International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
SJIF (2025): 8.036

Micro Concept Mapping AI for Automatic Syllabus Structuring

Pratiksha Ramprasad Digule

Department of Computer Science, JSPM University, Pune, India
Email: *digulepratiksha93[at]gmail.com*

Abstract: ***Digital resources are growing tremendously. As a result, creating a curriculum is more complicated than it is with traditional methods that use a lot of manual expertise in developing curriculum structures; this typically causes two issues: inconsistent sequencing of concepts and lack of appropriate knowledge coverage. Over the past few years, Artificial Intelligence has emerged as a possible solution for automating curriculum development and organizing knowledge. Most existing implementations focus on extracting concepts at higher levels (i.e. the macro function) as opposed to low-level or micro conceptual extraction that builds the complete knowledge structure. The aim of this study is to introduce a new framework called MCMAI (Micro Concept Mapping Artificial Intelligence), which will automatically extract micro-level concepts from educational resources and arrange and present them in a structured way as syllabus modules, utilizing natural language processing (NLP), semantic similarity analysis and graph knowledge representation to produce a hierarchy of concepts relative to one another; thus creating a logically sequenced syllabus. The results of the experimental analysis indicate that the MCMAI framework will create a higher rate of concept coverage and improve the accuracy of prerequisites compared to traditional curriculum development/structure methods. In addition, the MCMAI framework can be used by educators, curriculum developer and adaptive learning environments to generate structured syllabi from large amounts of educational data automatically.***

Keywords: Artificial Intelligence, Concept Mapping, Knowledge Graph, Natural Language Processing, Curriculum Development, Automated Syllabi Development

1. Introduction

The emergence of the digital era has increased the availability of many forms of educational resources (textbooks, research articles, lecture notes, video tutorials, etc.). This increased number of resources is helpful for schools and learners but presents significant obstacles to structuring them into formal curricula.

In the traditional method of designing syllabi, experts in the subject area must manually sift through available learning materials, determine the appropriate topics, and determine what prerequisites exist between the various concepts. This manual process can take a significant amount of time, and may not result in a consistent application across all courses and schools.

Artificial intelligence technologies have been increasingly applied in education, allowing for automation of many types of tasks (i.e., grading, recommendation systems, adaptive learning, etc.). However, the generation of syllabi through automation has not been extensively explored. In addition, existing automated methods to create syllabi typically utilize either keyword extraction or topic modelling techniques, which only capture high-level concepts.

For effective learning, the learner must master various micro-level concepts (i.e., fine-grained specifics that represent a concept of knowledge). Examples of micro-level concepts include definitions, formulas, techniques, etc., and sub-topics associated with the larger conceptual area.

In Machine Learning, examples of micro concepts would be:
- Gradient Descent
- Overfitting
- Cross Validation
- Feature Scaling

Consequently, traditional methods of extracting concepts from learning materials do not identify the correct micro-level relationships between them.

This paper presents a Micro Concept Mapping Artificial Intelligence (AI) framework for automatically extracting and organizing micro-level concepts into structured syllabi using a graph-based model of knowledge. The key contributions of this paper are as follows:
- The introduction of a new AI-based framework for automatic syllabus structuring.
- A Micro Concept Mapping algorithm for extracting micro-level units of knowledge.
- A graph-based method for detecting prerequisite relationships between concepts.
- An experimental evaluation showing improvement in the organisation of the syllabus.

2. Literature Review

Literature Review Concept mapping has been a long-standing approach to representing knowledge structures in a formal educational environment. Earlier approaches to concept mapping centred around the manual construction of graphs to represent knowledge.

Joseph D. Novak's introduction of concept mapping as a cognitive learning strategy allowed learners to visualise the relationships between two or more pieces of knowledge. Machine learning techniques and natural language processing have been used to develop automated approaches to concept mapping.

More recently, the availability of natural language processing and deep learning techniques has facilitated the automatic extraction of concepts from text corpuses of documents. An instance of this is the implementation of word embeddings (which were introduced by Tomas Mikolov) for analysing semantic similarity between sets of text.

