the paper introduces innovative quality metrics focused explicitly on property components. Additionally, it presents a conversion technique designed to map foundational ontologies into adaptable models suitable for software development purposes.

By introducing these quality metrics and conversion methods, the authors of the work presented in [2] aim to showcase their practical advantages and usability. It emphasizes their effectiveness through examples drawn from the realm of knowledge modelling. Through these examples, the paper demonstrates how a robust focus on properties, alongside classes, can enhance the re-usability and adaptability of ontology models across various fields. Ultimately, the intent is to provide developers with valuable tools and insights to create more versatile and applicable ontology models for different application areas, illustrating their potential impact within knowledge representation and software development.

The research discussed in [2] emphasizes the practical utility of ontologies in software and database development by highlighting the attainment of an optimal structure. The study validates the substantial advantages of incorporating these metrics in ontology modelling by thoroughly examining newly introduced metrics and applying them to diverse examples of different complexities. The results obtained from these instances underscore the notable potential of these quality metrics to support the creation of flexible and versatile ontology models significantly.

Moreover, these metrics offer valuable guidance for ontology designers in crafting models adaptable to diverse knowledge management applications. The resultant ontologies, with a more well-balanced class structure, exhibit suitability for software engineering and knowledge management. Furthermore, the paper identifies promising future application domains for these refined ontologies, envisioning their potential integration in semantic web decision support systems, biology, intelligent e-tutor systems, and diverse engineering applications. This suggests these metrics' wide-ranging implications and utility in advancing ontology design for varied practical contexts.

Another study [60] endeavours to delve into the potential avenues for enhancing online learning experiences by harnessing the capabilities of a customized recommendation system. The focus lies in leveraging user interaction histories within the system as the bedrock for crafting personalized recommendations. Employing a multifaceted approach, the methodology includes a comprehensive review conducted across multiple stages. The primary aim is to discern the determinant factors or variables significantly influencing users' learning experiences within the system. Consequently, the study aims to uncover latent opportunities within the recommendation system, precisely honing in on personalized development pathways that cater to individual user profiles. Education has witnessed a pervasive utilization of Learning Management Systems (LMS) in contemporary times, mainly within online learning. These systems, accessible through web interfaces or mobile applications, are fundamental tools for delivering educational content. Integrating personalized recommendation systems within LMS platforms has emerged as a pivotal advancement. These recommendation systems adeptly tailor learning materials based on user behaviour and preferences. They encompass a spectrum of suggestions ranging from learning models, methodologies, and patterns to different learning stages, all finely tuned to match individual users' unique profiles and characteristics. The findings of this research are poised to shed light on the influential factors shaping user interactions within learning management systems. Furthermore, it aims to unearth the untapped potential within recommendation systems, particularly in facilitating personalized learning journeys. By identifying and understanding these influential factors, the study aims to pave the way for a more tailored and practical approach to online learning, ultimately fostering improved user experiences and enhanced educational outcomes.

Through the development of a personalized Learning Management System (LMS), a significant stride is made in addressing the challenge of treating all students uniformly by offering diverse learning materials. The outcomes gleaned from this initiative highlight a strong correlation between the presentation of different learning materials and students' active or reflective learning performance. This underscores the importance of tailoring teaching materials with a proposed taxonomy that aligns with students' learning styles. The essence of personalization becomes paramount in light of the vast expanse of information available on the internet, which often overwhelms users in their quest to locate pertinent and essential information. Herein lies the pivotal role of recommendation systems on the web, aiming to foresee and provide items or information likely to pique the users' interest or be helpful to them.

Moving forward, subsequent studies could capitalize on the insights gleaned from this review by further developing and evaluating the perceived impact of utilizing personalized recommendation systems. Such endeavours could encompass a comprehensive examination of behavioural patterns, interest variables, and the barriers that influence decision-making concerning the recommendations offered. By exploring these dimensions, future work seeks to unravel the nuanced intricacies surrounding the adoption and efficacy of personalized recommendation systems, thereby contributing significantly to the evolution and optimization of tailored learning experiences within educational platforms.

The widespread use of digital libraries among students globally stems from their convenient access to digitized study materials, e-books, and multimedia content. However, to enhance user satisfaction and loyalty, deploying a personalized digital library becomes imperative to offer tailored services. This article introduces an innovative approach to designing personalized digital libraries using the Protégé editor and machine learning methodologies.

The proposed methodology by [52] involves leveraging the capabilities of the Protégé 4.3 tool to craft digital library ontologies and identify interconnected concepts. This process encompasses domain knowledge acquisition, organizing ontologies, elaborating ontology structures, ensuring information consistency, and validating the ontologies. In addition, the article introduces the utilization of a gated recurrent unit-recurrent neural network (GRU-RNN) coupled with a deep training tree (DTT). This

GRU-RNN model is employed to predict diverse user behaviour styles encompassing cognitive behaviour, learning speed behaviour, sedentary behaviour, and aggressive behaviour. The incorporation of DTT addresses the vanishing gradient problem often associated with GRU-RNN, replacing traditional gradient-based optimization methods. Moreover, a black widow optimization approach is integrated to enhance the accuracy of the GRU-RNN network by updating its weights.

The efficiency of the methodology [52] is assessed using various performance metrics such as F-score, accuracy, loss, precision, and recall scores. The results demonstrate the potential of the personalized digital library generated through this methodology to exhibit efficiency in usability and user-centric customization. This study contributes a novel approach to developing personalized digital libraries, offering tailored services to users, and enhancing their overall experience and satisfaction within digital learning environments.

Table 3.3.1: Related Work on Ontology for Personalized Tutoring

Title	Contribution	Drawbacks
Property-Based Quality Measures in Ontology Modeling [2]	The paper highlights the importance of properties in ontology modeling and emphasizes that ontology developers often concentrate more on classes while neglecting the role of properties. The paper introduces novel quality metrics specifically related to property usage parameters in ontology modeling.	The paper outlines potential application areas for future ontologies including semantic web decision support systems, biology, intelligent e-tutor systems, and engineering applications, yet lacks detailed exploration or empirical evidence for these areas, and introduces novel quality metrics without offering a comprehensive comparison to existing ontology quality metrics.
Personalized Recommendation System for Online Learning: An Opportunity [60]	The paper highlights the importance of personalized recommendation systems in online learning by analyzing user behavior through their interaction history with the Learning Management System (LMS) to understand how personalized recommendations affect learning styles and achievements.	The paper explores the potential advantages of personalized recommendation systems in online learning but lacks empirical evidence or evaluation of their perceived impact, while also not addressing the technical implementation aspects within learning management systems.
Personalized ontology and deep training tree-based optimal gated recurrent unit-recurrent neural network for prediction of students' behaviour [52]	The paper introduces a unique approach utilizing the Protégé editor and machine learning methods to craft personalized digital libraries, emphasizing the significance of tailored services for user satisfaction and loyalty, while also constructing a gated recurrent unit-recurrent neural network (GRU-RNN) integrated with a deep training tree (DTT) to forecast diverse user behavioral characteristics.	The paper lacks a comprehensive exploration of the unique context, challenges, requirements, and limitations associated with digital libraries and fails to compare the proposed scheme with existing approaches or systems for personalized digital libraries.

CHAPTER 4

Problem Statement and Proposed Method

4.1 Problem Statement and Contributions

This chapter begins by delineating the issue targeted for resolution within this thesis, then outlining the proposed methodology devised to address and overcome the identified problem.

4.1.1 Problem Statement

Effective personalized tutoring hinges on dynamically responding to learner progress and tailoring content difficulty accordingly. The development of scalable, efficient, and adaptable personalized tutoring systems is imperative to address the multifaceted challenges posed by the diverse and evolving educational landscape. Prior efforts within the research group yielded foundational ontologies, albeit needing more semantic richness. External research emphasizes recommendation aspects but often needs to improve in effective dialogue management, as noted by [3]. The absence of a scalable ontology emerges as a significant obstacle in crafting robust personalized tutoring systems, restricting their capacity to cater to a broad spectrum of learners [49]. The inherent challenge lies in the manual and time-intensive construction of educational ontologies from datasets, impeding the scalability and efficiency of personalized tutoring systems. Recognizing this, the thesis endeavours to pioneer an automated ontology construction for personalized tutoring to augment the knowledge base and introduce a novel, semantically rich approach to ontology building.

4.1.2 Contribution

After Maaz [46] integrated facial expression animation and lip-syncing capabilities into the e-tutoring system and the development of a Tutoring ontology by Ashwitha [43], the plan started to take some shape but lacks the robustness as the ontology was very finite.

This thesis endeavours to advance the landscape of personalized tutoring through a multifaceted approach. Firstly, we introduce a meta-level framework and a sophisticated algorithmic application. This framework is a pivotal contribution, laying the groundwork for creating highly specialized, domain-specific ontologies meticulously crafted for the unique demands of personalized tutoring systems.

Taking a significant step forward, our research extends its influence by presenting intricate algorithms explicitly designed for Automatic Ontology Construction. These algorithms are meticulously crafted to autonomously construct domain-specific ontologies, leveraging the wealth of information within publicly accessible datasets. This innovation ushers in a new era of efficiency and scalability in personalized education, addressing the evolving demands of modern learning environments.

The thesis's contribution doesn't stop at ontology construction, it delves deeper into interaction modes. Our research enhances personalized tutoring by seamlessly integrating domain ontologies into learning. This integration facilitates two distinct and versatile interaction modes: quiz style and question-answer style. These modes are not only adaptive to individual learner preferences but also serve to amplify the overall effectiveness of the proposed educational approach.

To conclude, this thesis aims to enhance personalized tutoring by introducing a meta-ontological framework, automating ontology construction, and enhancing interaction modes. By combining these elements, we aim to contribute to the evolution of educational methodologies, offering a more tailored, efficient, and interactive approach to personalized learning [16].

4.2 Algorithms for Ontology Construction

Within the paradigm of a POMDP dialogue manager, the conversational dynamics are intricately modelled as a probabilistic decision process. This framework operates on the premise that the system's actions are dictated by its belief state, which encapsulates knowledge and uncertainty about the user's intentions. At each step of the dialogue, this belief state undergoes dynamic updates, incorporating observations and prior knowledge, thus evolving in response to the evolving conversation. The POMDP dialogue manager orchestrates various methodologies from reinforcement learning, probabilistic reasoning, and decision theory. This amalgamation is strategically employed to ascertain the optimal action for the system in light of the prevailing belief state. A pivotal aspect of this decision-making process is the judicious consideration of the exploration-exploitation trade-off. This intricate balance is aimed at navigating the dual objectives of gathering more insights into the user's intentions—thereby reducing uncertainty—while maximizing the system's overall performance. The POMDP system, responsible for determining the system's text-based responses, consists of several vital components. The State Estimator generates the system's current believed state, forwarded to the Belief State History (BSH) database to maintain a comprehensive history of belief states. The BSH Storage retains the historical belief states, providing additional insights into user interactions. Within this system lies a pivotal ontology, the primary repository for storing domain knowledge and knowledge rules. The ontology assumes a critical role in the intelligent processing of user queries, playing a direct and decisive part in determining the system's responses and subsequent actions. Its responsibility extends beyond mere storage, the ontology performs the intricate task of deciphering the appropriate response tailored to the user's input. Moreover, it intricately outlines the systematic steps to be undertaken by the system, orchestrating a coherent and contextually relevant interaction. In essence, this ontology acts as the cognitive backbone, bridging the gap between user queries and system responses, ensuring a seamless and informed dialogue within the defined domain. The intricacies of this ontology's design and functionality form

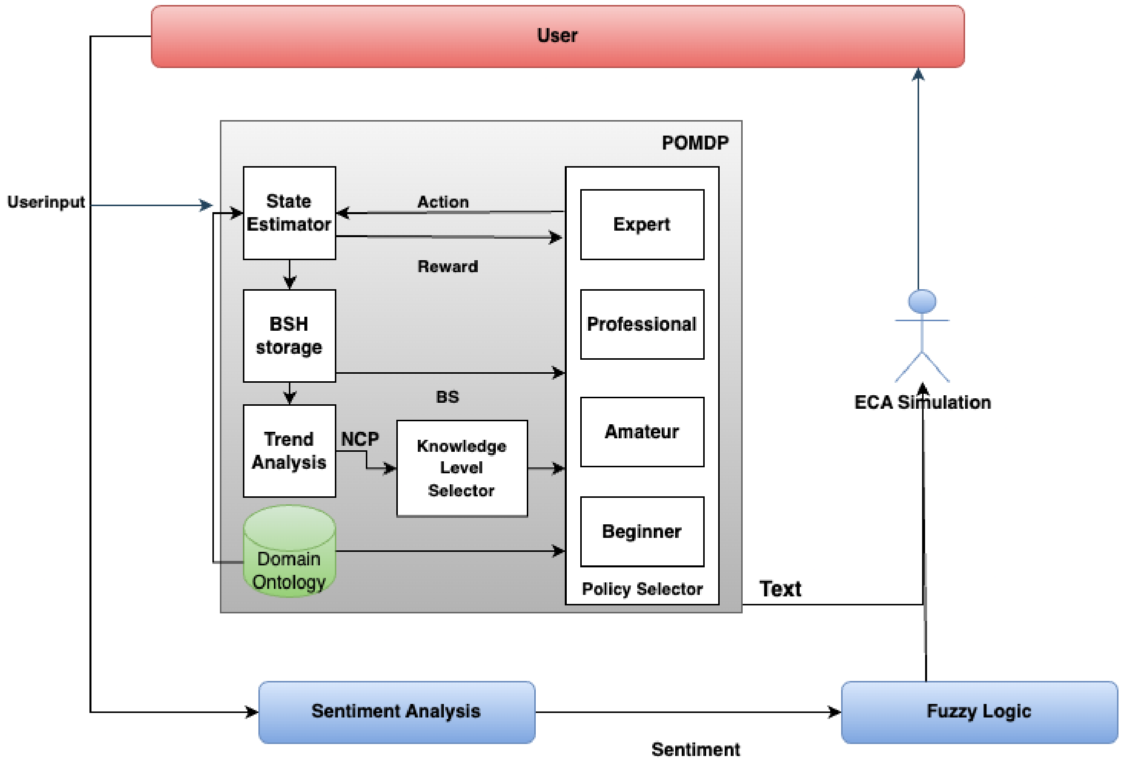

Fig. 4.2.1: Proposed Architecture

a cornerstone in the architecture, fostering a robust and intelligent system capable of navigating the complexities of user interactions with precision and efficacy. Trend Analysis, employing Discrete Wavelet Transformation (DWT), identifies sharp variation points, influencing the knowledge level determination. The Number of Change Points (NCP) then dictates the knowledge level, guiding the selecting of an appropriate knowledge policy by the Policy Selector module. Based on the knowledge level and rewards from the state estimator, this module determines the mode and optimal text actions for integration into the broader system. The model is subsequently incorporated into the architecture's lower right section, utilizing fuzzy logic to ascertain the emotion displayed by the Embodied Conversational Agent (ECA). Conversely, the sentiment analysis module uses fuzzy logic to extract sentiment from the user's input, with the fuzzifier determining the appropriate emotion for the ECA to express. In essence, the POMDP dialogue manager represents a fusion of advanced computational techniques, leveraging probabilistic reasoning to navigate the inherent uncertainties in user interactions. By incorporating insights from reinforcement

