Application of Ontology-Driven NLP for the Analysis of Text

—

Guide : Dr. Ujwala H. Bharambe

Group No. **35**
04 Eashan Bajaj
92 Aakash Ramchandani
122 Omkar Yadav

Motivation

Examination being the lone measure of competence in the current education system of our country, plays a decisive role in building of a student's career.

Hence, utmost care has to be taken in framing the question paper. However, setting up a good question paper for assessment is a challenging task.

Fairness, accuracy, consistency and elimination of bias are very important while selecting questions into the paper.

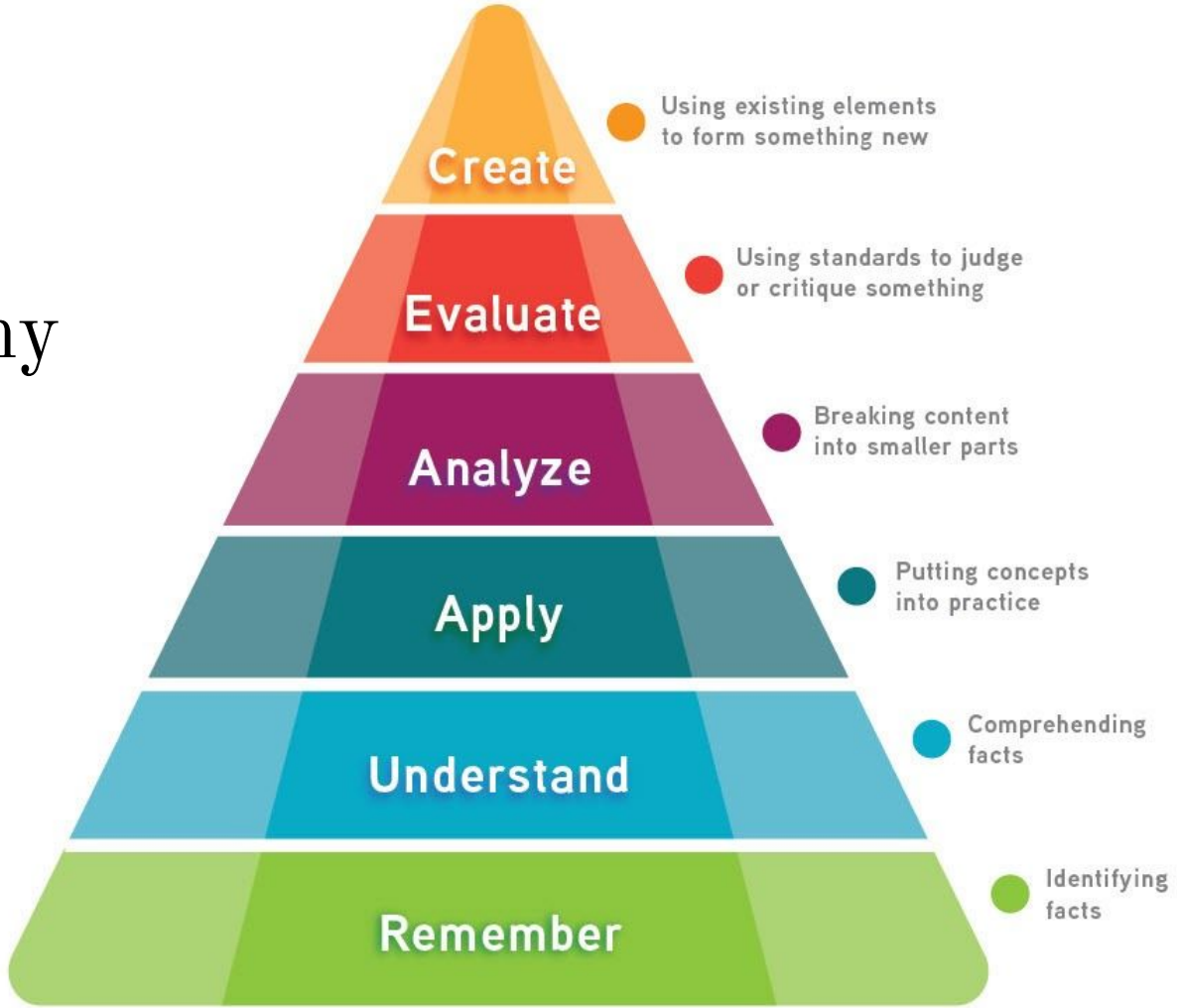
Our primary **objective** is to propose a feasible method, taking into consideration several aspects to **accurately** tackle the problem of judging syllabus **fairness.**