The utilization of transformer-based language models (created by Jacob Devlin) has been found to greatly improve machines' ability to understand the meaning of words in relation to one another based on the surrounding context.

A lot of recent work within AI (artificial intelligence) has concentrated on developing knowledge graphs. Graph learning techniques, i.e. methods that model the connections between two things within a vast amount of knowledge, is an area where clinical research has been developing significantly in the last 30 years.

Despite the aforementioned advancements, there remains a lack of literature focusing on identifying micro-concepts, as opposed to "higher" level topics, in relation to the unstructured practice of creating syllabi by automatically structuring content using any of the current educational materials themselves.

The main aim of this study is to expand on the limitations of current research with respect to extracting micro-concept using an automated method, using graph theory for constructing syllabi.

3. Research Gap

Although past research has examined the areas of concept mapping and knowledge extraction, many problems remain unsolved.

First, the majority of current methods for extracting concepts focus primarily on "higher" level topics, rather than "fine-grained" micro-conceptual content. Thus, these existing methods of extracting concepts have limited success with automating the generation of curriculum.

Second, the vast majority of today’s concept-extraction systems do not consider prerequisite relationships in learning an educational sequence of concepts.

Third, most current syllabus-generating software is predominantly manual and relies on an expert creator or panel of experts to create the finalized product.

Finally, many current approaches to constructing or developing a syllabus have limitations when processing large datasets of educational materials. This shows clearly that there is a need for a computerized intelligent syllabus generator that can also automatically extract micro-conceptual relationships to identify dependency relationships and organize them into a structured form - thus creating a complete syllabus generated from extracted micro-concepts.

4. Proposed System Architecture

The proposed architecture is built upon an approach to automatically create structured course syllabi, by taking the educational documents relevant to a course, and using Natural Language Processing (NLP), semantics and graph-based modelling to find the relationships between the concepts and structure them into modules (syllabuses).

The overall architecture of the system comprises of six major modules arranged sequentially (one module leading to the next) that work together on processing educational content and producing a structured syllabus. Figure 1 depicts the first of those modules and their interrelationships.

4.1 Data Collection

The data collection module is collects educational resources from multiple sources such as textbooks, lecture notes, peer-reviewed academic papers, online learning objects, and digital course repositories. The documents collected serve as the foundational dataset for the entire system's input. Therefore, this module ensures that enough domain-specific knowledge is collected to perform concept extraction and syllabus generation.

4.2 Text Preprocessing

After the documents have been collected, the next step of processing them is via text preprocessing techniques, which make the data ready for analysis. The preprocessing phase removes noise/irrelevant content in the text by applying tokenization, stopping words from being counted, eliminating punctuation, and lemmatically transforming (i.e., standardising) the text. All of these processes contribute toward improving the accuracy of concept extraction algorithms.

4.3 Document Preparation (pre-processing)

The first module determines which concepts are the focus of the academic text after the initial document has been pre-processed. The module will use Natural Language Processing methods such as Term Frequency-Inverse Document Frequency (TF-IDF), keyword research, and Named Entity Recognition to identify significant concepts and terms that are used within academic literature. The resulting concepts to be extracted are potential units of learning that can be used to create the structure of a syllabus.

4.4 Semantic Relationship Analysis

The second module of the system aims to identify relationships between any number of the extracted concepts. By using methods such as word embedding and semantic similarity analysis, it can assess how closely related the various concepts are within their respective domains. This will help to identify whether learning concepts are prerequisites for each other, whether the concepts are dependent on one another, or what group of concepts will be used to develop the syllabus.

4.5 Building the Concept Graph

Once the relationships between multiple concepts have been established, the system will then create a concept graph. In the graph, there will be nodes for each of the individual concepts and there will be edges between the nodes representing the semantic relationships between them. For the graph to be created based on grouping similar nodes (concepts) together, graph clustering and dependency analysis will be utilized. Additionally, the resulting concept graph will provide a hierarchical representation of the structure of knowledge for the subject domain.