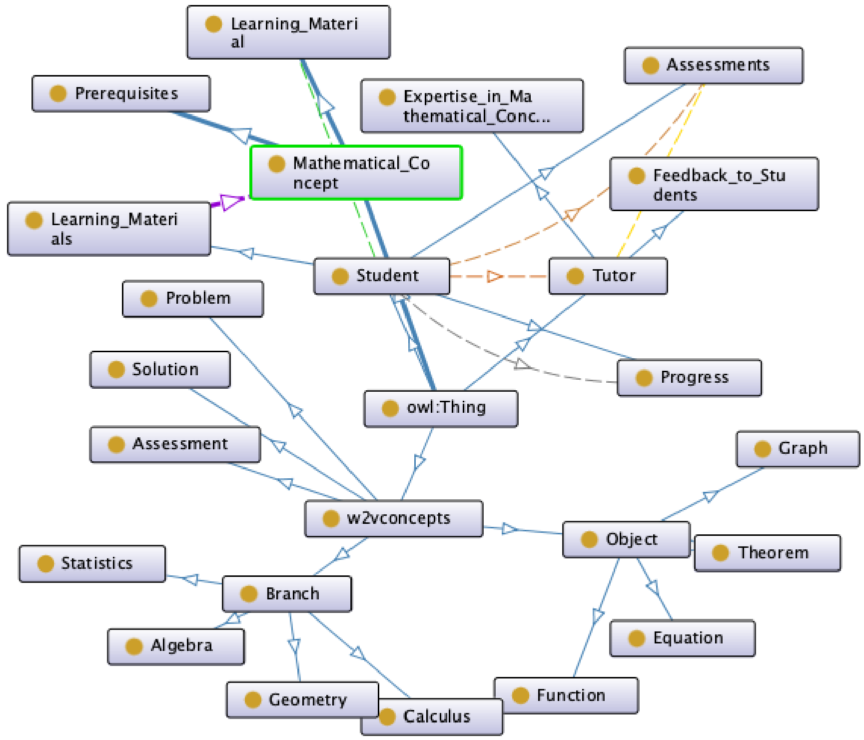

Fig. 4.2.2: Proposed Meta Ontology

learning and decision theory, the system adapts dynamically to the evolving dialogue, ensuring an optimal interplay between exploration, exploitation, and the pursuit of system efficiency. This approach, rooted in the principles of POMDP, stands as a testament to the intricate and nuanced nature of modelling conversational interactions within intelligent systems, as expounded in my thesis document.

4.2.1 Meta Ontology

A meta-ontology is a foundational blueprint for constructing diverse ontologies, offering a set of general concepts and relationships that transcend specific domains [47]. Unlike domain-specific ontologies, meta-ontologies abstain from incorporating detailed domain-specific information but provide a robust framework for developing such details. This role is pivotal in facilitating knowledge organization and sharing across various domains [41]. In our personalized tutoring system, meta-ontology is a crucial architectural component designed with scalability. Crafted explicitly for our

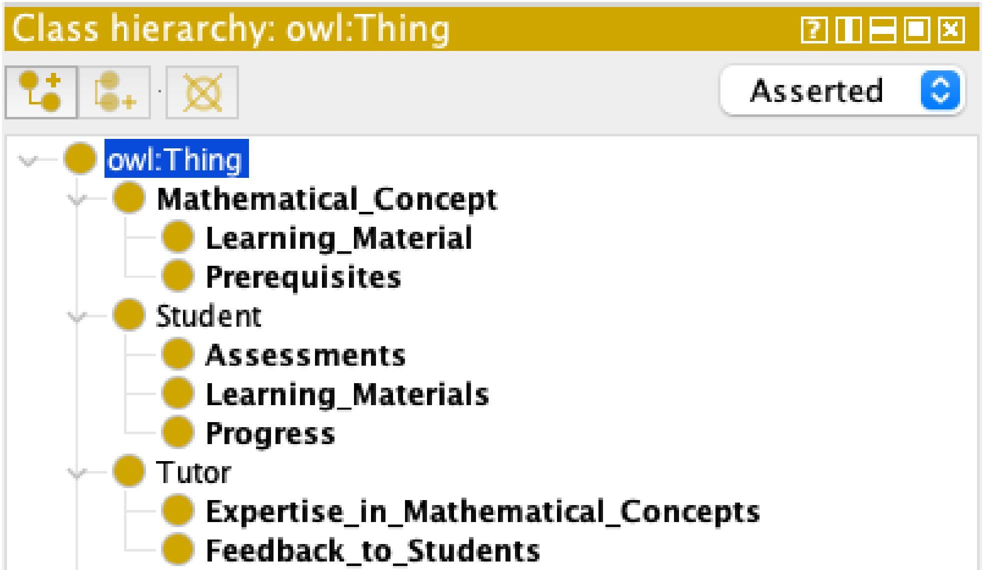

Fig. 4.2.3: Detailed Parent-Child Relationship in Meta Ontology

personalized tutoring system, it features intricate relationships and predefined classes tailored to the nuances of tutoring scenarios. This meta-ontology is not merely a static blueprint but a dynamic and adaptable foundation that underpins the construction of our Tutoring Ontology. The complexity within its relationships and the predefined classes cater to the unique demands of personalized tutoring, emphasizing its role as a cornerstone for the knowledge representation and organization within the Tutoring Ontology. This meta-ontology aligns with the broader objective of ontology engineering by providing a structured yet flexible framework that accommodates the evolving nature of the personalized tutoring domain. Its significance lies in its ability to capture the essential abstractions and relationships that characterize tutoring scenarios, ultimately contributing to the robustness and adaptability of the entire ontology framework within our personalized tutoring system.

4.2.2 Dataset Processing

The dataset under consideration comprises textual information and elements that are either non-textual or lack meaningful content. Non-textual information, exemplified by figures and similar components, may need to be considered during preprocessing to alleviate the overall file size. Regular expressions (Regex) can be tailored to match

```
print(keywords)
```

```
['mathematics', 'mensuration', 'shapes', 'squares', 'rectangles', 'circle
s', 'triangles', '3d', 'shapes', 'cubes', 'cylinders', 'area', 'side', 'l
ength', 'width', 'circumference', 'radius', 'pi', 'diameter', 'base', 'he
ight', 'perimeter', 'hypotenuse', 'leg', 'volume', 'tower', 'tank', 'heig
ht', 'sculpture', 'surface', 'area', 'prism', 'garden', 'figure', 'regio
n', 'path', 'center', 'centroid', 'container', 'portion', 'edge', 'spac
e', 'roof', 'slant', 'height', 'diameter', 'base', 'dimensions', 'solid']
```

Fig. 4.2.4: POS tagging & removal of stop words

```
print(Lemma)
```

```
['mathematics', 'mensuration', 'shape', 'square', 'rectangle', 'circle',
'triangle', '3d shape', 'cube', 'cylinder', 'area', 'side', 'length', 'wi
dth', 'circumference', 'radius', 'pi', 'diameter', 'base', 'height', 'per
imeter', 'hypotenuse', 'leg', 'volume', 'tower', 'tank', 'sculpture', 'su
rface area', 'prism', 'garden', 'figure', 'region', 'path', 'center', 'ce
ntroid', 'container', 'portion', 'edge', 'space', 'roof', 'slant height',
'base dimension', 'solid']
```

Fig. 4.2.5: Lemmatization of extracted tokens

specific patterns associated with figures, images, or other non-textual content, allowing for targeted removal. This strategic elimination of non-textual elements aims to streamline the dataset, ensuring a more compact representation focusing on textual content. Additionally, text fragments devoid of domain-specific information pose another facet for refinement. Text sections lacking substantial domain relevance can be selectively removed from the dataset. This deletion process contributes to creating a more condensed text dataset, enhancing its relevance and informativeness for subsequent analyses. The overarching goal is to optimize the dataset's composition, retaining only the text elements that carry meaningful domain-specific information while discarding extraneous or redundant components. This meticulous curation process ensures that the resultant dataset aligns more closely with the specific requirements and objectives of the research outlined in my thesis. The implementation involves the utilization of the SpaCy module [39] to execute POS-tagging, a crucial linguistic analysis process. This operation transforms the input string into a nested list structure, encapsulating each sentence within a distinct list entry. Tokenization is then carried out, extracting individual tokens from the text. Notably, tokens are separated based on interpunction and space characters. Further refinement of the linguistic analysis is achieved through lemmatization, leveraging the vocabulary provided by [85]. This comprehensive step involves reducing each token to its base or root form, facilitating a standardized and uniform representation of words within the text. The lemmatization process contributes to the normalization of the data, enhancing the accuracy and consistency of subsequent analyses. Moreover, each dish is categorized within every sentence by assigning it to a specific lexical category, such as a noun, verb, number, etc. This categorization provides valuable insights into the grammatical and semantic structure of the text, laying the foundation for more in-depth linguistic and computational analyses.

4.2.3 Annotation of Extracted Tokens

A crucial step in the ontology construction process involves filtering and selecting relevant terms to serve as ontology classes. Since ontology classes predominantly consist of nouns, the dataset undergoes a meticulous filtering process, retaining only tokens categorized as 'nouns' and 'proper nouns'. This targeted selection ensures that the subsequent procedures focus exclusively on linguistically appropriate elements for ontology construction. To annotate extracted tokens, We remove and store ontology class definitions in hierarchical dictionaries for easy access. Each ontology has its glossary of class names and meanings, ensuring efficient retrieval. Extracted tokens are matched with class names, and tickets and definitions are stored for review.

Following the token selection, a comprehensive search is conducted within the ontologies to identify and quantify the presence of these selected tokens as classes. This step provides valuable insights into the distribution of relevant terms across different ontologies, aiding in the strategic decision-making process for choosing a base ontology for further extension steps. The quantitative analysis of token occurrences within each ontology is a foundational metric for determining the ontological groundwork that best accommodates the intended expansion. An additional layer of information is extracted by capturing the definitions associated with these classes within the ontologies to enhance the understanding of the selected classes. The extraction of class definitions, stored as string values, contributes to the interpretability

```
print(math_dictionary)
```

```
{'Abacus': 'An early counting tool used for basic arithmetic.', 'Absolute Value': 'Always a p
ositive number, absolute value refers to the distance of a number from 0.', 'Acute Angle': 'A
n angle whose measure is between 0° and 90° or with less than 90° (or pi/2) radians.', 'Adden
d': 'A number involved in an addition problem; numbers being added are called addends.', 'Alg
ebra': 'The branch of mathematics that substitutes letters for numbers to solve for unknown v
alues.', 'Algorithm': 'A procedure or set of steps used to solve a mathematical computatio
n.', 'Angle': 'Two rays sharing the same endpoint (called the angle vertex).', 'Angle Bisecto
```

Fig. 4.2.6: Reference Dictionary created from Meta Ontology & Ontario Ministry of Education Website

```
{'area': 'The measure of the space inside a two-dimensional shape.', 'side': 'One of the segm
ents that make up the boundary of a geometric figure.', 'length': 'The measurement of somethi
ng from end to end.', 'width': 'The measurement of something from side to side.', 'circumfere
nce': 'The distance around the edge of a circle or any curved geometric figure.', 'radius':
'The distance from the center of a circle to its outer edge.', 'pi': 'A mathematical constant
approximately equal to 3.14159, used in calculations involving circles.', 'diameter': 'The di
stance across a circle through its center, or the length of a straight line segment that pass
es through the center of a circle.', 'base': 'The bottom side or face of a geometric figur
e.', 'height': 'The vertical measurement from the base to the top of a geometric figure.', 'p
```

Fig. 4.2.7: Annotated tokens using keyword matching

and semantic richness of the ontology. This extraction facilitates subsequent steps in the ontology construction process and provides a basis for easy evaluation and refinement by domain experts. Including expert input becomes pivotal in ensuring the accuracy and relevance of the ontology's conceptual framework.