Module No.	Topics	Hrs.
1.0	**Introduction to Data Warehouse and Dimensional modelling:** Introduction to Strategic Information, Need for Strategic Information, Features of Data Warehouse, Data warehouses versus Data Marts, Top-down versus Bottom-up approach. Data warehouse architecture, metadata, E-R modelling versus Dimensional Modelling, Information Package Diagram, STAR schema, STAR schema keys, Snowflake Schema, Fact Constellation Schema, Factless Fact tables, Update to the dimension tables, Aggregate fact tables.	8
2.0	**ETL Process and OLAP:** Major steps in ETL process, Data extraction: Techniques, Data transformation: Basic tasks, Major transformation types, Data Loading: Applying Data, OLTP Vs OLAP, OLAP definition, Dimensional Analysis, Hypercubes, OLAP operations: Drill down, Roll up, Slice, Dice and Rotation, OLAP models : MOLAP, ROLAP.	8
3.0	**Introduction to Data Mining, Data Exploration and Preprocessing:** Data Mining Task Primitives, Architecture, Techniques, KDD process, Issues in Data Mining, Applications of Data Mining, Data Exploration :Types of Attributes, Statistical Description of Data, Data Visualization, Data Preprocessing: Cleaning, Integration, Reduction: Attribute subset selection, Histograms, Clustering and	10
4.0	**Classification, Prediction and Clustering:** Basic Concepts, Decision Tree using Information Gain, Induction: Attribute Selection Measures, Tree pruning, Bayesian Classification: Naive Bayes, Classifier Rule - Based Classification: Using IF-THEN Rules for classification, Prediction: Simple linear regression, Multiple linear regression Model Evaluation & Selection: Accuracy and Error measures, Holdout, Random Sampling, Cross Validation, Bootstrap, Clustering: Distance Measures, Partitioning Methods (*k*-Means, *k*-Medoids), Hierarchical Methods(Agglomerative, Divisive)	12
5.0	**Mining Frequent Patterns and Association Rules:** Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rule, Frequent Pattern Mining, Efficient and Scalable Frequent Item set Mining Methods: Apriori Algorithm, Association Rule Generation, Improving the Efficiency of Apriori, FP growth, Mining frequent Itemsets using Vertical Data Format, Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules	8
6.0	**Spatial and Web Mining:** Spatial Data, Spatial Vs. Classical Data Mining, Spatial Data Structures, Mining Spatial Association and Co-location Patterns, Spatial Clustering Techniques: CLARANS Extension, Web Mining: Web Content Mining, Web Structure Mining, Web Usage mining, Applications of Web Mining	6
	Total	52

Fairness

What makes a question paper fair?