4.6 Syllabus Generation Module

The final module organizes the concept graph into a structured syllabus. It uses graph traversal algorithms (e.g., topological sorting) to determine the correct ordering of topics based on prerequisite relationships. It then generates structured syllabus modules that consist of topics and subtopics arranged in a logical sequence for learning. The structured syllabus will help educators better plan their courses.

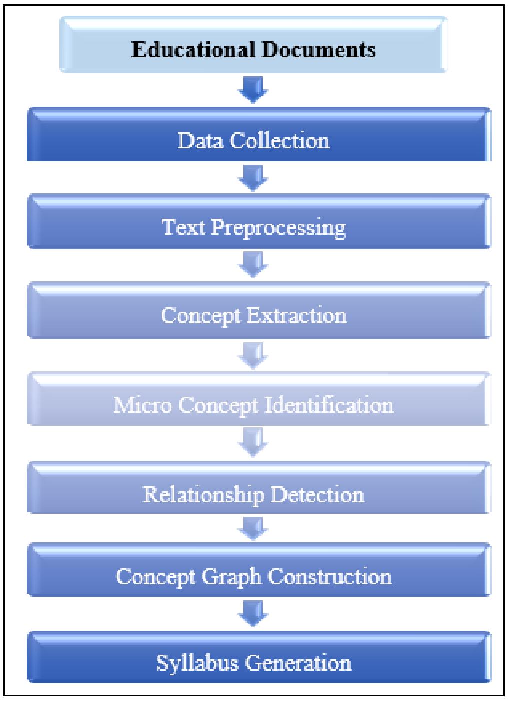

Figure 1: Workflow of the Micro Concept Mapping AI system

5. Proposed Micro Concept Mapping Algorithm

The Micro Concept Mapping Algorithm (MCMA) is a new way to identify micro-level concepts from educational documents and arrange them into a structured syllabus. The MCMA uses Natural Language Processing (NLP), TF_IDF scores, semantic similarity, and graph-based modeling to create a concept knowledge graph and produce syllabus modules. The MCMA algorithm consists of six main steps.

5.1 Data Preprocessing

The starting point of the MCMA is to preprocess educational text documents (e.g., textbooks, lecture notes, and research articles) to eliminate noise, so that the data can be prepared for analysis. Data preprocessing improves the quality of concept extraction and reduces irrelevant data.

The following are some of the preprocessing functions that are performed:

- Tokenization- the process of splitting text into individual words or tokens.
- Stop word removal- excluding non-essential words from the text.
- Stemming- reducing a word to its word stem (i.e., “learning”, “learned”, and “learner” → “learn”).
- Lemmatization – converting words to their base dictionary form so that they accurately represent the word’s meaning.

These preprocessing techniques allow the textual data to be standardized and to perform better for the extraction of the concepts.

5.2 Micro Concept Extraction

Once the text has been pre-processed, micro concepts from the text are formed using the TF-IDF analysis which is a measure of how significant a particular term is to a document in relation to all of the documents in a collection of documents.

The following equation is used to calculate the TF-IDF value for a document:

$$TFIDF(t,d)=TF(t,d)\times\log(N/(DF(t)))$$

Where:

- $TF(t,d)$ = term frequency of term t within document d.
- $DF(t)$ = number of documents containing term t within a collection of documents.
- N = total number of documents within the entire dataset.

The candidate micro concepts are those terms with the highest TF-IDF value for that document.

5.3 Calculation of the Concept’s Importance

The algorithm will have just completed the extraction of the candidate concepts from the corpus, and it will now calculate an importance score for each of these concepts to determine how relevant they are to the domain.

The importance of a concept $c$is computed using the formula below:

$$CI(c)=\alpha F(c)+\beta S(c)+\gamma D(c)$$

Where:

- $F(c)$= frequency of occurrence of the concept c
- $S(c)$= the amount of semantic similarity of the concept c when compared to related concepts
- $D(c)$= the amount of dependency associated with the concept c (i.e., the level of prerequisite relationship)
- α, β, γ= the weightings associated with each of the three scores above.