In selecting a suitable ontology for the dataset and enhancing it through integrating concepts acquired during pre-processing, a crucial step involves understanding the existing definitions of tokens within the ontologies. To achieve this, Python code has been developed, leveraging the owlready2 library [39] to load ontologies from a local database dynamically. Subsequently, the code extracts all class labels and their corresponding definition strings from the ontologies, organizing this information into key-value pairs within dictionaries. The structure of these dictionaries is designed to store the class names and their definitions efficiently. A nested approach is employed, where a dictionary is dedicated to each ontology, with the ontology name as the key and the nested dictionary containing class names and their definitions as the corresponding value. This hierarchical organization facilitates the systematic retrieval of class information based on ontology names. Upon completing this data extraction and organization process, the next step involves processing the tokens identified through text extraction. The developed Python code systematically navigates the dictionaries,

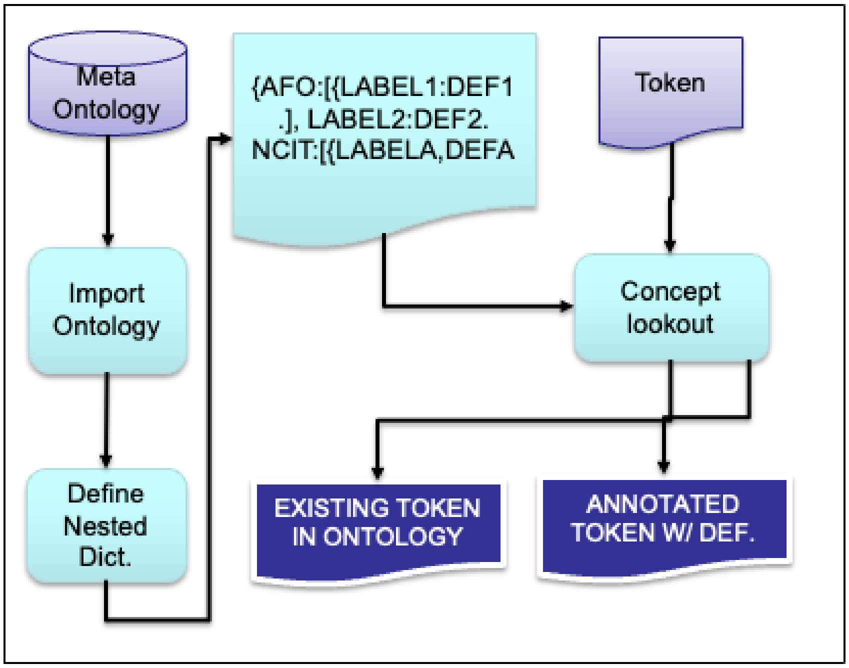

Fig. 4.2.8: The code's workflow developed for annotating the extracted tokens.

specifically targeting class names, to identify and match these extracted tokens. This meticulous matching process ensures the accurate association of tokens with their corresponding class names within the ontologies. Moreover, the tokens extracted during the text processing phase are systematically stored in a dedicated table, accompanied by their corresponding definitions. Each token is explicitly associated with its source ontology, facilitating a streamlined review process by domain experts at a later stage. The workflow designed for annotating these extracted tokens is visually represented in Figure 4.2.4. The components requiring input, namely the Meta ontology and the tokens acquired through text extraction. On the other hand, the output elements, depicted in dark blue, signify the results and insights generated through the workflow. This structured approach ensures efficient token management and linkage to source ontologies and establishes a clear pathway for expert evaluation and validation. The integration of these processes contributes to the overall robustness and reliability of the annotation system, aligning with the quality standards expected in my research as outlined in my thesis document.

4.2.4 Automatic Creation of New Class

The training of the Word2Vec model involves utilizing the textual data acquired through the methods detailed in Section 4.2.2. A vector size 300 was chosen for stability and consistency in representing words or phrases. Notably, while the Word2Vec model can be applied to hierarchical clustering, the resultant clusters do not inherently yield ontological or semantic hierarchies. This limitation arises from the nature of relations between tokens extracted through the vectorization of concepts. Despite capturing semantic similarities within text clusters relevant to knowledge domains, the Word2Vec model does not inherently provide classification or hierarchical information conducive to establishing ontology classes and their respective subclasses. Consequently, employing hierarchical clustering techniques such as dendrograms may not necessarily yield structured concepts into ontological hierarchies. However, the strength of Word2Vec lies in its ability to generate tokens with high cosine similarity to an initial input concept, offering valuable insights into semantic associations within the textual data.

The output generated by the workflow outlined in Section 4.2.3 is employed to leverage the functionality of similar tokens. This workflow serves a dual purpose by not only annotating tokens in a text dataset with definitions sourced from ontologies but also providing information on tokens already present in each investigated ontology. By selecting the ontology with the most common classes, the tokens already contained in these classes serve as input for the Word2Vec model, which has been trained on the text dataset. Subsequently, this model is utilized to identify the closest tokens based on cosine similarity to the input word. The process incorporates a threshold value that regulates the number of output tokens, considering the minimum allowable cosine similarity. For instance, setting a minimum cosine similarity of 0.999 would yield only tokens closely aligned with the input. In contrast, a minimum similarity of 0.8 would encompass broader tokens positioned farther away in the vector space. Given that these tokens exhibit the highest similarity to the already contained ontology class, the relationship between the ontology class and the tickets retrieved

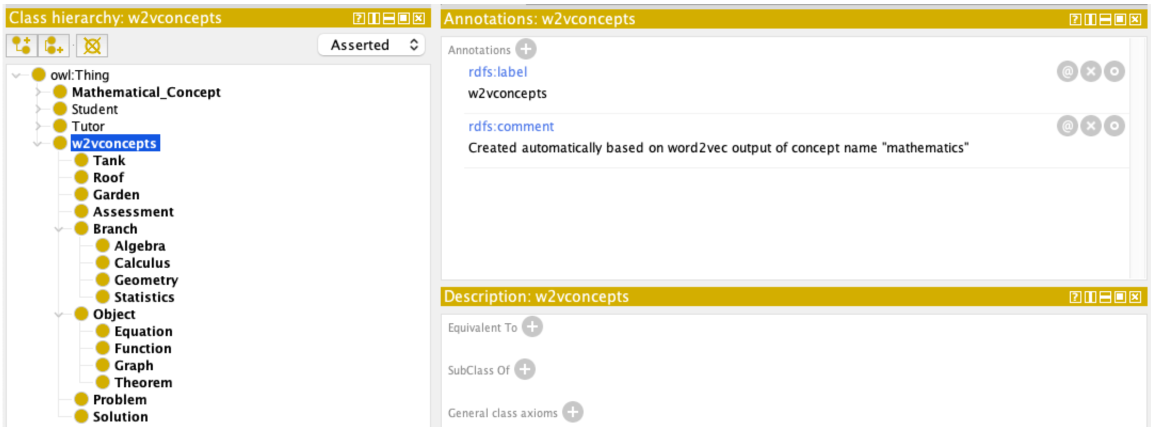

Fig. 4.2.9: Annotations of new class w2v visualized in Protégé for later review by domain expert

through Word2Vec is presumed to have some semantic relevance. This approach enhances the semantic understanding and association between ontology classes and corresponding keys, contributing to a more nuanced and contextually rich representation. In the context of Word2Vec output, the process involves generating tokens that may not already exist within the ontology. When such a scenario arises, a proactive approach is taken to create new classes, each reflecting a specific token. A comprehensive category named "w2vConcept" is established to efficiently manage these newly introduced classes, functioning as a subclass of the "owl: Thing" class. Tokens produced by the Word2Vec model, which still need to be present in the ontology, are then instantiated as classes. Additionally, they are designated as subclasses of the automatically generated "w2vConcept" class, which, in turn, is positioned as a subclass of the ontology root class "owl: Thing" This structuring facilitates the subsequent revision of the automatically generated classes, providing ease of identification within an ontology editor such as Protégé, mainly when listed as subclasses of the same category. Moreover, this approach ensures that the integration of new classes preserves the semantic integrity of the ontology. An object property named "conceptually related to" is automatically created to establish relationships between these unique classes. This property aims to simplify the subsequent definition of the precise relationship between these classes, streamlining the process of identifying conceptual associations as indicated by the Word2Vec model.

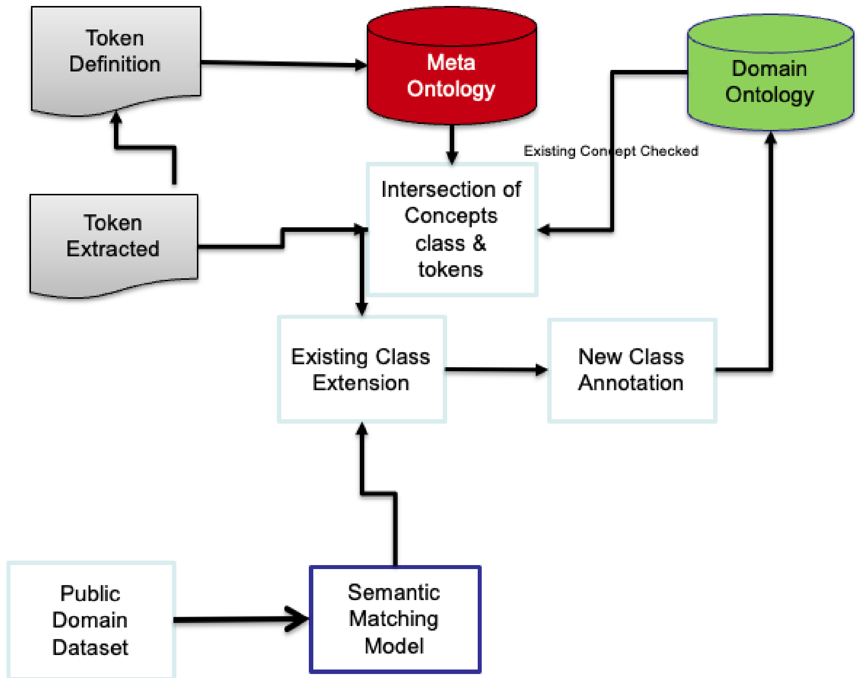

Fig. 4.2.10: Overall Algorithm

The methodology outlined in Section 4.2.3 is employed in annotating classes with missing definition strings. This workflow entails searching for definition strings for the newly generated classes within other semantic artifacts. It's important to note that the code cannot autonomously determine the most suitable definition when confronted with multiple options. Consequently, each obtained definition string is meticulously listed in a distinct 'rdfs: comment' field associated with the respective class. Additionally, a note detailing the source of the definition is appended to provide transparency and context regarding the origin of the information. This approach ensures a comprehensive annotation of classes, incorporating relevant definition strings and their sources, thus contributing to the clarity and comprehensiveness of the semantic artifacts in the ontology. Domain experts can systematically review the newly introduced classes after storing the resultant extended ontology. This review process involves a comprehensive assessment, allowing experts to make informed decisions regarding accepting or rejecting these added classes. Moreover, domain experts have the flexibility to refine the conceptual relationships associated with these classes, tailoring them to better align with the specific requirements or nuances of the given domain. This iterative workflow, depicted in Figure 4.2.10 of the thesis, outlines the steps in automatically extending an ontology. The initial input is represented by the meta ontology, indicated in red, which serves as the foundation for the extension process. The resulting domain ontology, coloured in green, represents the output of this workflow. This systematic approach ensures that the extended ontology undergoes expert scrutiny and refinement, aligning it more closely with the domain's nuanced understanding and specific needs.

4.2.5 Construction of Domain Ontology

The algorithm for constructing a domain ontology through an automated extension process involves several key steps to seamlessly integrate new classes based on semantic similarities identified by the Word2Vec model. The workflow is designed to ensure the preservation of semantic integrity and the facilitation of subsequent expert review and refinement. The step by step process is detailed as follows:

1. Word2Vec Model and Clustering: The Word2Vec model is initially trained on textual data to capture semantic relationships. However, the resulting clusters lack ontological or semantic hierarchies due to the nature of the relations between extracted tokens. Hierarchical clustering techniques are not employed, as they do not provide classifications of concepts or subclasses of ontologies. 2. Utilizing Word2Vec for Semantic Similarity: Word2Vec is harnessed to identify tokens with high cosine similarity to an initial input concept. The workflow annotates these tokens with definitions from existing ontologies, determining which tokens are already present in each ontology. 3. Automatic Creation of New Classes: If a token identified by Word2Vec is not already present in the ontology, a new class is dynamically generated to represent that token. This new class is categorized under an overarching class called "w2vConcept", positioned as a subclass of "owl: Thing". This accommodation ensures the inclusion of newly introduced classes that may still need to be defined through semantic means. 4. Preservation of Semantic Integrity: The newly created classes are explicitly set as subclasses of the "w2vConcept" class establishing

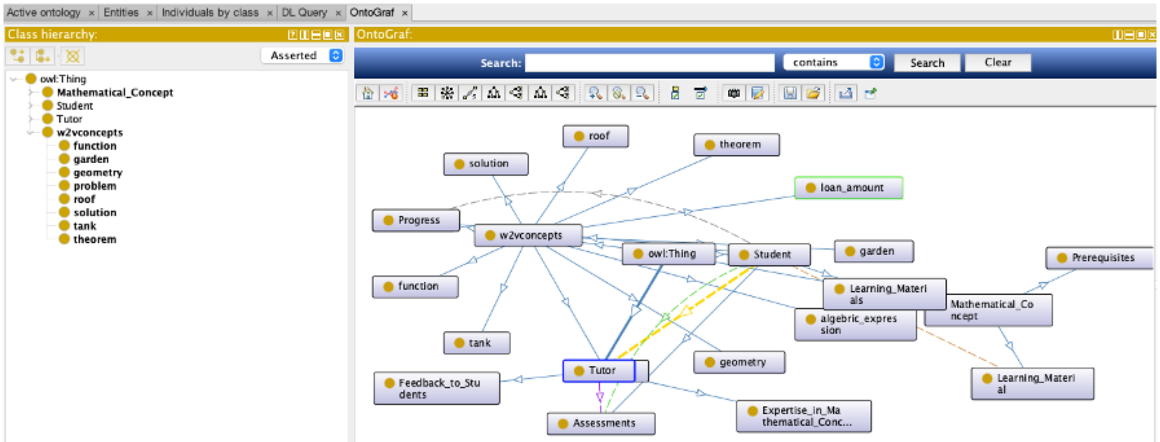

Fig. 4.2.11: New Concepts added to the Class w2vconcepts

a hierarchical structure that preserves the semantic integrity of the ontology. This structured organization facilitates the subsequent revision of automatically created classes using popular ontology editing tools like Protégé. 5. Conceptual Relationships and Definitions: Unique classes are interconnected through an automatically established relationship termed "conceptually related to". This relationship assists in precisely defining the associations between the classes. The workflow searches for definitions in other semantic artifacts to enrich classes with missing definition strings, documenting each obtained definition string along with its source. These definitions are then included as "rdfs: comment" annotations of the respective classes.

The overall outcome of this comprehensive workflow is the automatic extension of the ontology, incorporating new classes derived from semantic similarities identified by the Word2Vec model. The resulting extended ontology provides a foundation for domain experts to review and refine the newly added classes, ensuring alignment with the domain's nuanced understanding and specific requirements.

4.2.6 Automatic Creation of New Classes

To automate the creation of new classes within the domain ontology, our methodology was applied to a comparable dataset, specifically the HotPotQA, renowned for featuring natural, multi-hop questions [97]. The iterative process was initiated afresh, beginning from Section 4.2.2 and progressing through Section 4.2.5. Focusing on a limited dataset segment, our approach demonstrated its capability to discern and extract essential concepts such as “interest” and “loan amount” from the dataset.

