



A Comprehensive Survey on Automated Concept Extraction and Prerequisite Dependency Detection in Educational Texts

Aditi Patil

aditi030604@gmail.com

COEP Technological University,
Maharashtra

Nidhi Chaudhari

nidhichaudhari2505@gmail.com

COEP Technological University,
Maharashtra

Riya Hankare

hankareriya@gmail.com

COEP Technological University,
Maharashtra

Siddharth K. Gaikwad

SkG.comp@coeptech.ac.in

COEP Technological University,
Maharashtra

Shreya Araganji

araganjishreya@gmail.com

COEP Technological University,
Maharashtra

ABSTRACT

Automated extraction of concepts and detection of dependencies are important for analyzing educational texts and supporting applications like curriculum development, learning path suggestions, and intelligent tutoring systems. The aim of these tasks is to automatically extract key domain concepts from educational materials and find dependencies or prerequisites between them. Several approaches are put forward in recent studies, such as weakly supervised methods based on semantic embeddings and clustering, distant supervision techniques that make use of domain-specific glossaries, and supervised transformer-based models. Supervised models frequently exhibit high accuracy, but they rely significantly on large, manually labeled datasets. On the other hand, weakly and distantly supervised approaches drastically lower annotation costs, but they also face issues with seed quality, dictionary coverage, and cross-domain applicability. In order to guide future work in automated educational concept analysis, this study offers a thorough overview of current methods, frequently used datasets, assessment metrics, and recognized limits. It also addresses important issues and unresolved research gaps.

Keywords: Concept Extraction, Prerequisite Learning, Educational Data Mining, Knowledge Graphs, Natural Language Processing, Large Language Models.

1. INTRODUCTION

Educational resources that contain the structured topic information necessary for efficient teaching and learning include textbooks, lecture notes, syllabi, and online learning resources [4]. With the proliferation of digital educational resources, there is an increasing need for automated methods that can effectively evaluate and arrange this content. Two essential tasks in this context are concept extraction, which focuses on extracting significant domain concepts from text [1], and dependency identification, which seeks to uncover prerequisite or dependence links among these concepts [10]. While dependency detection helps in understanding the logical learning sequence among these topics, automated concept extraction assists in discovering the fundamental topics that make up a subject. Collectively, these activities facilitate various educational uses, such as curriculum development, adaptive learning technologies, learning pathway suggestions, and the creation of educational knowledge graphs [9], [11]. Identifying concepts and their relationships manually is labor intensive, subjective, and challenging to scale for extensive and varied educational datasets.

Recent studies have utilized Natural Language Processing and machine learning methods to tackle these issues. Supervised methods, especially transformer-based models, demonstrate great effectiveness in extracting concepts from educational materials but they depend on large, annotated datasets [5], [6]. Weakly supervised and distant supervision techniques that use domain glossaries, seed concepts, or external knowledge sources have been investigated in order to minimize annotation work [2], [3]. Although these methods increase scalability, they frequently face drawbacks like domain dependence, inadequate coverage, and decreased generalization. Automated concept extraction and dependency detection are still difficult despite advancements in this field because of differences in terminology among subjects, the intricate nature of educational language, and the absence of uniform datasets [8]. The aim of this survey is to study the current approaches, find their advantages and disadvantages, and address the research gaps. The report aims to provide an organized summary of existing methods.

2. BACKGROUND AND PROBLEM DEFINITION

The need for automated methods to arrange and comprehend this knowledge is growing due to the expansion of digital educational resources. These platforms provide a large amount of useful data, but this data is largely unstructured, making manual analysis and categorization difficult at scale [4].

In this context, automated concept extraction and dependency detection help turn unstructured educational content into organized knowledge. This simplifies curriculum design and supports intelligent tutoring systems and personalized learning experiences [10].