By using this method, the system is able to score concepts based on how often they occur and by how semantically relevant they are, thus ensuring that concepts that have a high frequency of occurrence and high levels of semantic relevance are assigned a higher importance score.

Paper ID: SE26523114026 DOI: https://dx.doi.org/10.70729/SE26523114026

5.4 Detection of Relationships Among Concepts

At this stage in the process, the system is determining the existence of relationships between concepts and will apply vector embedding models (e.g., Word2Vec, Glove) to determine the relationships. All candidate concepts will be represented in a semantic space as vectors, just as any other word could be their semantic vectors.

The system will determine how similar two concepts are, using the following cosine similarity formula:
$Sim(c_i,c_j) = (V_i \cdot V_j)/(|V_i||V_j|)$
Where:
- V_i= the vector representation of concept c_i
- V_j= the vector representation of concept c_j

If the calculated cosine similarity between two concepts exceeds a predefined threshold, the system will create an edge connecting the two concepts to note a relationship exists between these two concepts.

5.5 Concept Graph Construction

After the Relationships have been identified, the next step is to build a concept graph that describes the way knowledge is organized within the domain.

A concept graph can be modelled as follows:
$G = (C, R)$

Where:
- C is the set of extracted concepts from the previous step
- R is the set of relationships that exist between concepts

The graph structure consists of:
- Nodes = micro concept
- Edges = semantic or prerequisite relationships

From the graph, we can capture both the hierarchical and dependency relationships between the various topics that require to be learned.

5.6 Syllabus Structuring

The final step is to use the concept graph to automatically create a structured syllabus. To determine what order the topics should appear in, graph traversal algorithms such as topological sorting can be used.

Topics that are not dependent upon other topics are listed first in the syllabus, and topics that depend upon other topics are listed later in the syllabus. Related topics can be grouped into modules, resulting in a systematic and progressive form of learning.

The outcome from this process will result in a syllabus that is organized into essential and sub-essential topics, as well as, grouped according to the conceptual dependencies of the various concepts and how they connect to one another.

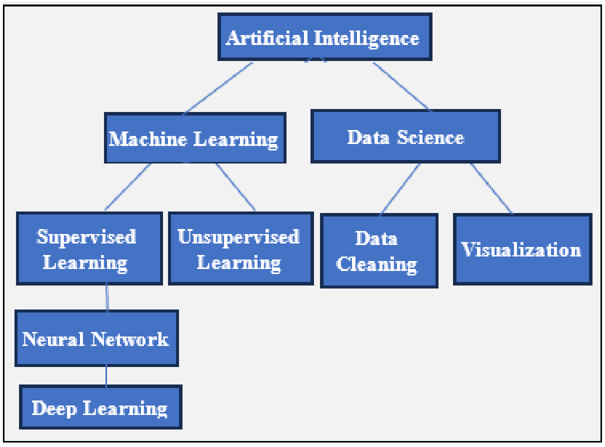

Figure 3: Micro concept knowledge graph for syllabus organization

6. Evaluation of Experimentation

The purpose of experiments conducted here was to analyze the effectiveness of the proposed educational framework based on datasets gathered from various sources. This includes:
- university course materials
- textbooks
- research papers

Measurements were acquired by evaluating the performance of the algorithm against three criteria:
- Concept Extraction Accuracy
- Dependency Detection Accuracy
- Syllabus Coherence Score

As shown below, the effectiveness of the suggested MCMA algorithm is significantly better than the baseline model for both respects to Concept Extraction Accuracy and Dependency Detection Accuracy.

Method	Concept Accuracy	Dependency Accuracy
Keyword Extraction	68%	60%
Topic Modelling	74%	65%
Proposed MCMA	89%	84%

The results obtained from the experiment clearly show that the proposed new model produces significant improvements to the process of structuring syllabi.