This application of our technique illustrates its adaptability and effectiveness in scaling up ontologies across diverse datasets. The meticulous execution of Sections 4.2.2 to 4.2.5 showcases how the process seamlessly extends to comprehend intricate concepts embedded in the dataset. Notably, this ability to traverse varying datasets underscores the methodology’s potential to generate more elaborate domain ontologies across various domains. This success in automating the creation of new classes exemplifies the versatility and efficiency of our approach in handling diverse data sources and producing sophisticated domain ontologies tailored to specific domains.

4.3 Algorithm of Personalized Tutoring

The following section delineates the intricate details of the algorithm for implementing personalized tutoring. Within this section, we delve into the systematic structure of the algorithm and its integration into the broader system architecture. Emphasizing the pivotal role of ontology in personalized tutoring, we explore how it serves as a foundational component, enabling the system to tailor educational interactions based on individual user needs and preferences.

4.3.1 System Structure

The foundation of the tutoring system’s functionality lies in the intricacies of its system structure, which orchestrates a dynamic and adaptive learning environment. Commencing with the user’s initial knowledge level as the point of origin, the tutoring session unfolds through a meticulous knowledge assessment process and tailored question retrieval. Estimating the user’s knowledge level is a multifaceted task, intricately woven with factors like NCP scores and user interactions, encompassing a comprehensive evaluation of previous responses and performance metrics. The semantically enriched ontology emerges as a linchpin in this process, serving as the backbone for categorizing and organizing questions. It is pivotal in ensuring the retrieved ques-

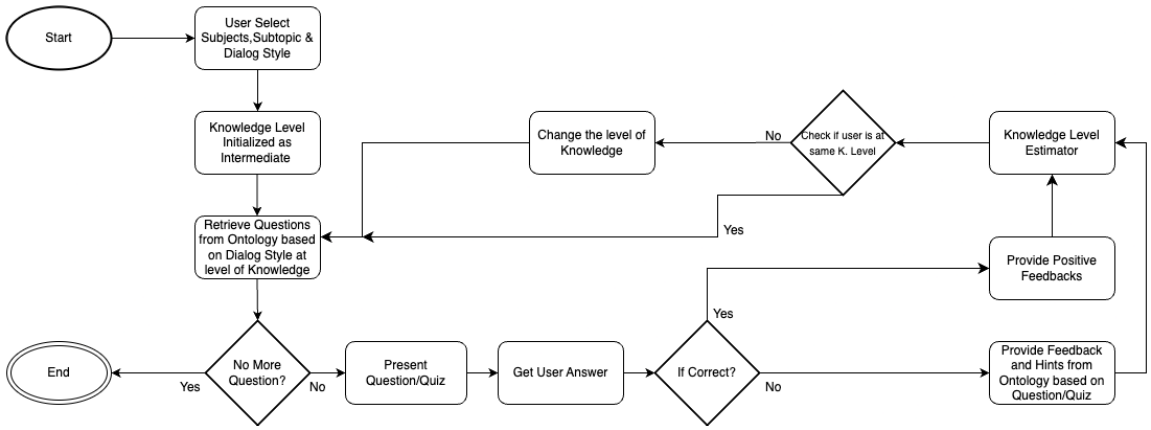

Fig. 4.3.1: Improved Personalized E-Learning Flow

tions align precisely with the user's estimated knowledge level. By leveraging the contextual relevance encoded in the ontology, the system adeptly tailors the learning experience, providing questions that are appropriate and align with the user's proficiency.

Evaluating user responses represents a critical juncture where the system employs predefined criteria to discern correctness. Driving by user performance, knowledge level adjustments introduce a dynamic element to the learning trajectory. Consecutive correct answers propel an upward adjustment, signifying enhanced proficiency, while a series of incorrect responses may prompt a downward adjustment, signalling the need for foundational content. In tandem with these processes, the system actively engages the user through informative feedback. Positive reinforcement is judiciously employed for correct answers, fostering a positive learning experience. Simultaneously, the system provides guidance and hints for incorrect responses, nurturing a supportive and instructive environment. This intricate system structure exemplifies a nuanced approach to personalized tutoring, where adaptability, context-awareness, and user-centric feedback converge to optimize the learning journey.

4.3.2 Use of Ontology for Personalized Tutoring

In personalized tutoring, utilizing ontology forms a foundational component, orchestrating a dynamic and adaptive learning experience for users. The initiation of the tutoring session hinges upon establishing the user's initial knowledge level as a baseline. This critical starting point sets the stage for a tailored educational journey. The system employs a sophisticated approach to estimate the user's knowledge level, drawing insights from NCP scores and user interactions. This estimation considers multifaceted factors, including the user's historical responses and overall performance.

The semantically enriched ontology emerges as a critical player in the subsequent phases of the tutoring process. With a nuanced understanding of the user's estimated knowledge level, the system navigates the ontology to retrieve questions meticulously tailored to the user's proficiency. The ontology categorizes and organizes questions, ensuring their relevance and appropriateness within the user's contextual learning journey. This strategic use of ontology enriches the tutoring experience by aligning content with the user's proficiency, fostering a targeted and effective learning environment.

A crucial evaluation process ensues as the user engages with the system by responding to questions. The system employs predefined criteria to assess the correctness of user responses. Based on this evaluation, the user's knowledge level adjustments are dynamically made. A series of consecutive correct answers may lead to an upward adjustment, reflecting the user's enhanced proficiency. Conversely, straight incorrect answers prompt a downward adjustment, signalling the need for a more foundational understanding of the content.

Integral to the personalized tutoring paradigm is the continuous provision of feedback to the user. The system strategically integrates feedback mechanisms, encouraging positive responses and offering guidance and hints for incorrect answers. This feedback loop contributes to the user's understanding and shapes the tutoring system's adaptive nature, tailoring the learning experience in real time based on user interactions.

In summary, incorporating ontology in personalized tutoring is the linchpin orchestrating a seamless and adaptive educational journey. From the initial estimation of the user's knowledge level to the dynamic adjustment based on performance, the interplay between ontology and user engagement ensures a personalized and practical tutoring experience.

CHAPTER 5

Implementation and Experiments

This chapter delves into the intricate details of the implementation process, offering comprehensive insights and running examples that vividly showcase the utilization of a Domain Ontology within the context of personalized tutoring. At the core of this thesis lies the pivotal concept of constructing a Robust Domain Ontology meticulously designed for integration into an e-tutoring system. The primary objective is to empower the tutoring system with the capability to furnish users with constructive feedback, hints, and valuable suggestions. This innovative approach not only involves the creation of a comprehensive ontology but also extends to incorporating an e-tutoring model that adeptly observes and monitors the learner's historical interactions. The outcome is a dynamic system that delivers tailored actions that enhance the learning experience. By leveraging the Robust Domain Ontology, the e-tutoring system becomes adept at deciphering the learner's trajectory, adapting its responses, and providing targeted interventions to optimize the learning journey. This chapter serves as a detailed exploration of the practical implementation of these concepts, offering concrete examples to underscore the effectiveness and functionality of the proposed approach in the realm of personalized tutoring.

5.1 Software and Tools

Table 5.1.1 furnishes an exhaustive inventory of software applications and tools employed throughout the thesis undertaking, accompanied by a delineation of their specific functions and objectives. The subsequent section delineates the primary software and tools harnessed during the project's implementation. The project primarily relies

Table 5.1.1: Software and Tools

Function	**Software/ Library**
Programming Language	Python 3.6.1
Ontology Language	OWL/RDF
IDE	Jupyter
Ontology Editor	Protégé, Owlready2, RDFLib, PyPDF2
Dataset Format	Json
Avatar Simulation	PyopenGL
Text Library	Gensim, Glob2, Spacy
Datasets	SQUAD2.02.0, HotPotQA, CoQA, CIMA

on Python as its core programming language, providing the foundational framework for implementing diverse algorithms, managing data processing tasks, and orchestrating the overarching project logic. The Owlready2 package is pivotal in this Python ecosystem, facilitating ontology-oriented programming. Owlready2 can seamlessly load OWL 2.0 ontologies as Python objects, enabling their modification, preservation, and execution of reasoning operations through HermiT, which is integrated into the package. Notably, Owlready2 ensures transparent access to OWL ontologies, a departure from conventional Java-based API approaches.

The project extensively employs RDF/OWL files in the context of ontology file formats. These formats play a crucial role in structuring and representing ontological knowledge, offering a standardized and interoperable means of encoding semantic information. The RDF/OWL file format ensures compatibility with the ontology-oriented objectives of the project, supporting effective data representation and exchange.

Additionally, the project leverages Protégé, a widely recognized ontology development tool. Protégé provides a user-friendly environment for creating, editing, and visualizing ontologies. Its intuitive interface facilitates ontology modelling, enabling users to define classes, properties, and relationships. By incorporating Protégé into the project workflow, ontology construction and management are streamlined, contributing to the overall efficiency and coherence of the development process.

5.2 Resources and Format of Datasets

This section provides a comprehensive overview of the diverse resources employed in the research, elucidating their significance in shaping the investigation's trajectory. Additionally, it delves into the intricate details of dataset formats, shedding light on the structures and conventions adhered to in representing empirical data. A meticulous exploration of this section unfolds the essential groundwork, guiding the reader through the intricacies of the datasets that form the backbone of the empirical analyses conducted in this research endeavour.

SQUAD2.0, derived from Wikipedia articles, features a compilation of questions linked to specific excerpts within the articles, serving as a valuable resource for training and assessing models focused on natural language understanding, particularly in question answering and comprehension tasks. Renowned for its meticulous annotations, SQUAD2.0 stands as a benchmark for evaluating the efficacy of machine reading comprehension systems. Conversely, CoQA introduces a layer of complexity by incorporating conversational dynamics, encompassing multi-turn interactions and the resolution of coreference. Both datasets are subject to ongoing enhancements, with the release of new versions and updates aimed at addressing limitations and furnishing a more diverse and challenging set of examples for training and evaluation purposes.

The CIMA [85] collection, which we openly share, is innovative in that it exposes students to intersecting foundational concepts across exercises, collecting multiple pertinent tutoring responses for the same input. CIMA possesses several noteworthy attributes from an educational standpoint: student role-players complete activities with fewer turns during the conversation, and tutor players adopt strategies that align with certain academic conversational norms, such as offering hints versus posing questions in appropriate contexts. This dataset facilitates a model's training to generate the subsequent tutoring utterance in a conversation contingent upon a provided action strategy.

CoQA is a substantial dataset designed for constructing Conversational question-answering systems, presenting a challenge that gauges machine proficiency in comprehending a given text passage and responding to a series of interconnected questions embedded within a conversation. This dataset encompasses an extensive compilation of over 127,000 questions paired with corresponding answers derived from more than 8,000 distinct conversations. Each talk is curated through the collaboration of two crowd workers engaging in a dialogue centred around a text passage structured in the format of questions and answers. CoQA boasts distinctive attributes, including the conversational nature of questions, the free-form text format of solutions, the provision of evidence subsequences highlighted in the passage alongside each answer, and the incorporation of courses from seven diverse domains. Notably, CoQA introduces challenging elements that need to be included in conventional reading comprehension datasets, such as coreference resolution and pragmatic reasoning.

HotpotQA introduces a novel dataset comprising 113,000 question-answer pairs derived from Wikipedia, characterized by four distinctive attributes: (1) questions necessitate locating and reasoning across multiple supporting documents for answers, (2) the questions exhibit diversity without constraints to existing knowledge bases or schemas, (3) the dataset provides sentence-level supporting facts, offering intense supervision for reasoning and facilitating explanatory predictions by QA systems, (4) it includes factoid comparison questions, designed to assess QA systems' capability to extract pertinent facts and perform relevant comparisons. This research demonstrates the challenging nature of HotpotQA for contemporary QA systems, highlighting the role of supporting facts in enhancing performance and promoting explainability in predictions.

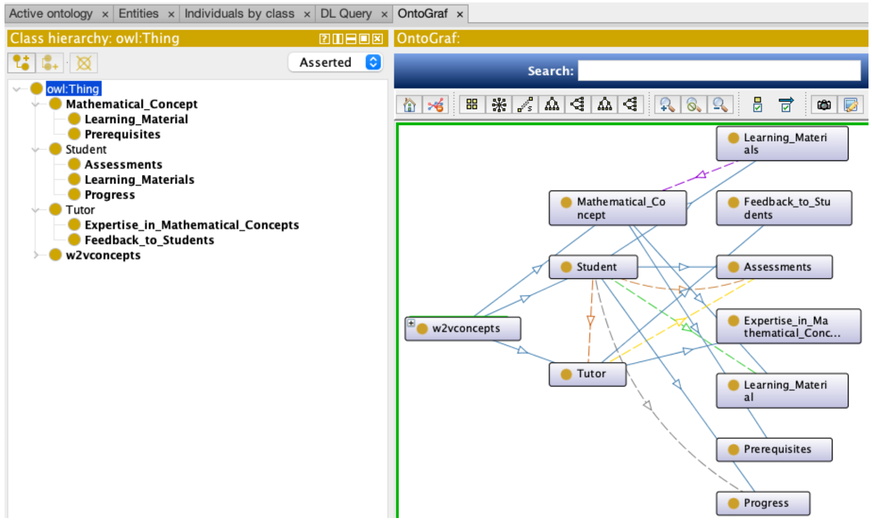

Fig. 5.3.1: Format of Meta Ontology.

5.3 Format of Meta-Ontology

A comprehensive Meta-Ontology has been meticulously crafted in personalized tutoring, laying the groundwork for developing a Domain Ontology explicitly tailored for mathematics. This endeavour culminates in an extensive literature review and a deliberate expansion upon prior research efforts. Meta-Ontology is intricately designed to align with the unique requirements of the dataset under consideration, reflecting purposeful customization for the targeted educational domain.

The formulation of the Meta-Ontology results from thorough analysis and thoughtful consideration of prevalent concepts and relationships within the field. It serves as more than a mere representation of data, it operates as a source of guiding principles and structural templates. This dual role significantly enhances the quality and integrity of the ontologies subsequently generated.