Concept extraction involves identifying the key terms or phrases that represent important learning units within a subject [1]. Dependency detection, on the other hand, concentrates on discovering the connections between these ideas, particularly the prerequisite links that specify the logical sequence in which they should be taught [7]. When combined, these methods allow for practical applications such as adaptive course sequencing, knowledge graph construction, and learning path recommendation [11]. Concepts may be implicit, abstract, or stated in a variety of ways, and educational language can range significantly between disciplines. The connections between concepts are difficult to identify with basic text analysis since they are frequently instructional rather than explicitly stated [9]. Large, annotated datasets are necessary for supervised learning algorithms to produce strong results [8]. Weakly or distantly supervised techniques can minimize the requirement for significant labeling even though they come at the cost of accuracy and domain flexibility [3]. These challenges illustrate the need for reliable, scalable, and adaptable methods that are capable of handling a variety of educational content.

3. EXISTING METHODOLOGIES

Many research methods have been used in the large study of automated idea extraction and dependence detection in educational texts. Based on supervised transformer-based methods, independent monitoring techniques, weakly supervised and embedding-based solutions, unsupervised many-criteria methods, blended iterative learning frameworks, graph neural network-based models, interpretable learner models, and large language model (LLM) based systems are general groups into which existing approaches can be divided based on the surveyed literature. This section provides a structured review of different procedures, focusing their main concepts, advantages, and disadvantages.

3.1 Supervised Transformer-Based Methods

One of the best methods for educational idea collection is supervised transformer-based methods. These methods usually use annotated educational texts to improve pre-trained transformer models, like BERT, and present idea removal as an order labeling or text classification problem [1]. An expert system called YodKw was introduced by Rico and Espada [1]. It uses a specific educational dataset made up of biology, physics, and history textbooks to improve a BERT-based token classification model. The system is tested against neural models like BiLSTM-CRF and BERT-KPE, as well as traditional statistical fundamentals like TF-IDF, RAKE, and TextRank. According to experimental results, transformer finetuning beats classical methods in terms of precision and F1 score and increases situational understanding of instructional content. But performance varies by discipline, with story-driven fields like history showing more difficulties [1]. Supervised transformer-based methods depend heavily on big, high-quality annotated datasets, despite their outstanding performance. These methods' scalability and usefulness are limited by the time-consuming and expensive nature of creating such datasets for a variety of disciplines and educational levels [1].

3.2 Distant Supervision with Discipline-Aware Refinement

Distant supervision techniques use external resources like dictionaries or domain knowledge to automatically produce noisy labels with the goal to minimize the effort required for human annotation. A three-step pathway for course concept retrieval from MOOC data is put forward by the DS-MOCE framework [2]. In order to improve specific domain usage, the first step provides discipline-aware dictionary empowerment through interactive methods. By deleting incorrect or confusing annotations, the second stage concentrates on fixing noisy labels created via distant supervision. To increase the extraction quality, discipline-aware linking models and self-training using positive-unlabeled learning are used in the last phase [2]. DS-MOCE shows significant increases in F1-score over earlier poorly supervised baselines when tested on a largescale MOOC sample from XuetangX [2]. Although this method reduces manual labeling and successfully scales to huge datasets, it still depends on exact discipline classification and dictionary usage. Further validation is also required for its success outside of MOOC-style topics or non-English context.

3.3 Weakly Supervised Embedding and Clustering Approaches

Poorly directed methods depend on little guidance, such as short initial lists, in an effort to find a balance between retrieval precision and annotation efficiency. GlossEx is a typical method for retrieving specific field language from textbooks using contextualized embeddings [3]. Candidate terms are extracted, relevant embeddings are created using BERT, these embeddings are clustered, and a small selection of initial terms from external sources such as Wikipedia are used to identify the clusters. After that, the grouped language is examined to identify domain-related concepts [3]. GlossEx obtains good agreement with expert opinions and performs well on German economics textbooks. Reduced annotation time, clarity through clustering, and suitability for textbook-style content are among its advantages. Yet, the method's scope is limited because it has only been tested on a small domain and language and is vulnerable to the quality of initial seeds [3].