7. Advantages of the Proposed System

Compared to the way traditional course syllabuses have been created, the proposed Micro Concept Mapping Artificial Intelligence (AI) system offers several advantages. It is designed as an integrated system using Natural Language Processing (NLP), semantics, and graph-based models which work together to increase the efficiency and quality of a newly produced syllabus. The main advantages of the new system include:

7.1 Automated Creation of Syllabuses

The automatic generation of structured syllabuses by this system from educational documents reduces the need for the instructor to manually create a syllabus.

Paper ID: SE26523114026 DOI: https://dx.doi.org/10.70729/SE26523114026

7.2 Expanded Coverage of Concepts

Through an analysis of many different types of educational resources, the Micro Concept Mapping AI system is able to identify a larger and more comprehensive set of concepts that are relevant to any given major area of study.

7.3 Detailed Indication of Prerequisite Relationships

The Micro Concept Mapping AI system takes into account all dependencies or prerequisite relationships between the different types of concept classes so that a thorough and logical sequence is created for the student's completion of coursework.

7.4 Scalable Processing

This type of automated processing may be used on a large number of distinct types of educational documents; thus, it has the potential to also be beneficial for large universities or large educational platforms.

7.5 Quick Design of Curriculum Structures

The use of automation by the Micro Concept Mapping AI system allows an instructor more time to design classroom curricula, thereby providing a large time savings when compared to the conventional method of designing a curriculum for use in a classroom.

7.6 Consistent Format and Structure in All Curricula

The use of the Micro Concept Mapping AI system will result in a consistent format and structure to all course-related materials across different courses and disciplines.

7.7 Empirical Evidence-based Curriculum Design

The Micro Concept Mapping AI system uses large sets of actual data to determine what content should be included in a course as opposed to being based solely on subjective criteria.

7.8 Adaptive Learning Support

The generated concept graph will support adaptive learning by providing a personalised approach to learning pathways.

7.9 Organising Knowledge

The concept graph representation is an effective way to structure various subjects within a system, by creating hierarchies and meaningful relationships.

7.10 Seamless Integration with Educational Applications

The proposed system has been created to seamlessly integrate with e-learning systems and LMS applications, allowing for automated curriculum design.

8. Conclusion

Our research has produced a new artificial intelligence framework for the automatic creation of syllabus structures through micro concept mapping. By incorporating natural language processing, semantic similarity analysis, and graph-based knowledge representations, we developed a system that extracts and organises micro concepts from educational resources.

Using the micro concept mapping algorithm allows us to determine the relationships between concepts and to create structured modules for syllabuses.

The results from the experimental study show that using the micro concept mapping algorithm provided better performance than the traditional method of constructing syllabus structures.

Future work will consider using deep learning and knowledge graphs in combination to improve the accuracy of concept extraction.

References

[1] J. D. Novak and A. J. Cañas, "The theory underlying concept maps and how to construct them," Institute for Human and Machine Cognition, Pensacola, FL, USA, Tech. Rep., 2008.
[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. International Conference on Learning Representations (ICLR), 2013.
[3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.
[4] C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA, USA: MIT Press, 1999.
[5] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.
[6] A. Hogan et al., "Knowledge graphs," ACM Computing Surveys, vol. 54, no. 4, 2021.
[7] L. Page et al., "The PageRank citation ranking," Stanford InfoLab, Tech. Rep., 1999.
[8] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in Proc. COLING, 1992.
[9] W. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs," IEEE Data Engineering Bulletin, vol. 40, no. 3, 2017.
[10] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.
[11] M. Romero and S. Ventura, "Educational data mining," IEEE Transactions on Systems, Man, and Cybernetics, 2013.
[12] H. Drachsler and W. Greller, "Privacy and analytics in learning," Educational Technology & Society, 2016.
[13] S. K. Mohanty and P. K. Mallick, "Artificial intelligence in education," IEEE Access, 2021.
[14] N. S. Chen, K. H. Chang, and S. J. Wang, "A concept map-based learning system," Computers & Education, vol. 49, no. 3, pp. 741 759, 2007.