The decision-making process behind the design of the Meta-Ontology was driven by the strategic need to establish a robust and flexible foundation for the ontology construction process. By structuring the Meta-Ontology to encompass a diverse array of high-level concepts and relationships, the objective was to ensure that the automated ontology generation could effectively capture a broad spectrum of knowledge domains. This approach allows for the accommodation of various subject areas and

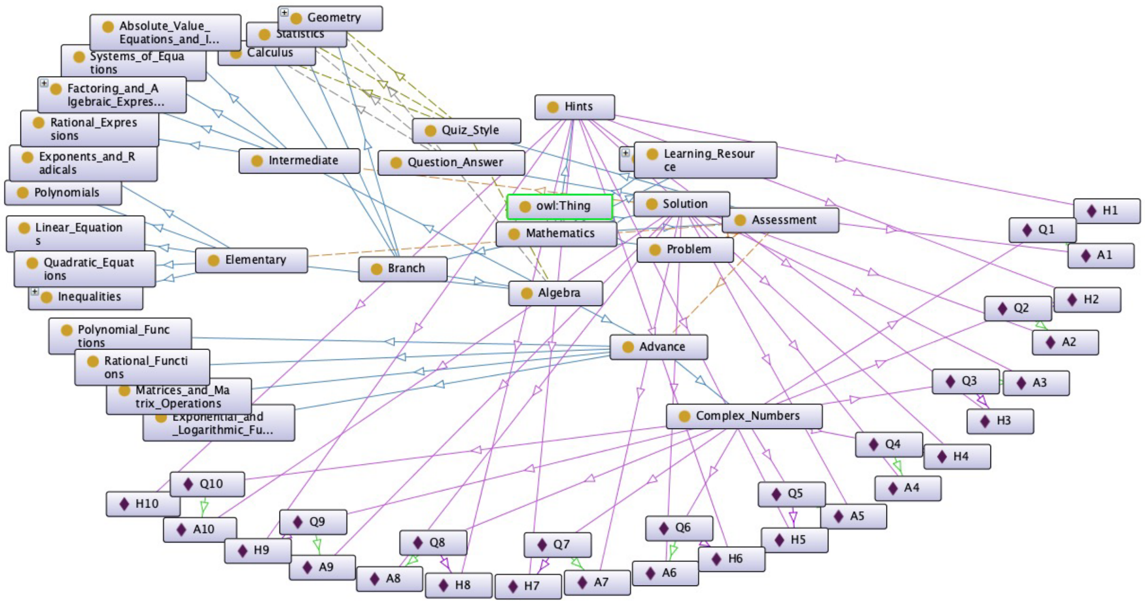

Fig. 5.4.1: Format of Domain Ontology.

facilitates adaptation to potential shifts in focus, ensuring the longevity and adaptability of the ontological framework. The Meta-Ontology, thus, serves as a pivotal component in the ontology construction process, providing a versatile and comprehensive framework that underpins the subsequent development of domain-specific ontologies.

5.4 Format of Domain-Ontology

After going through our four-step algorithm process, the initial domain ontology constructed is as shown in Figure 5.4.1. The domain ontology employed in this research is structured to encapsulate the underlying knowledge of the targeted subject area, ensuring a comprehensive and well-organized representation. The format adheres to established ontology principles, incorporating a hierarchical structure, clearly defined classes, and explicit relationships

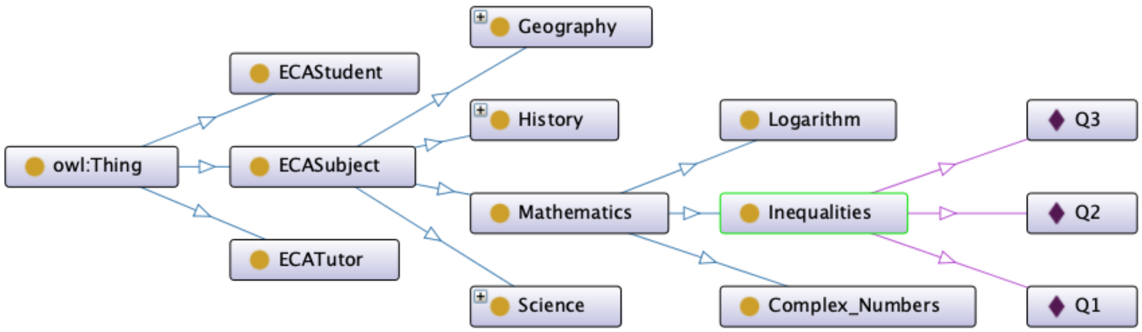

Fig. 5.5.1: Format of a simple Ontology

5.5 Use of Ontology in E-Tutoring

5.5.1 E-Tutoring Without Personalization

"E-Tutoring Without Personalization" is a critical aspect that warrants thorough consideration in educational technology. The system in question relies on an ontology to efficiently categorize educational content into distinct subjects and subtopics, contributing to a well-structured library of learning materials. However, a notable limitation arises from the system's need for adaptation to users' knowledge levels.

The inherent drawback manifests as an absence of personalized learning experiences, as the ontology-based approach must account for users' diverse proficiency levels. This oversight results in learners encountering material that may surpass their current understanding or prove too introductory for their skill level. The consequence is a potential disengagement from the learning process, as users may find the content either overly challenging or insufficiently stimulating.

The significance of personalization becomes evident in addressing these limitations. With tailored adaptation to users' knowledge levels, the e-tutoring platform can avoid causing frustration among learners who grapple with overly complex material or, conversely, disinterest when confronted with content perceived as too elementary. The absence of personalization also hinders the platform from delivering tailored learning experiences that align with each learner's specific needs and goals. This example underscores the imperative of incorporating personalization into e-tutoring platforms to enhance adaptability, engagement, and overall effectiveness in delivering

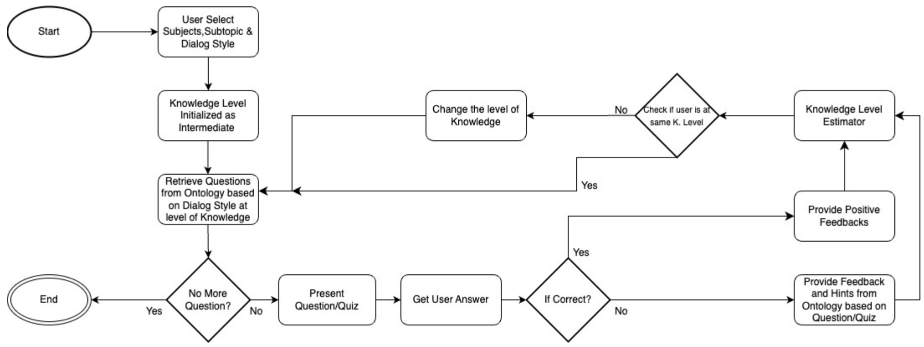

Fig. 5.5.2: Algorithm Flow Chart

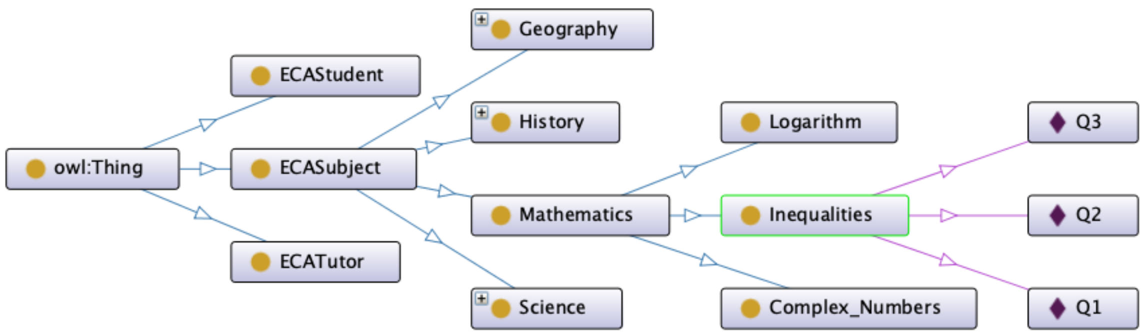

Fig. 5.5.3: Basic Ontology

a more meaningful and tailored learning experience. The analysis emphasizes the pivotal role of personalization in mitigating the challenges associated with rigid, non-adaptive e-tutoring systems, ultimately contributing to a more positive and impactful learning journey for users.

5.5.2 Personalized E-Tutoring with Basic Ontology and Algorithm

Algorithm-driven adaptation is a pivotal component in the personalized e-tutoring system, employing a sophisticated algorithm designed to monitor and adjust the learner's knowledge level dynamically. This rule-based adaptive approach enables continuous evaluation of the learner's performance. When the algorithm identifies a learner struggling with a specific concept or question, it initiates actions to modify the knowledge level accordingly. Notably, the algorithm can decrease the associated knowledge level for the content in response to challenges the learner faces. For ex-

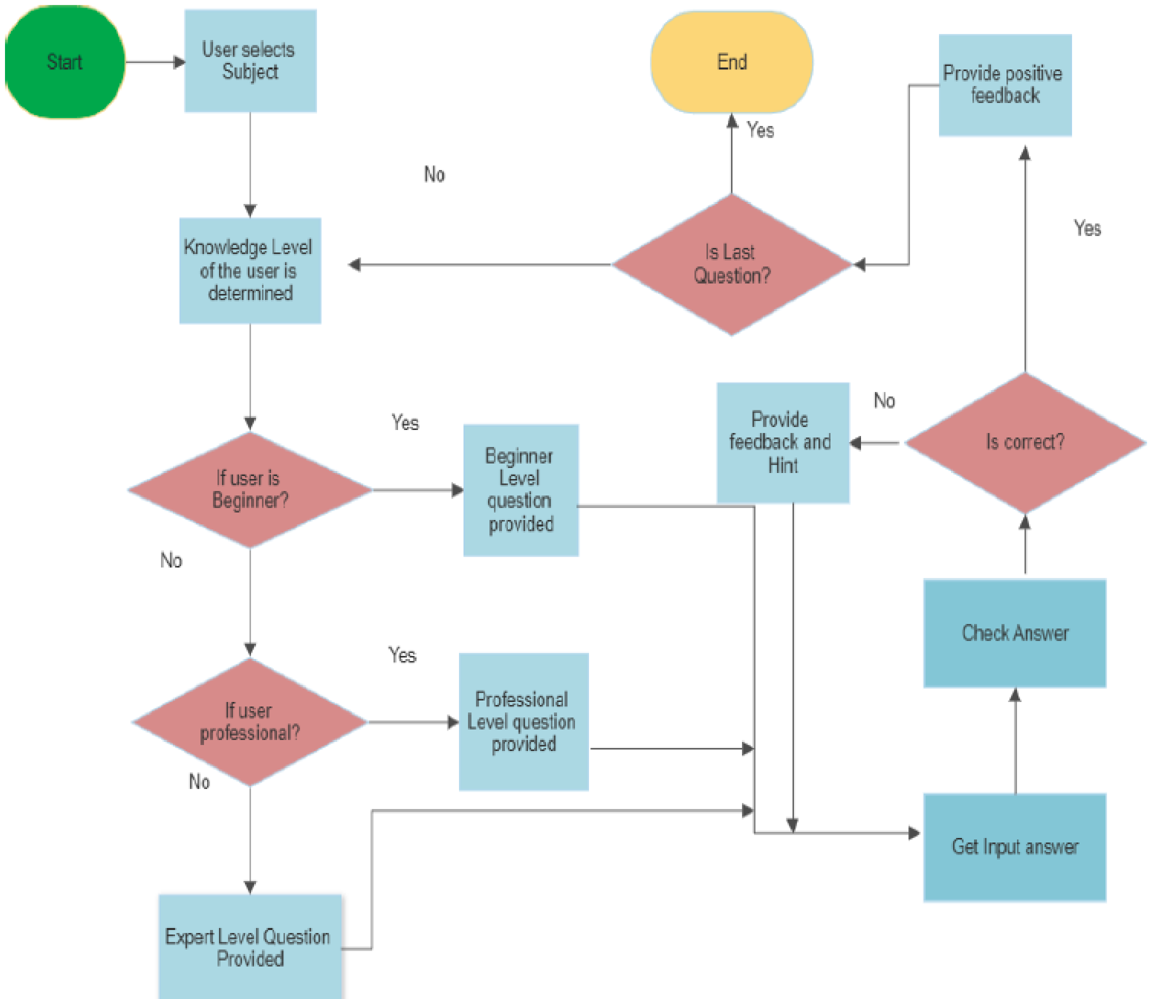

Fig. 5.5.4: Algorithm to utilize the basic Ontology

ample, if a learner encounters difficulty with a challenging question and provides an incorrect answer, the algorithm adjusts the difficulty level downward. This ensures a more accessible learning experience, aligning with the learner's proficiency. In the earlier methodology, question complexity was contingent on the user's proficiency level, accompanied by hints when difficulties arose. Successful user responses triggered affirmative feedback, fostering accelerated knowledge acquisition. The user engagement process involves selecting a subject before proceeding with the task, as the algorithm outlines. Once the knowledge level and current task are determined, the system iterates through each question based on the user's knowledge. Each question is presented individually, and the user enters their response. The ontology then

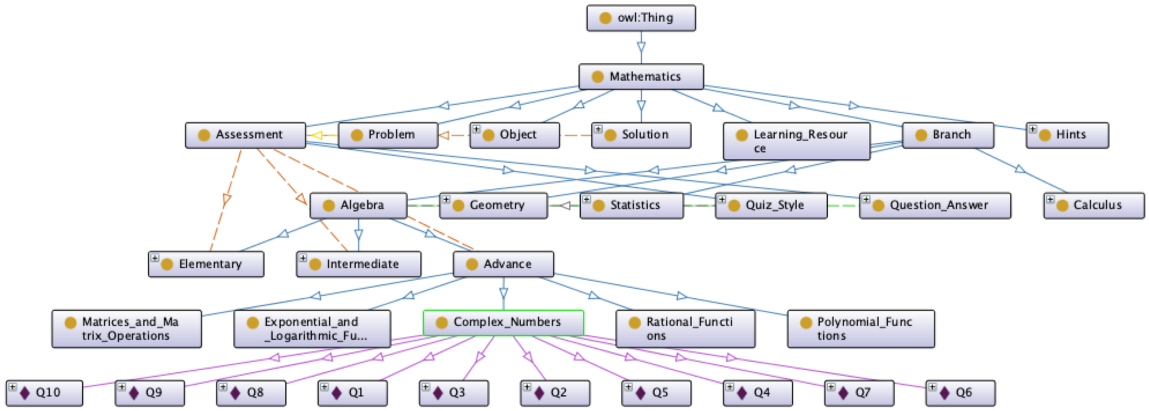

Fig. 5.5.5: Improved Domain Ontology

evaluates the user-provided answer for accuracy, providing constructive feedback for correct responses and negative feedback with motivation and hints for incorrect ones. Figure 5.5.4 visually depicts the entire interaction and conditional actions, showcasing the comprehensive nature of the personalized e-tutoring system's algorithm-driven adaptation.