3.4 Unsupervised and Multi-Criteria Feature-Based Methods

By using mathematical, structural, and semantic signals found in educational resources, unsupervised techniques completely eliminate the requirement for labeled data. The UPreG framework mixes reference-based algorithms, relatedness values, and subject modeling to create initial concept graphs [7]. Latent Dirichlet Allocation (LDA) is used to extract concepts from course descriptions. A balanced combination of relative mutual information from Stack Overflow data, cosine similarity from Word2Vec embeddings, and asymmetric reference scores calculated from Wikipedia articles are used to calculate prerequisite connections. Learning maps are created by scanning the resulting concept graphs [7]. UPreG generates excellent cooperation with subject matter experts and good concept matching with platform-issued tags. However, potential bias is introduced by concept name confusion and dependence on external knowledge sources like Wikipedia. Also, even with experts, required links may be selective [7]. Another unsupervised method employs a complex voting algorithm that integrates document-based, graph-based, tree based, and text-based elements. By choosing only the most trusted prerequisite links, this method improves precision at the cost of low recall, limiting the completeness of the resulting dependency graph [10].

3.5 Hybrid and Iterative Learning Frameworks

To make the most of the benefits of both paradigms, hybrid frameworks combine supervised and unsupervised learning. A well-known example is the DsCE + iPRL system, which integrates iterative prerequisite relation learning with high quality phrase searching [11]. AutoPhrase is used to retrieve domain-specific concepts at the beginning of the process, followed by graph-based ranking to identify important concepts. An unsupervised recovery based model that exploits relationships between learning materials is paired with a supervised classifier to iteratively perform prerequisite learning. Joint evaluation reduces the mismatch between the two models [11].

This framework reduces dependence on large, labeled datasets and performs well on both textbook and MOOC datasets. However, the recovery process requires solving computationally expensive optimization problems, and phrase searching may introduce noisy candidates.

3.6 Graph Neural Network-Based Dependency Modeling

Graph neural networks (GNNs) have been used in recent research to directly model prerequisite relationships as directed graphs. A widely used approach employs the Weisfeiler–Leman (WL) graph isomorphism test to guide a permutation-equivariant directed GNN [8]. Textual explanations encoded using BERT are used to initialize concept nodes. A directed GNN aggregates information from neighboring nodes while preserving permutation consistency. Prerequisite relations between concept pairs are then predicted using a Siamese network [8]. Experiments on datasets such as LectureBank and MOOC corpora demonstrate strong F1-scores. However, the method depends on the quality and availability of textual explanations and requires careful parameter tuning [8].

Another advanced model, GKROM, formulates prerequisite learning as a global multi-objective optimization problem using graph attention networks to jointly optimize concept–concept, document–concept, and document–document interactions. While performance is significantly improved, training time and computational complexity increase [11].

3.7 Interpretable Learner Models for Dependency Discovery

Learner-focused methods infer prerequisite relations from student interaction data rather than textual content. The E-PRISM framework models causal prerequisite links between knowledge components using an interpretable Dynamic Bayesian Network [9]. By analyzing learning and forgetting behaviors, prior knowledge, and concept difficulty, E-PRISM produces human interpretable dependency structures that align well with expert knowledge. However, scalability remains a limitation, as the framework typically focuses on pairwise concept relationships [9].

3.8 Large Language Model Based Approaches

Large language models are increasingly used for both concept extraction and prerequisite detection. Prompt-based systems such as Cognitext and GPT-based pipelines generate concept maps directly from educational content [4]–[6]. LLMs demonstrate a strong ability to uncover implicit and nuanced concepts often overlooked by traditional NLP methods. Human evaluations indicate high accuracy and improved interpretability. However, challenges such as hallucinations, prompt sensitivity, reproducibility issues, and high computational cost affect the reliability of LLM-based approaches [4]–[6].

4. COMPARISON AND PERFORMANCE EVALUATION

The table provides a comprehensive comparative evaluation of representative approaches for automated concept extraction and dependency detection in educational texts. The comparison encompasses a wide range of methodological paradigms. Rule-based expert systems [1], supervised transformer-based models, distantly supervised frameworks that take advantage of curriculum glossaries [2], and weakly or semi-supervised techniques relying on embeddings, bootstrapping, and statistical cues [3] are just a few of the many methodological paradigms that are compared. Each approach is analyzed in terms of the datasets used, the evaluation metrics used, and the practical advantages and limitations reported in existing studies. Overall, the comparison shows distinct tradeoffs between extraction accuracy, annotation effort, scalability, and interpretability. While transformer-based and graph neural network models [8], [11] achieve strong performance in capturing contextual and complex relationships, they require large, annotated datasets and considerable computational resources. Glossary-driven, heuristic [10], and weakly supervised approaches, on the other hand, [7] offer improved scalability and lower annotation cost but often struggle with implicit concepts and dependency inference. By jointly analysing concept extraction and dependency detection techniques [11], the analysis reveals the strong interdependence between these tasks and highlights persistent challenges, particularly the lack of standardized evaluation benchmarks and robust modeling of implicit prerequisite relations, thereby motivating future research in integrated and scalable educational text analysis.

Table-1: Comparison of Concept Extraction and Dependency Detection Approaches in Educational Texts

Approach	Reference	Method	Dataset	Metrics	Strengths	Limitations
Transformer Supervised	Expert system for extracting keywords in educational texts and textbooks (YodKw) (Rico-Juan et al., 2025)	Fine-tuned BERT-base with sequence labeling head	EduSet (363 Spanish units)	$F_1 \approx 0.21$ (Strict) $F_1 \approx 0.65$ (Relaxed)	High precision on formal terminology Domain-adapted training	Requires token level manual annotation Limited zero-shot transfer
Distant Supervision	Distantly Supervised Course Concept Extraction in MOOCs with Academic Discipline (Lu et al., 2025)	Discipline-aware PU Learning + Knowledge-base grounding	XuetangX MOOCs (315 courses)	F_1 improvement $\approx +7.2\%$ over PU-learning baseline	Reduces manual labeling effort Scales across academic disciplines	Dependent on dictionary quality Struggles with polysemous terms
Weakly Supervised	GlossEx: Weakly supervised concept extraction using contextualized embeddings (De Kuthy et al., 2025)	Contextualized BERT Embeddings + iterative clustering (GlossEx)	28 German Economics textbooks	$F_1 \approx 0.61$ Precision ≈ 0.72	Minimal annotation requirement Interpretable cluster centroids	Sensitive to seed selection Sensitive to cluster size k

LLM Concept Mapping	LLM-powered concept mapping (Cognitext) (Han et al., 2025)	Zero-shot Chain-of-Thought prompting (GPT-4o-mini)	10 academic disciplines	Precision ≈ 0.84 Recall ≈ 0.75 $F_1 \approx 0.79$	Domain-agnostic approach Human-like hierarchical structures	Output stochasticity Potential concept hallucination
LLM Prompt Evaluation	Examining GPT's capability to generate and map course concepts (Yang et al., 2025)	Few-shot ICL vs. Zero-shot prompting	XuetangX MOOCs Coursera MOOCs	F_1 : 0.74 – 0.88 (across GPT-4 variants)	Captures implicit concepts Strong prerequisite inference	High token cost Performance drops with dense graphs
LLM Extractive	ConExion: Concept Extraction with LLMs (Norouzi et al., 2025)	LLM-as-a-Judge for candidate ranking	Inspection SemEval-2017	$F_1@5 \approx 0.71$ $F_1@10 \approx 0.69$	High reproducibility State-of-the-art extractive accuracy	Extractive-only approach No abstractive concept generation
Unsupervised	UPreG: An Unsupervised approach for building the Concept Prerequisite Graph (Zhang et al., 2025b)	LDA topic modeling PMI-based similarity Wikipedia grounding	64k MOOC courses	Concept Accuracy ≈ 0.73 Prerequisite Accuracy ≈ 0.82	Fully unsupervised Highly scalable	Strong dependence on Wikipedia Weak for niche domains
Directed GNN	Concept Prerequisite Relation Prediction using Permutation Equivariant Directed GNNs (Qu et al., 2023)	Permutation Equivariant Directed GNN (PE-DGNN)	LectureBank MOOC-cube	F_1 : 0.78 – 0.91 AUC-ROC ≈ 0.94	Captures directionality Models non-linear dependencies	High computational complexity Large data requirement
Learner-Model	Discovering prerequisite relationships from an interpretable learner model (Allegre et al., 2023)	Dynamic Bayesian Networks (Causal discovery from logs)	ASSISTments (2009) Eedi	RMSE ≈ 0.12 High CPVD / NEC scores	Models true learning dependencies Data-driven causal inference	Requires massive student logs Limited to observed skills
Multi-criteria	Inferring Prerequisite Knowledge Concepts in Educational Knowledge Graphs (Alatrash et al., 2025)	Voting ensemble of lexical temporal and semantic features	AL-CPL Biology datasets	Precision: 0.48 – 1.0 Recall: 0.13 – 0.35	Highly transparent decisions Explainable Whitebox model	Extremely low recall Misses latent prerequisites
Global GNN	Learning Concept Prerequisite Relation via Global Knowledge Relation Optimization (Zhang et al., 2025a)	Multi-task GAT Global transitive consistency loss	LectureBank MOOCs	$F_1 \approx 0.82-0.88$ Accuracy ≈ 0.88	Ensures global graph consistency Transitive closure enforcement	High GPU memory usage Costly global optimization