5.5.3 Improvement with Semantically Enriched Ontology

The developed ontology serves as a sophisticated tool for content categorization, surpassing traditional methods by incorporating expertise and domain knowledge into its classification framework. Our comprehensive review of Ontario's curriculum website involved meticulously categorizing topics into elementary, intermediate, and advanced levels, employing a keyword-matching approach to ensure classification accuracy based on complexity and relevance. The domain ontology, a key component of our system, exhibits a high level of granularity, enabling a nuanced differentiation between fundamental concepts and advanced topics within a given subject. This capability enhances content organization, ensuring a structured and informed presentation to users. One of the distinctive features of our system is its implementation of Adaptive Learning. This functionality optimizes the user's learning journey by aligning with their proficiency levels and objectives. The system intelligently tailors content delivery, striking a balance that is neither easy nor difficult, thereby fostering

Table 5.5.1: Personalized E-Tutoring with semantically enriched ontology

Q. Range	Knowledge Level	Action
1-2	Elementary	Incorrect. {Hint}, Repeat & Next Q}
3-9	Intermediate	Correct. {noHints} & Next Q}
10-13	Elementary	Incorrect. {Hint}, Repeat & Next Q}
14-20	Intermediate	Correct. {noHints} & Next Q}
21-24	Elementary	Incorrect. {Hint}, Repeat & Next Q}
25-35	Intermediate	Correct. {noHints}

optimal user growth and comprehension conditions.

Our semantically enriched ontology and Adaptive Learning form a robust foundation for improved personalized e-tutoring. The ontology's advanced categorization capabilities and the system's adaptive approach collectively contribute to a more refined and tailored educational experience, addressing users' individualized needs and growth trajectories.

5.5.4 Scalability using a different Dataset

The exploration of the scalability of our proposed method involved the application of our approach to a distinct dataset known as HotPotQA. Through this experimentation, our process demonstrated the capability to generate new concepts intricately related to the domain ontology. The hotpotQA dataset, renowned for featuring natural, multi-hop questions [97], was a robust testing ground for our approach. Focusing on a specific dataset section, we successfully extracted and comprehended complex concepts such as "interest" and "loan amount". This insightful endeavour showcased our method's effectiveness in scaling up the ontology across different datasets, emphasizing its adaptability and capacity to produce more intricate domain ontologies for diverse domains. The results underscore the versatility of our approach, illustrating its potential to extend its applicability and contribute to the development of nuanced

Level	Question	Context	Answer	_id	Type
medium	A father gave $500 to his two sons. He gave X dollars to one son. Which of the following expressions correctly shows the amount he gave to the other son.	If the father gave $500 to his two sons and X dollars to one son, the amount he gave to the other son can be found by subtracting X from the total	500- X	5a7a0693 55429901 98eaf050	algebra
medium	Irene is N years old. If Tom is twice as old, which of the following algebraic statements shows his age now?	If Irene is N years old and Tom is twice as old as Irene, we can represent Tom's age as 2N.	T = 2N	5a879ab0 5542996e 4f30887e	algebra
hard	John has B bats and buys C bats more. He then gives away D bats. Which of the following algebraic statements corretly shows how many bats John has left?	To determine how many bats John has left after buying C bats and giving away D bats, we can start with the initial	B + C - D	5a8d7341 55429944 1c6b9fe5	algebra
medium	Joe is lending Frank $500 and Frank has 4 years to pay it back. Joe is charging 6% interest. Frank wants to determine how much interest he'll pay over the 4 years. He sets up	To find the interest (i) that Frank will pay over the 4 years, we can use the formula: i = I * r * t	120$	5a82171f5 542990a1 d231f4a	algebra

Fig. 5.5.6: The HotPotQA dataset

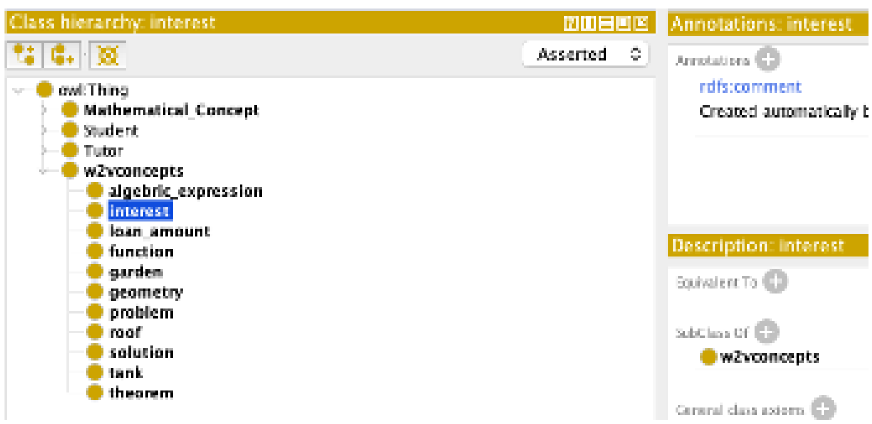

Fig. 5.5.7: Added Classes

and comprehensive domain representations.

CHAPTER 6

Experimental Results

In this chapter, we present the outcomes of the experimental phase, delving into the automatically constructed domain ontology and providing insights into its structure, coverage, and effectiveness. The results are systematically organized to showcase the contributions made, supported by evidence highlighting the impact and significance of the developed ontology. Additionally, we address any identified limitations and engage in discussions to comprehensively understand the experimental findings.

6.1 Automatically Constructed Domain Ontology

Constructing the Automatically Constructed Domain Ontology involves a robust four-step algorithm that ensures the ontology's sustainability and adaptability. This algorithm, characterized by its versatility and conciseness, paves the way for a comprehensive ontology that evolves dynamically based on the input data. The algorithm's innate ability to adapt to varying data characteristics contributes to the ontology's robustness. The previous ontology that was built was a very simple one [43].

Semantic Enrichment with Word2Vec is a crucial step in constructing the ontology, emphasizing the depth and semantic richness of the knowledge representation. Word2Vec, a robust language model, is employed to imbue the ontology with semantically meaningful word embeddings, elevating the expression of concepts to a more nuanced and contextually aware level. Each word or concept in the ontology is transformed into a high-dimensional vector, capturing its semantic relationships with other words. Word2Vec operates on the principle that words appearing in similar contexts will have vectors closer to the vector space. As a result, the enriched ontology reflects the inherent meanings of individual concepts and the intricate web of relationships and contextual associations between them. Incorporating semantically meaningful word embeddings enhances the ontology's ability to capture the subtleties of language and the complex interplay of concepts within the defined domain. This nuanced understanding contributes to a more sophisticated and context-aware representation of knowledge, ensuring that the ontology is comprehensive and reflective of the intricate semantics inherent in the domain.

Using Word2Vec embeddings goes beyond mere concept enrichment, it plays a pivotal role in visualizing semantic relationships within the ontology. By leveraging Word2Vec's ability to capture semantic nuances, the system creates a visual representation that illuminates the intricate connections among existing concepts. This visualization offers a profound understanding of how concepts are interlinked and provides nuanced insights into the semantic associations between these concepts and the extracted tokens. The visual representation is a dynamic and insightful resource, offering a comprehensive view of the ontology's structure and dynamics. It enables stakeholders to discern not only the conceptual hierarchy but also the nuanced relationships that exist between different elements. This visual aid proves invaluable in deciphering the intricacies of the ontology, fostering a deeper comprehension of the semantic landscape and facilitating informed decision-making in the ontology construction process.

Integrating Word2Vec in our ontology construction process significantly enhances scalability and efficiency. Word2Vec's ability to provide contextualized word meanings is pivotal in the system's adaptability to varying datasets. The model's contextual understanding allows it to efficiently identify and propose new concepts for ontology expansion, ensuring the system scales seamlessly across diverse data environments. By leveraging Word2Vec, the system can effectively accommodate more extensive, intricate datasets without compromising performance. The model's understanding of contextual relationships allows it to scale up the ontology by recognizing patterns and semantic nuances in diverse data. On the other hand, efficiency highlights the system's ability to perform these tasks with optimized computational resources. Word2Vec's contextualized word meanings contribute to the efficiency of the ontology construction process. The model efficiently identifies relevant concepts and proposes meaningful expansions, ensuring the system operates resourcefully and minimizes computational overhead.

In assessing the effectiveness of our methodology, a comprehensive and rigorous evaluation process will be employed, relying on quantitative measures to gauge the system's performance. This evaluation encompasses scrutinizing annotated and newly generated classes across diverse datasets. The objective is to meticulously analyze the system's capacity to understand and annotate existing classes and, more critically, generate novel classes pertinent to specific topics. The quantitative measures will include metrics that capture the annotated and newly developed classes' accuracy, precision, recall, and F1 scores. Accuracy measures the overall correctness of predictions, precision focuses on the accuracy of positive predictions, recall gauges the ability to capture all relevant instances, and the F1 score provides a balance between precision and recall [73]. By employing these metrics, we aim to derive a nuanced understanding of the system's proficiency in recognizing existing concepts and proposing innovative ones. We have also used several other precise QA datasets to measure our score. The model consistently demonstrates exceptional performance

Table 6.1.1: Model Quantitative Measure

	Accuracy	**Precision**	**Recall**	**F1 Score**
SQUAD2.0	93.000%	97.778%	94.624%	96.175%
HotPodQA	84.000%	94.595%	85.366%	89.744%
COQA	86.000%	87.654%	94.667%	91.026%
IWT Edit Study (Niyati)	85.000%	91.463%	90.361%	90.909%

across diverse datasets, particularly excelling in the SQUAD2.0 dataset. It achieves high accuracy and precision here, showcasing its proficiency in recognizing and suggesting pertinent concepts. The commendable recall and F1 score further underscore the model's ability to balance precision and sensitivity. Moving to the HotPotQA dataset, the model exhibits slightly lower accuracy than SQUAD2.0. However, it maintains robust precision and a well-balanced F1 score, indicating its reliability in recognizing and handling relevant concepts. Despite a marginal dip in the recall, the model remains proficient in capturing pertinent information within the HotPotQA context.

Consistency is a hallmark of the model's performance across various datasets. Its ability to sustain a favourable equilibrium between precision and recall is noteworthy. In the COQA dataset, the model attains high accuracy and memory, showcasing its versatility in adapting to diverse question-answer scenarios. The model's reliability in accuracy and precision persists across datasets, emphasizing its pivotal role in generating semantically enriched ontologies. The balanced recall and F1 score further affirm the model's effectiveness in addressing diverse scenarios, thus emphasizing the scalability and versatility inherent in the proposed approach. Furthermore, this evaluation process scrutinizes different datasets deliberately chosen to represent varying topics and domains. This diversity in datasets ensures a robust assessment of the approach's adaptability and effectiveness across various scenarios. The multifaceted evaluation aims to validate the system's performance and provide insights into its robustness and adaptability in real-world, diverse data situations.

6.2 Evidence of Contribution

The textual data from two publicly available datasets undergoes a meticulous preprocessing and extraction process, aligning with the guidelines outlined in Section 4.2.2. This comprehensive procedure results in a dataset comprising 147,125 symbols, with 2,223 noun tokens identified for further integration into the workflows detailed in Section 4.2.3. The exploration of various mincount parameters within the range of mincount=[1...25] reveals varying quantities of tokens. Notably, higher mincount parameters yield fewer tokens, considered more significant as they exhibit greater frequency within the dataset. The ensuing tokens serve as concept names for the subsequent search for suitable classes in the mathematical dictionary curated for annotation.

This systematic approach not only provides the count of tokens already present in the respective ontology classes but also extracts textual definitions for these classes in an automated fashion. Additionally, leveraging the count of classes already incorporated enables the suggestion of the ontology that best aligns with the given text dataset. This evidence underscores the meticulous and data-driven process employed to develop and enrich ontologies, emphasizing the practical application and relevance of the proposed methodologies.

The focus is on presenting tangible results from the annotation process on semantic artifacts. Table 6.2.1 provides a comprehensive breakdown of the identified numbers of classes resulting from the annotation effort across four distinct quantity ranges, spanning from 1 to 100. The annotation approach emphasizes the significance of annotating each token with a textual definition at least once. The cumulative count of annotated tokens is calculated for each set, accounting for situations where a token may receive annotations from multiple semantic artifacts. However, to ensure an accurate representation of the overall contribution, a token is counted only once in the row depicting the sum of annotated tokens.

The critical metric for assessment involves dividing the sum of annotated tokens by the total number of tokens, thereby generating the rate of annotated tokens. A high rate of annotated tokens is a desirable outcome, as it indicates a more comprehensive coverage of definitions for classes. This accomplishment holds strategic importance in minimizing future workloads associated with ontology revision, given that devising definitions for classes is inherently more challenging than reaching a consensus on pre-existing ones. Simultaneously, a high sum of annotated tokens is also advantageous, contributing to increased expressivity in the ontology. Integrating more classes into the ontology enhances its capability to articulate a broader range of concepts, reinforcing its effectiveness and relevance in capturing the intricacies of the represented domain. In exploring different parameter settings, it is observed that sets

Table 6.2.1: Quantity of tokens represented as classes in semantic artifacts

	Min_Count				
Dataset	1	10	20	50	100
SQUAD2.0	522	302	151	97	47
HotPodQA	190	101	65	31	18
COQA(Ashwitha)	107	45	21	9	7
IWT Edit Study (Niyati)	255	145	88	33	25
Total Annotated Tokens Found	1074	593	325	170	97
Total Tokens Found	2223	1097	756	515	322

obtained with lower mincount values contain more tokens than those with higher mincount values. However, an interesting trend emerges as the rate of annotated tokens increases with higher mincount parameters, suggesting a potential higher relevance of tokens in sets with elevated mincount values. The relatively low rate of annotated tokens is particularly noteworthy when the mincount is set to 1, standing at 28.25% compared to other rates. This disparity may be attributed to including typing errors and non-domain-relevant tokens at lower mincount values, as a single occurrence is sufficient for a token to be included in the text dataset.