5. RESEARCH GAPS AND CHALLENGES

Despite notable advances in automated concept extraction and dependency detection, several key research gaps still persist, especially in the context of educational applications. The majority of current research treats concept extraction and dependency detection as separate tasks, with the latter often handled using simple heuristics or post-processing techniques [7], [10]. This division may result in error propagation, where mistakes made during concept extraction negatively affect dependency detection, resulting in incomplete or unreliable prerequisite frameworks. Another significant drawback is the strong reliance on supervised learning approaches [1] that require large, domain-specific annotated datasets. The creation of such datasets is expensive and time-consuming, and they are typically limited to a small number of subjects, which hampers scalability and cross-domain applicability.

Although weakly supervised and distantly supervised methods [2], [3] help in reducing annotation effort, their effectiveness often depends on the quality of seed concepts, glossary coverage, or external knowledge sources, which makes them less reliable in diverse or rapidly evolving educational settings. The absence of standardized, publicly available benchmark datasets further complicates progress in this area. Fair comparison and reproducibility are challenging since many studies rely on the use of custom-built databases [7], [10]. Moreover, evaluation practices still remain inconsistent: concept extraction is commonly assessed using precision, recall, and F1-score, whereas dependency detection often lacks clear quantitative metrics and is frequently evaluated solely through qualitative analysis. Domain adaptability continues to remain a major challenge. Approaches that perform well within a single subject area often struggle to generalize across disciplines due to differences in terminology, teaching styles, and curriculum structure. Moreover, despite its growing importance in global learning environments, multilingual educational content has received little attention, though LLM-based approaches [4], [6] are beginning to address this through zero-shot capabilities.

Ultimately, explainability and interpretability still remain under explored, especially for models based on deep learning. In educational settings, it is essential for educators to understand why particular concepts or dependencies are identified in order to foster trust and facilitate meaningful adoption. Addressing these challenges will require integrated, scalable, and interpretable approaches that jointly model concept extraction and dependency detection [9], supported by robust and standardized evaluation frameworks.

6. PROPOSED SOLUTION

Through reviewing existing studies on automated concept extraction and prerequisite dependency detection, several shared shortcomings become apparent. Many current methods depend significantly on extensive annotated datasets, have difficulties generalizing across various subject areas, and provide limited interpretability. Furthermore, educational concepts are frequently implicit or abstract, which makes them challenging to identify with solely surface-level text analysis. To address these issues, this section presents a collection of practical and research-focused solutions designed to enhance the robustness, scalability, and real-world relevance of educational concept analysis systems.

6.1 Hybrid Rule-Based and Learning-Based Concept Extraction

One promising approach involves the implementation of hybrid frameworks that merge rule-based methodologies with machine learning algorithms. Rule-based techniques, such as linguistic patterns, part-of-speech rules, and syllabus-driven heuristics [1], can help produce an initial group of candidate concepts with high accuracy. These candidates can subsequently be improved through the use of supervised or weakly supervised models, like transformer-based classifiers or embedding similarity techniques [2], [3]. By integrating human expertise with data-driven learning, such hybrid strategies lessen the reliance on extensive annotated datasets while ensuring both accuracy and interpretability.

6.2 LLM-Assisted Concept Refinement with Human-in-the Loop Validation

Recent developments show that large language models (LLMs) perform well in recognizing implicit, abstract, or less common concepts that traditional natural language processing models often overlook [6]. However, LLMs also present challenges. These include producing incorrect outputs, being sensitive to prompt design, and having limited reproducibility [5]. To address these issues, LLMs should be seen as tools for refinement rather than just standalone extractors. In this way, LLMs can improve or reorganize concepts created by conventional methods [4]. This should be followed by a simple validation process by humans. This human-in-the loop approach balances automation and reliability, ensuring accuracy in the field while keeping the annotation workload manageable.

6.3 Semantic Embedding-Based Clustering for Domain Adaptability

Embedding-based clustering methods provide a versatile approach for managing differences in terminology across various fields. By utilizing contextualized embeddings produced by transformer models, suggested concept candidates can be organized into clusters that are semantically meaningful [3]. This facilitates the automatic detection of related subjects and aids systems in adjusting to new areas without the need for extensive retraining. Additionally, implementing adaptive clustering thresholds and seed expansion strategies can enhance stability, especially in educational domains that are low resource or have limited labeling [7].

6.4 Graph Neural Network-Based Dependency Inference

For identifying prerequisite dependencies, graph neural networks (GNNs) provide a useful framework for showing complex relationships among concepts. In these models, concepts appear as nodes with their related textual embeddings. Edges represent possible relationships of prerequisites [8]. GNNs can detect global structural patterns within the concept graph. This ability allows them to uncover dependencies that go beyond simple co-occurrence or sequencing indicators [11]. To keep things scalable, lightweight GNN designs or pruning techniques can reduce computational costs without greatly affecting performance.

6.5 Multi-Source Evidence Integration for Dependency Validation

The reliability of prerequisite inference can be improved by combining information from various sources. These include textbook structure, course syllabi, external knowledge repositories like Wikipedia, and data from learner interactions [10]. Merging these different signals helps reduce bias that might come from relying on just one source. This leads to a better understanding of concept connections. Techniques like confidence-based fusion or weighted voting methods can be used to identify only the most reliable prerequisite links. This process increases overall accuracy.

6.6 Explainability and Interpretability-Oriented Models

Understanding model decisions is important in education. Teachers and curriculum developers need to trust and understand what these systems produce. Future research should work on creating models that explain clearly how they extract concepts and determine dependencies. Techniques like attention visualization, rule tracing, and interpretable learner models, such as Bayesian networks, can help support automated predictions with educational logic and teaching standards [9].

6.7 Development of Standardized and Diverse Educational Benchmarks

Advancements in this area are also obstructed by the absence of standardized, publicly accessible datasets. This challenge is especially significant for curricula outside the Western context and for multilingual educational materials. Future efforts should aim at creating benchmark datasets that correspond with national education systems, such as the syllabi of Indian schools and universities, and should encompass multilingual and code-mixed texts. Having standardized benchmarks and evaluation protocols would facilitate fair comparisons among different methods and support reproducible research.

6.8 Scalable and Modular System Design

Ultimately, the effective implementation of educational concept analysis systems necessitates scalable and modular structures. By creating systems where concept extraction, refinement, dependency inference, and validation are not tightly linked, it becomes possible to update or replace specific components without needing to redesign the whole pipeline. This kind of modularity enhances long-term maintenance, scalability, and the ability to integrate into actual educational platforms.