Moreover, lower mincount parameters encompass more undefined concepts in the ontologies, leading to a more significant number of new class candidates in the respective ontologies. Interestingly, the ontologies exhibit a lower token count than the dictionary created for mathematics tutoring. This observation suggests a meaningful intersection between the topics in the text dataset and the conceptual framework outlined in the QA datasets. This evidence underscores the contribution of parameter settings, especially the impact of ‘mincount,‘ in shaping the relevance and alignment of ontologies with the dataset. Analyzing the rate of annotated tokens about the mincount parameters reveals a notable increase between mincount = 1 and mincount = 2. Considering the number of tokens within each ontology, it is observed that the Squad2.0 dataset consistently contains the highest number of tokens for every mincount. Word2Vec models undergo training on token sets within the specified range of mincount parameters (mincount = [1...25]). Subsequently, class labels from the meta ontology in the token set serve as input to identify the most similar words. The determination of word similarity is based on cosine similarity, allowing the setting of thresholds to limit the output words based on their similarity to the input word. A maximum of five output words per input word is established, and the point is adjusted within the range of [0.8, ..., 0.999]. Due to the potential overlap of words in multiple output sets for distinct input words, the count of unique tokens generated by Word2Vec is calculated by considering each word as a class candidate only once.

The impact of the cosine similarity threshold on the generation of unique tokens for low mincount is notable, particularly within the range of [0.8, ..., 0.995] and mincount greater than 5, effectively mitigating the effects. When employing various mincount values and a cosine similarity threshold of 0.999, the automatic extension by adding new classes is suggested by the Word2Vec model. These new classes are subsequently annotated with textual definitions derived from the classes and concepts in these public-domain datasets. Object properties conceptually related to these new classes are asserted, referencing the ontology classes pre-existing in the meta-ontology. Table 6.2.2 provides the count of new classes inserted into the domain ontology by setting the cosine similarity threshold to 0.999 and applying different mincount parameters ranging from [1, ..., 25]. Notably, a mincount of 10 emerges as the most promising, resulting in the highest number of new and annotated classes. Thus, The Domain Ontology is extended by 96 classes automatically created based on the text dataset, achieving an annotation rate of 74.73%.

The new class flow is introduced into the workflow as a subclass of “w2vConcept”. It is linked through the relation conceptually related to (indicated by dashed orange arrows) to the class “Mathematics”. The resulting annotations for the class “Interest” are illustrated in Figure 5.7.7, where the first entry denotes the class label, and

Table 6.2.2: The number of additional classes introduced into the domain ontology when establishing the cosine similarity threshold at 0.999

	Min_Count				
Dataset	1	10	20	50	100
SQUAD2.0	23	30	17	11	9
HotPodQA	15	14	9	7	1
COQA(Ashwitha)	14	8	5	1	0
IWT Edit Study (Niyati)	19	11	9	5	2
Total Annotated Tokens Found	56	45	33	25	15
Total Tokens Found	78	65	54	48	39

the subsequent entries highlight the word-input leading to the class generation. Automatically generated new classes can have multiple "rdfs: comment" assignments, but only one "rdfs: label" is assigned. This automated ontology extension based on open datasets allows an arbitrary number of comments to be associated with a class. However, an evaluation by domain experts is essential to validate the accuracy of resulting definitions and relations.

The automated ontology extension process can be employed for annotated tutoring data. E-learning platforms uploading their data and accompanying textual documentation to a database can utilize this workflow to select the most suitable ontology and extend it automatically. Subsequently, the extended ontology can annotate previously uploaded tutoring data for any subject, establishing connections between data entries and relations articulated in the textual documentation.

6.3 Limitations and Discussions

The automatic construction of ontologies using publicly available domain datasets presents numerous advantages. However, it is crucial to acknowledge and address certain limitations. One primary limitation is the quality and completeness of the input datasets. Publicly available datasets may vary significantly in accuracy, relevance, and coverage, impacting the resulting ontology's comprehensiveness and correctness. Additionally, the inherent biases present in publicly sourced data may be transferred to the ontology, potentially leading to skewed or incomplete representations of certain concepts. Moreover, the challenge of domain specificity limits the applicability of automatically constructed ontologies. Public datasets may not capture the intricacies and nuances of highly specialized domains, determining the effectiveness of the constructed ontology in such contexts. This limitation necessitates careful consideration of the target domain and potential adjustments to ensure the ontology aligns with specific requirements. The scalability of the automatic ontology construction process is another consideration. While publicly available datasets may suit certain domains, scaling the approach to more extensive or rapidly evolving fields could be challenging. The adaptability of the methodology to diverse domains and its ability to accommodate evolving knowledge landscapes require continuous evaluation and refinement.

Furthermore, it introduces challenges by relying on natural language processing techniques for extracting concepts and relationships. Ambiguities, polysemy, or variations in language use within the datasets may lead to inaccuracies or misinterpretations during the extraction process. Rigorous validation and refinement processes are essential to mitigate these challenges and enhance the overall quality of the constructed ontology.

In discussions surrounding automatic ontology construction, it is imperative to consider the trade-off between automation and human expertise. While automation streamlines the process, the involvement of domain experts remains critical for ensuring the ontology's relevance, accuracy, and alignment with domain-specific requirements. Striking the right balance between automated techniques and human validation is pivotal for the success of the ontology construction process.

In conclusion, the automatic construction of ontologies from publicly available datasets offers a promising avenue for knowledge representation. However, researchers and practitioners must remain vigilant about the inherent limitations, addressing issues related to data quality, domain specificity, scalability, and the interplay between automation and human expertise. Ongoing discussions and advancements in the field will contribute to refining these methodologies and expanding their applicability across diverse domains.

CHAPTER 7

Conclusion and Future Work

7.1 Research Summary

In this research endeavour, an innovative methodology was devised to construct a domain ontology from a meta-ontology, employing advanced techniques like word2vec. The primary objective was to address the challenge of enhancing personalized tutoring systems with domain-specific knowledge, ultimately elevating the quality of tailored learning experiences. The methodology involved successfully extracting domain-specific concepts and relationships from publicly available datasets, leveraging natural language processing techniques. A comprehensive four-step algorithm was employed to create a robust domain ontology, ensuring the incorporation of contextually relevant knowledge. Notably, the integration of word2vec for semantic enrichment played a pivotal role in expanding the ontology's coverage of domain-specific terms, adding depth and nuance to its representation. The research findings emphasize the significant potential of the developed methodology to shape the future landscape of educational technology, offering transformative possibilities for the delivery of personalized tutoring and revolutionizing the way learners engage with educational content.

7.2 Future Works

Future work could encompass several pivotal areas to further enhance the efficacy and scope of personalized tutoring systems.

1. Ontology Refinement and Expansion: Building upon the automated ontology

construction approach outlined in the thesis, future endeavours could focus on continuous refinement and expansion of the ontology. This involves incorporating feedback mechanisms that dynamically allow the system to adapt and improve based on user interactions. Additionally, enriching the ontology with more intricate relationships and nuanced domain-specific knowledge will bolster the capability of personalized tutoring systems to cater to diverse learner needs.

2. Integration of Advanced AI Techniques: Incorporating cutting-edge AI techniques, such as reinforcement learning or advanced natural language processing models, could significantly elevate the capabilities of Embodied Conversational Agents (ECAs) used within tutoring systems. These techniques could enable the agents to provide tailored content and adapt their tutoring strategies in real-time based on the learner's cognitive responses, emotional cues, and learning patterns. Moreover, incorporating sophisticated natural language processing models empowers ECAs to understand and address nuanced linguistic cues, emotional nuances, and learners' cognitive states. This integration enables ECAs to provide personalized, adaptive tutoring experiences that better match individual learning paths and preferences, resulting in a more captivating and efficient educational process.

3. Evaluation and User Studies: Conducting comprehensive user studies and evaluations will be crucial to assess the effectiveness, usability, and user satisfaction of the personalized tutoring system implemented with the automated ontology. Gathering feedback from learners, educators, and stakeholders will provide invaluable insights into the system's strengths, weaknesses, and areas for improvement, guiding iterative refinements.

4. Collaboration and Interoperability: Future work could explore strategies enabling interoperability and collaboration between personalized tutoring systems. Developing standards or protocols for exchanging information and knowledge between diverse systems could facilitate a more cohesive and comprehensive learning experience for learners accessing multiple platforms or educational resources.

5. Ethical Considerations and Bias Mitigation: Addressing ethical considerations, such as ensuring fairness, transparency, and mitigating biases within the personalized

tutoring systems, is crucial. Future research should explore methodologies to identify and rectify biases in the generated ontologies or tutoring content to provide an inclusive and equitable learning environment for all learners.

6. Natural Language Integration: Investigating methods to enhance the interaction and resource efficiency of automated ontology construction processes will be pivotal. This involves exploring techniques like natural language integration to handle larger datasets efficiently and expedite ontology robustness.

By delving into these future directions, researchers can contribute to advancing and refining automated ontology construction methodologies, bolstering the foundation of personalized tutoring systems and fostering a more adaptive, responsive, and enriching learning experience for learners across diverse educational landscapes.

REFERENCES

[1] Abebe, S., Qin, T., Zhang, X., and Yan, D. (2022). Wavelet transform-based trend analysis of streamflow and precipitation in upper blue nile river basin. *Journal of Hydrology: Regional Studies*, 44:101251.

[2] Agárdi, A. and Kovács, L. (2022). Property-based quality measures in ontology modeling. *Applied Sciences*, 12(23):12475.

[3] Ahmed, S. K., Jeffries, D., Chakraborty, A., Carslake, T., Lietz, P., Rahayu, B., Armstrong, D., Kaushik, A., and Sundarsagar, K. (2022). Teacher professional development for disability inclusion in low-and middle-income asia-pacific countries: An evidence and gap map. *Campbell systematic reviews*, 18(4):e1287.

[4] Ahmend, M. U., Begum, S., Funk, P., Xiong, N., and Von Schéele, B. (2008). Case-based reasoning for diagnosis of stress using enhanced cosine and fuzzy similarity. *Transactions on Case-Based Reasoning for Multimedia Data*, 1(1):3–19.

[5] Al-Adwan, A. S., Nofal, M., Akram, H., Albelbisi, N. A., and Al-Okaily, M. (2022). Towards a sustainable adoption of e-learning systems: The role of self-directed learning. *Journal of Information Technology Education: Research*, 21.

[6] Alabri, A., Al-Khanjari, Z., Jamoussi, Y., and Kraiem, N. (2019). Mining the students' chat conversations in a personalized e-learning environment. *International Journal of Emerging Technologies in Learning (iJET)*, 14(23):pp. 98–124.

[7] Albukhitan, S., Helmy, T., and Alnazer, A. (2017). Arabic ontology learning using

deep learning. In *Proceedings of the international conference on web intelligence*, pages 1138–1142.

[8] Alenezi, A. (2020). The role of e-learning materials in enhancing teaching and learning behaviors. *International Journal of Information and Education Technology*, 10(1):48–56.

[9] Allouch, M., Azaria, A., and Azoulay, R. (2021). Conversational agents: Goals, technologies, vision and challenges. *Sensors*, 21(24):8448.

[10] Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The journal of the learning sciences*, 4(2):167–207.

[11] Bengag, A., Bengag, A., Moussaoui, O., and Blej, M. (2023). *A Fuzzy Logic-Based Intrusion Detection System for WBAN Against Jamming Attacks*, pages 3–11.

[12] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009). Dbpedia-a crystallization point for the web of data. *Journal of web semantics*, 7(3):154–165.

[13] Browarnik, A. and Maimon, O. (2015). Ontology learning from text: why the ontology learning layer cake is not viable. *International Journal of Signs and Semiotic Systems (IJSSS)*, 4(2):1–14.

[14] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

[15] Buitelaar, P., Cimiano, P., and Magnini, B. (2005). *Ontology learning from text: methods, evaluation and applications*, volume 123. IOS press.

[16] Capuano, N. and Toti, D. (2019). Experimentation of a smart learning system for law based on knowledge discovery and cognitive computing. *Computers in Human Behavior*, 92:459–467.

[17] Cassandra, A. (2003). The pomdp page. *POMDP. org,[Online]. Available: http://pomdp. org/.[Accessed 21 2 2020].*

[18] Cela, K. L., Sicilia, M. Á., and Sánchez, S. (2015). Social network analysis in e-learning environments: A preliminary systematic review. *Educational Psychology Review*, 27:219–246.

[19] Chimalakonda, S. and Nori, K. V. (2020). An ontology based modeling framework for design of educational technologies. *Smart learning environments*, 7(1):1–24.

[20] Cimiano, P. and Völker, J. (2005). Text2onto: A framework for ontology learning and data-driven change discovery. In *International conference on application of natural language to information systems*, pages 227–238. Springer.

[21] Dascalu, M.-I., Bodea, C.-N., Moldoveanu, A., Mohora, A., Lytras, M., and de Pablos, P. O. (2015). A recommender agent based on learning styles for better virtual collaborative learning experiences. *Computers in Human Behavior*, 45:243–253.

[22] Davis, J. and Domingos, P. (2011). Deep transfer: A markov logic approach. *AI Magazine*, 32(1):51–51.

[23] Downes, S. (2007). E-learning 2.0//elearn magazine. 2005. *Retrieved June*, 25:29–1.

[24] Drissi, A., Khemiri, A., Sassi, S., and Chbeir, R. (2021). A new automatic ontology construction method based on machine learning techniques: Application on financial corpus. In *Proceedings of the 13th International Conference on Management of Digital EcoSystems*, MEDES '21, page 57–61, New York, NY, USA. Association for Computing Machinery.

[25] Fenza, G., Orciuoli, F., and Sampson, D. G. (2017). Building adaptive tutoring model using artificial neural networks and reinforcement learning. In *2017 IEEE*

17th international conference on advanced learning technologies (ICALT), pages 460–462. IEEE.

[26] Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M., and Soria, C. (2006). Lexical markup framework (lmf). In *International Conference on Language Resources and Evaluation-LREC 2006*.