7. CONCLUSION AND FUTURE WORK

This survey examined and synthesized existing research on automated concept extraction and dependency detection in educational texts, with particular attention to the methods used, the datasets considered, the evaluation strategies adopted, and the limitations reported in prior work. The reviewed studies show that substantial progress has been made in concept extraction, especially through supervised transformer based models [1] that are capable of capturing rich contextual and semantic information. Weakly supervised and distantly supervised approaches [2], [3] further support this progress by reducing the need for large, annotated datasets and improving scalability, which is especially valuable in educational settings where labeled data is often limited. Nevertheless, these approaches tend to be sensitive to factors such as seed quality, glossary coverage, and the availability of domain specific resources. The comparative analysis also indicates that most existing research prioritizes concept extraction, while dependency detection is commonly treated as a secondary task and addressed using heuristic, structural, or graph-based methods [7], [10] with limited quantitative evaluation. Consequently, dependency detection remains less mature and less standardized than concept extraction. In addition, the absence of publicly available benchmark datasets and consistent evaluation protocols makes it difficult to directly compare results across studies and restricts reproducibility. Domain dependence continues to be a recurring challenge, as many approaches perform well only within specific subject areas or instructional contexts [5]. Overall, the survey suggests that no single method fully meets the combined requirements of accuracy, scalability, domain adaptability, and interpretability. Supervised approaches often achieve high performance but rely heavily on large, labeled datasets, whereas weakly and distantly supervised methods offer greater scalability at the cost of reduced accuracy. These observations highlight the need for more comprehensive solutions that jointly address concept extraction and dependency detection [11] rather than treating them as independent problems. Looking forward, several promising directions for future research emerge. Developing integrated models that perform concept extraction and dependency detection simultaneously could help reduce error propagation and improve overall system reliability. Advances in representation learning and graph based modeling [8], [11] also offer opportunities to better capture complex relationships among educational concepts. Equally important is the creation of standardized, publicly available datasets that span multiple disciplines and educational levels, enabling fair evaluation and improved cross domain generalization. Future research should also focus on domain-adaptive and multilingual approaches to better reflect the diversity of educational content. Incorporating explainability and interpretability into automated systems is particularly crucial in educational contexts, where transparency plays a key role in building trust and encouraging adoption by educators [9]. Finally, combining large language models with structured knowledge representations [4], [6] presents a promising path toward more flexible and scalable educational concept analysis, provided that challenges related to control, evaluation, and reliability are carefully addressed. By consolidating existing work and identifying key open challenges, this survey aims to support and guide future efforts toward the development of more effective, scalable, and trustworthy automated systems for concept extraction and dependency detection in educational applications.

REFERENCES

- [1] J. Rico and J. P. Espada, "YodKw: An expert system for extracting keywords in educational texts and textbooks," *Expert Systems with Applications*, 2025.
- [2] Anonymous, "DS-MOCE: Distantly supervised course concept extraction," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.
- [3] K. De Kuthy, L. Gurrbach, and D. Meurers, "GlossEx: Weakly supervised concept extraction using contextualized embeddings," in *Proc. BEA Workshop*, 2025.
- [4] S. Han and Y. Choi, "LLM-powered concept mapping for educational content," *IEEE Trans. Learning Technologies*, 2025.
- [5] Z. Yang et al., "Examining GPT's capability to generate and map course concepts," *arXiv preprint arXiv:2504.08856*, 2025.
- [6] M. Norouzi et al., "ConExion: Concept extraction with large language models," *arXiv preprint arXiv:2504.12915*, 2025.
- [7] Anonymous, "UPreG: An unsupervised approach for building concept prerequisite graphs," *IEEE Trans. Knowledge and Data Engineering*, 2025.
- [8] H. Liang et al., "Concept prerequisite relation prediction using permutation-equivariant directed graph neural networks," *IEEE Trans. Neural Networks and Learning Systems*, 2023.
- [9] L. Allegre et al., "Discovering prerequisite relationships between knowledge components using interpretable learner models," *IEEE Trans. Learning Technologies*, 2023.
- [10] Alatrash et al., "Inferring prerequisite knowledge concepts in educational knowledge graphs: A multi-criteria approach," *Knowledge-Based Systems*, 2025.
- [11] J. Zhang et al., "Learning concept prerequisite relations via global knowledge relation optimization," *IEEE Trans. Knowledge and Data Engineering*, 2025.