[27] Gašić, M. and Young, S. (2013). Gaussian processes for pomdp-based dialogue manager optimization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(1):28–40.

[28] George, G. and Lal, A. M. (2019). Review of ontology-based recommender systems in e-learning. *Computers & Education*, 142:103642.

[29] Glass, J., Polifroni, J., Seneff, S., and Zue, V. (2000). Data collection and performance evaluation of spoken dialogue systems: The mit experience.

[30] Gómez-Pérez, A., Manzano-Macho, D., et al. (2003). A survey of ontology learning methods and techniques. *OntoWeb Deliverable D*, 1(5).

[31] Graesser, A. C. (2016). Conversations with autotutor help students learn. *International Journal of Artificial Intelligence in Education*, 26:124–132.

[32] Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., and Louwerse, M. M. (2004). Autotutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36:180–192.

[33] Haridy, S., Ismail, R. M., Badr, N., and Hashem, M. (2023). An ontology development methodology based on ontology-driven conceptual modeling and natural language processing: Tourism case study. *Big Data and Cognitive Computing*, 7(2):101.

[34] Harms, J.-G., Kucherbaev, P., Bozzon, A., and Houben, G.-J. (2018a). Approaches for dialog management in conversational agents. *IEEE Internet Computing*, 23(2):13–22.

[35] Harms, J.-G., Kucherbaev, P., Bozzon, A., and Houben, G.-J. (2018b). Approaches for dialog management in conversational agents. *IEEE Internet Computing*, PP:1–1.

[36] Hashim, H. and Tasir, Z. (2014). E-learning readiness: A literature review. In *2014 international conference on teaching and learning in computing and engineering*, pages 267–271. IEEE.

[37] Heisler, M. and Becker-Asano, C. (2023). An android robot head as embodied conversational agent.

[38] Hellmann, S., Lehmann, J., and Auer, S. (2012). Linked-data aware uri schemes for referencing text fragments. In *Knowledge Engineering and Knowledge Management: 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012. Proceedings 18*, pages 175–184. Springer.

[39] Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.

[40] Huang, R. (2019). *Educational technology a primer for the 21st century*. Springer.

[41] Hwang, Y.-Y., Son, J., and Shin, S. (2022). Design of ontology model for knowledge navigator. In *2022 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 161–166. IEEE.

[42] Janarthanam, S. (2017). *Hands-on chatbots and conversational UI development: build chatbots and voice user interfaces with Chatfuel, Dialogflow, Microsoft Bot Framework, Twilio, and Alexa Skills*. Packt Publishing Ltd.

[43] Jawahar, A. V. (2023). *Personalized ECA Tutoring with Self-Adjusted POMDP Policies and User Clustering*. PhD thesis, University of Windsor (Canada).

[44] Kamp, H. (1981). A theory of truth and semantic representation. *Formal semantics-the essential readings*, pages 189–222.

[45] Kay, J., Bartimote, K., Kitto, K., Kummerfeld, B., Liu, D., and Reimann, P. (2022). Enhancing learning by open learner model (olm) driven data design. *Computers and Education: Artificial Intelligence*, 3:100069.

[46] Khan, M. M. (2023). *An Approach of Lip Synchronization With Facial Expression Rendering for an ECA*. PhD thesis, University of Windsor (Canada).

[47] Kim, S., Ahn, K., and Kim, S. (2019). A method of educational quality administration based on hyper meta ontology. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, pages 237–242. IEEE.

[48] Li, H., Lv, Z., Li, J., Xu, Z., Wang, Y., Sun, H., and Sheng, Z. (2023). Traffic flow forecasting in the covid-19: A deep spatial-temporal model based on discrete wavelet transformation. *ACM Trans. Knowl. Discov. Data*, 17(5).

[49] Liu, J., Zhang, X., Li, Y., Wang, J., and Kim, H.-J. (2019). Deep learning-based reasoning with multi-ontology for iot applications. *IEEE Access*, 7:124688–124701.

[50] Macina, J., Daheim, N., Chowdhury, S. P., Sinha, T., Kapur, M., Gurevych, I., and Sachan, M. (2023). Mathdial: A dialogue tutoring dataset with rich pedagogical properties grounded in math reasoning problems. *arXiv preprint arXiv:2305.14536*.

[51] Maedche, A. and Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent systems*, 16(2):72–79.

[52] Mary Harin Fernandez, F., Venkata Ramana, T., Shabana, M., Kannagi, V., and Nalini, M. (2023). Personalized ontology and deep training tree-based optimal gated recurrent unit-recurrent neural network for prediction of students' behavior. *Concurrency and Computation: Practice and Experience*, 35(1):e7420.

[53] McCrae, J., Spohr, D., and Cimiano, P. (2011). Linking lexical resources and ontologies on the semantic web with lemon. In *The Semantic Web: Research*

and Applications: 8th Extended Semantic Web Conference, ESWC 2011, Heraklion, Crete, Greece, May 29-June 2, 2011, Proceedings, Part I 8, pages 245–259. Springer.

[54] Meziane, F. and Rezgui, Y. (2004). A document management methodology based on similarity contents. *Information Sciences*, 158:15–36.

[55] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[56] Mishra, S. and Jain, S. (2015). A study of various approaches and tools on ontology. In *2015 IEEE International Conference on Computational Intelligence & Communication Technology*, pages 57–61. IEEE.

[57] Monachesi, P., Simov, K., Mossel, E., Osenova, P., and Lemnitzer, L. (2008). What ontologies can do for elearning. *Proceedings of IMCL 2008*.

[58] Morris, T. H. and Rohs, M. (2021). Digitization bolstering self-directed learning for information literate adults–a systematic review. *Computers and Education Open*, 2:100048.

[59] Mulpuri, V. K. (2016). *Trend Analysis of Belief-State History with Discrete Wavelet Transform for Improved Intention Discovery*. PhD thesis, University of Windsor (Canada).

[60] Murad, D. F., Toha, M., Mayatopani, H., Wijanarko, B. D., Heryadi, Y., Dewi, M. A., and Leandros, R. (2023). Personalized recommendation system for online learning: An opportunity. In *2023 8th International Conference on Business and Industrial Research (ICBIR)*, pages 128–132. IEEE.

[61] Mustafa, G., Abbas, M. A., Hafeez, Y., Khan, S., and Hwang, G.-J. (2019). Effectiveness of ontology-based learning content generation for preschool cognitive skills learning. *Interactive Learning Environments*, 27(4):443–457.

[62] Nandwani, P. and Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11.

[63] Nguyen, T. H., Nguyen, H. D., Pham, V. T., Tran, D. A., and Selamat, A. (2022). Legal-onto: An ontology-based model for representing the knowledge of a legal document. In *ENASE*, pages 426–434.

[64] Nye, B. D., Graesser, A. C., and Hu, X. (2014). Autotutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24:427–469.

[65] Omran, M. G., Engelbrecht, A. P., and Salman, A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, 11(6):583–605.

[66] Osman, M. A., Mohd Noah, S. A., and Saad, S. (2022). Ontology-based knowledge management tools for knowledge sharing in organization—a review. *IEEE Access*, 10:43267–43283.

[67] Pazienza, M. T. (2012). *Semi-Automatic Ontology Development: Processes and Resources: Processes and Resources*. IGI Global.

[68] Pei, P., Xuejing, D., and Deqing, Z. (2020). Construction of curriculum knowledge map based on ontology. In *Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence*, pages 259–265.

[69] Peters, W., Montiel Ponsoda, E., Aguado de Cea, G., and Gómez-Pérez, A. (2007). Localizing ontologies in owl.

[70] Pezzini, M. C. and Thomas, P. (2022). A review of intelligent tutoring in secondary education. In *Argentine Congress of Computer Science*, pages 33–46. Springer.

[71] Qushem, U. B., Christopoulos, A., and Laakso, M.-J. (2022). The value proposition of an integrated multimodal learning analytics framework. In *2022 45th*

Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), pages 666–671. IEEE.

[72] Raccanello, D., Balbontín-Alvarado, R., da Silva Bezerra, D., Burro, R., Cheraghi, M., Dobrowolska, B., Fagbamigbe, A. F., Faris, M. E., França, T., González-Fernández, B., et al. (2022). Higher education students' achievement emotions and their antecedents in e-learning amid covid-19 pandemic: A multi-country survey. *Learning and Instruction*, 80:101629.

[73] Rahayu, N. W., Arifin, A., Ferdiana, R., and Kusumawardani, S. S. (2022a). An ontology model for formative assessments. In *2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pages 130–135. IEEE.

[74] Rahayu, N. W., Ferdiana, R., and Kusumawardani, S. S. (2022b). A systematic review of ontology use in e-learning recommender system. *Computers and Education: Artificial Intelligence*, 3:100047.

[75] Rajaonarivo, L., Mine, T., and Arakawa, Y. (2022). Coupling of semantic and syntactic graphs generated via tweets to detect local events. In *2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 128–133. IEEE.

[76] Raval, R. R. (2019). *An Improved Approach of Intention Discovery with Machine Learning for POMDP-based Dialogue Management*. PhD thesis, University of Windsor (Canada).

[77] Ruan, S., Jiang, L., Xu, J., Tham, B. J.-K., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E., and Landay, J. A. (2019). Quizbot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

[78] Sandars, J. (2021). Cost-effective e-learning in medical education. In *Cost effectiveness in medical education*, pages 40–47. CRC Press.

[79] Serrano-Guerrero, J., Romero, F., and Olivas, J. (2021). Fuzzy logic applied to opinion mining: A review. *Knowledge-Based Systems*, 222:107018.

[80] Shute, V. J., Sun, C., and Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational research review*, 22:142–158.

[81] Singh, S., Kearns, M., Litman, D. J., Walker, M. A., et al. (2000). Empirical evaluation of a reinforcement learning spoken dialogue system. In *AAAI/IAAI*, pages 645–651.

[82] Smith, A. (2022). Simple knowledge organization system (skos). *KO KNOWLEDGE ORGANIZATION*, 49(5):371–384.

[83] Soltanifar, M., Sharafi, H., Lotfi, F., Pedrycz, W., and Allahviranloo, T. (2023). *Introduction to Fuzzy Logic*, pages 31–45.

[84] Sonkar, S., Le, M., Chen, X., Liu, N., Mallick, D. B., and Baraniuk, R. G. (2023). Code soliloquies for accurate calculations in large language models. *arXiv preprint arXiv:2309.12161*.

[85] Stasaski, K., Kao, K., and Hearst, M. A. (2020). CIMA: A large open access dialogue dataset for tutoring. In Burstein, J., Kochmar, E., Leacock, C., Madnani, N., Pilán, I., Yannakoudakis, H., and Zesch, T., editors, *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 52–64, Seattle, WA, USA → Online. Association for Computational Linguistics.

[86] Sun, H., Zheng, Q., Zhang, H., and Pan, H. (2022). Learner modeling framework based on learning analytics. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, pages 64–68. IEEE.

[87] Szucs, T. (2019). *Lip Synchronization for ECA Rendering with Self-Adjusted POMDP Policies*. PhD thesis, University of Windsor (Canada).

[88] Tank, M. (2018). Multi-modal methods: Visual speech recognition (lip reading).

[89] Teixeira, M. S., Maran, V., and Dragoni, M. (2021). The interplay of a conversational ontology and ai planning for health dialogue management. In *Proceedings of the 36th annual ACM symposium on applied computing*, pages 611–619.

[90] THOMAS, J. (2023). Evaluation of personalized learning.

[91] Tomek Strzalkowski, S. M. H. (2008). *Advances in Open Domain Question Answering*. USA.

[92] VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., and Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive science*, 31(1):3–62.

[93] Vesin, B., Ivanović, M., Klašnja-Milićević, A., and Budimac, Z. (2012). Protus 2.0: Ontology-based semantic recommendation in programming tutoring system. *Expert Systems with Applications*, 39(15):12229–12246.

[94] Vyas, N. (2022). *An Approach of Using Embodied Conversational Agent for Personalized Tutoring*. PhD thesis, University of Windsor (Canada).

[95] Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3):33–35.

[96] Wong, W., Liu, W., and Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM computing surveys (CSUR)*, 44(4):1–36.

[97] Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W. W., Salakhutdinov, R., and Manning, C. D. (2018). Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *arXiv preprint arXiv:1809.09600*.

[98] Yuan, X. and Bian, L. (2010). A modified approach of pomdp-based dialogue management. In *2010 IEEE International Conference on Robotics and Biomimetics*, pages 816–821. IEEE.

[99] Yuan, X., Kaler, M., and Mulpuri, V. (2017). Personalized visualization based upon wavelet transform for interactive software customization. In *Machine Learning and Data Mining in Pattern Recognition: 13th International Conference, MLDM 2017, New York, NY, USA, July 15-20, 2017, Proceedings 13*, pages 361–375. Springer.

[100] Zakrzewska, D. (2008). Cluster analysis for users' modeling in intelligent e-learning systems. In *New Frontiers in Applied Artificial Intelligence: 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2008 Wrocław, Poland, June 18-20, 2008 Proceedings 21*, pages 209–214. Springer.

[101] Zakrzewska, D. (2009). Cluster analysis in personalized e-learning systems. In *Intelligent systems for knowledge management*, pages 229–250. Springer.

[102] Zhang, J. and Rasmussen, E. M. (2001). Developing a new similarity measure from two different perspectives. *Information Processing & Management*, 37(2):279–294.

[103] Zhang, L., Basham, J. D., and Yang, S. (2020). Understanding the implementation of personalized learning: A research synthesis. *Educational Research Review*, 31:100339.

[104] Zhou, L. (2007). Ontology learning: state of the art and open issues. *Information Technology and Management*, 8:241–252.

[105] Zhu, Q., Geishauser, C., Lin, H.-c., van Niekerk, C., Peng, B., Zhang, Z., Heck, M., Lubis, N., Wan, D., Zhu, X., et al. (2022). Convlab-3: A flexible dialogue system toolkit based on a unified data format. *arXiv preprint arXiv:2211.17148*.

VITA AUCTORIS

NAME:	Asim Jamal
PLACE OF BIRTH:	Varanasi, India
YEAR OF BIRTH:	1997
EDUCATION:	Jamia Hamdard, B.Tech in Computer Science Engineering, New Delhi, India, 2019
	University of Windsor, M.Sc in Computer Science, Windsor, Ontario, 2024