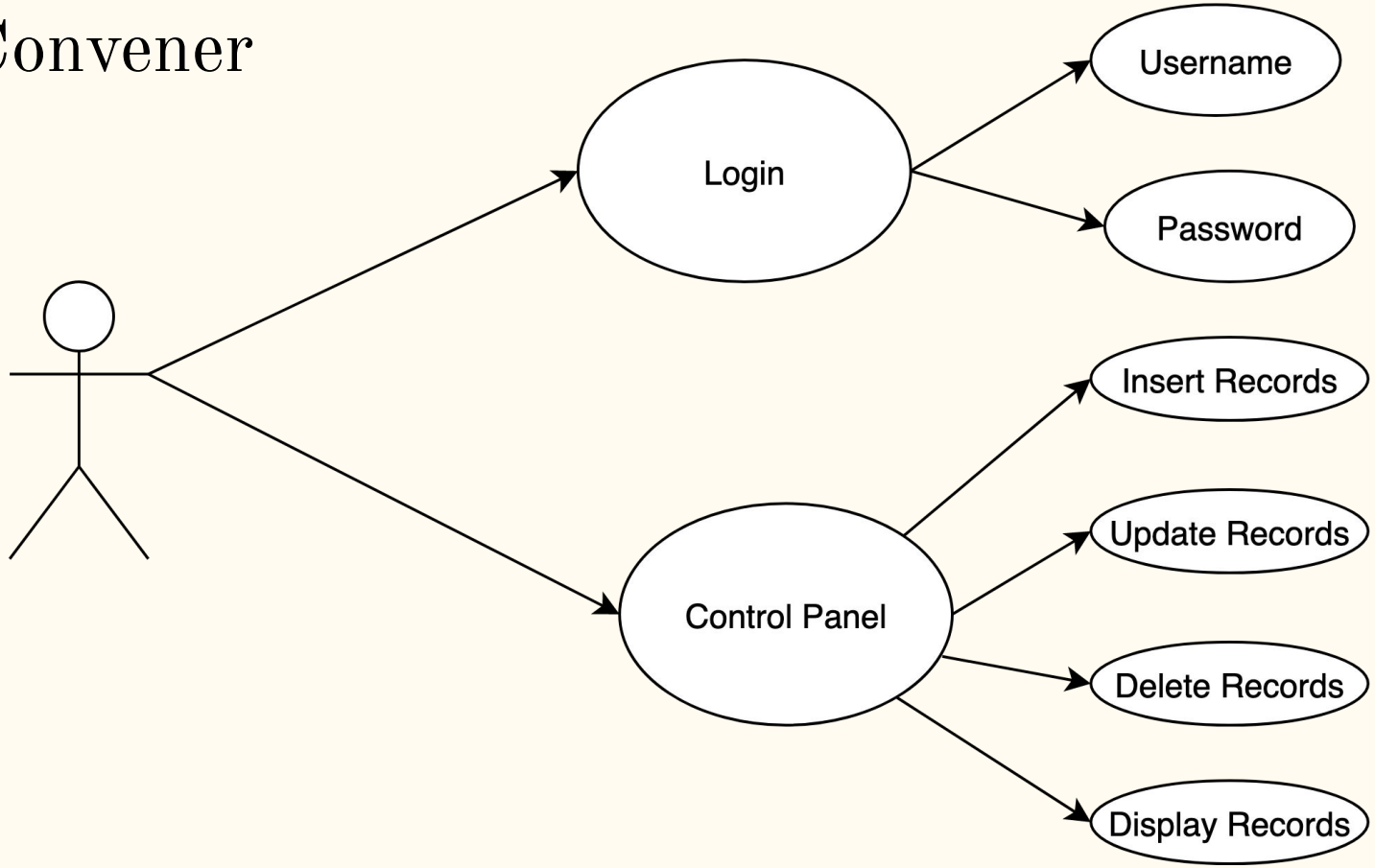
Bloom's Taxonomy



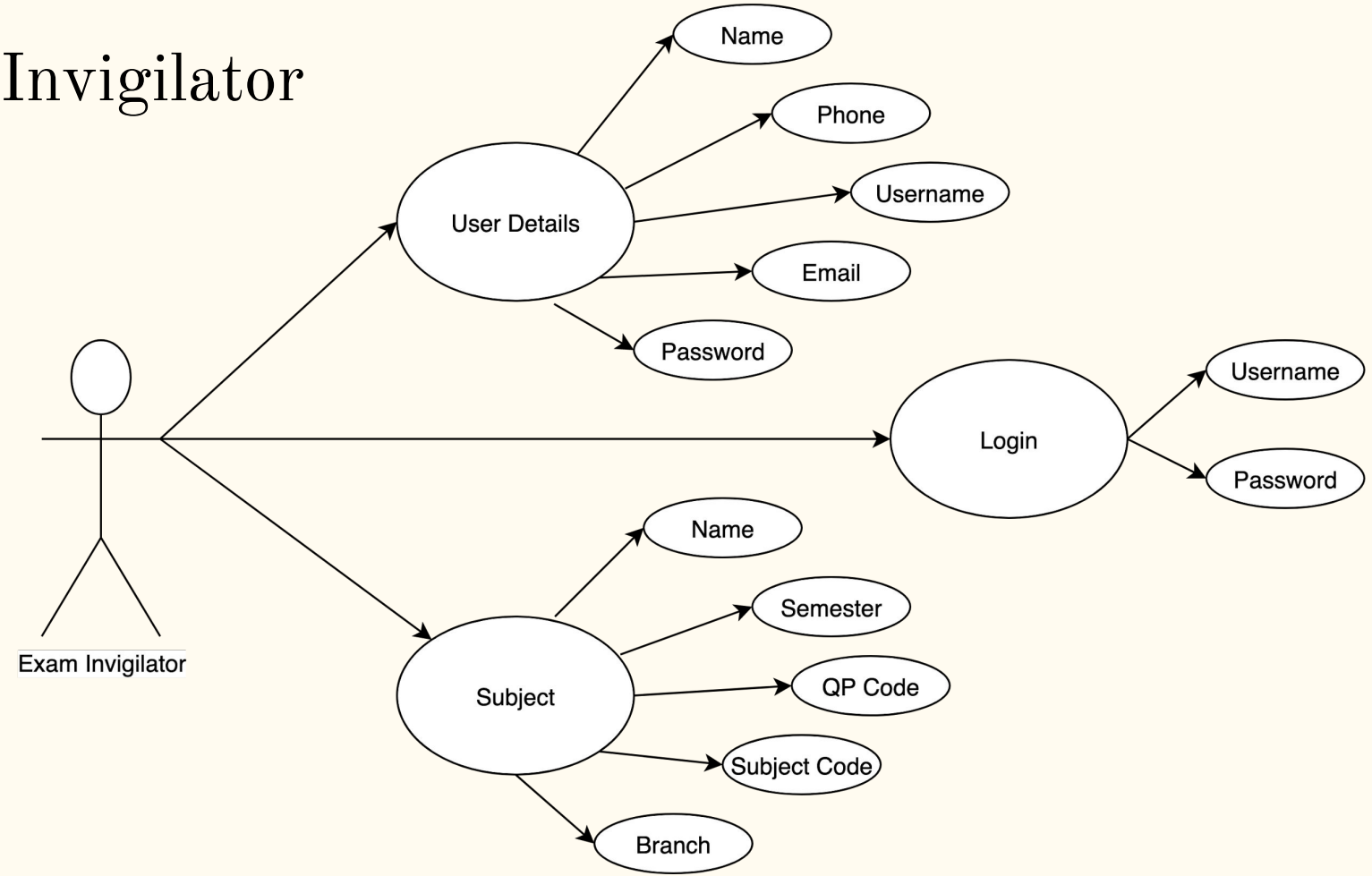
Solution

Our approach to tackle the problem.

Exam Convener



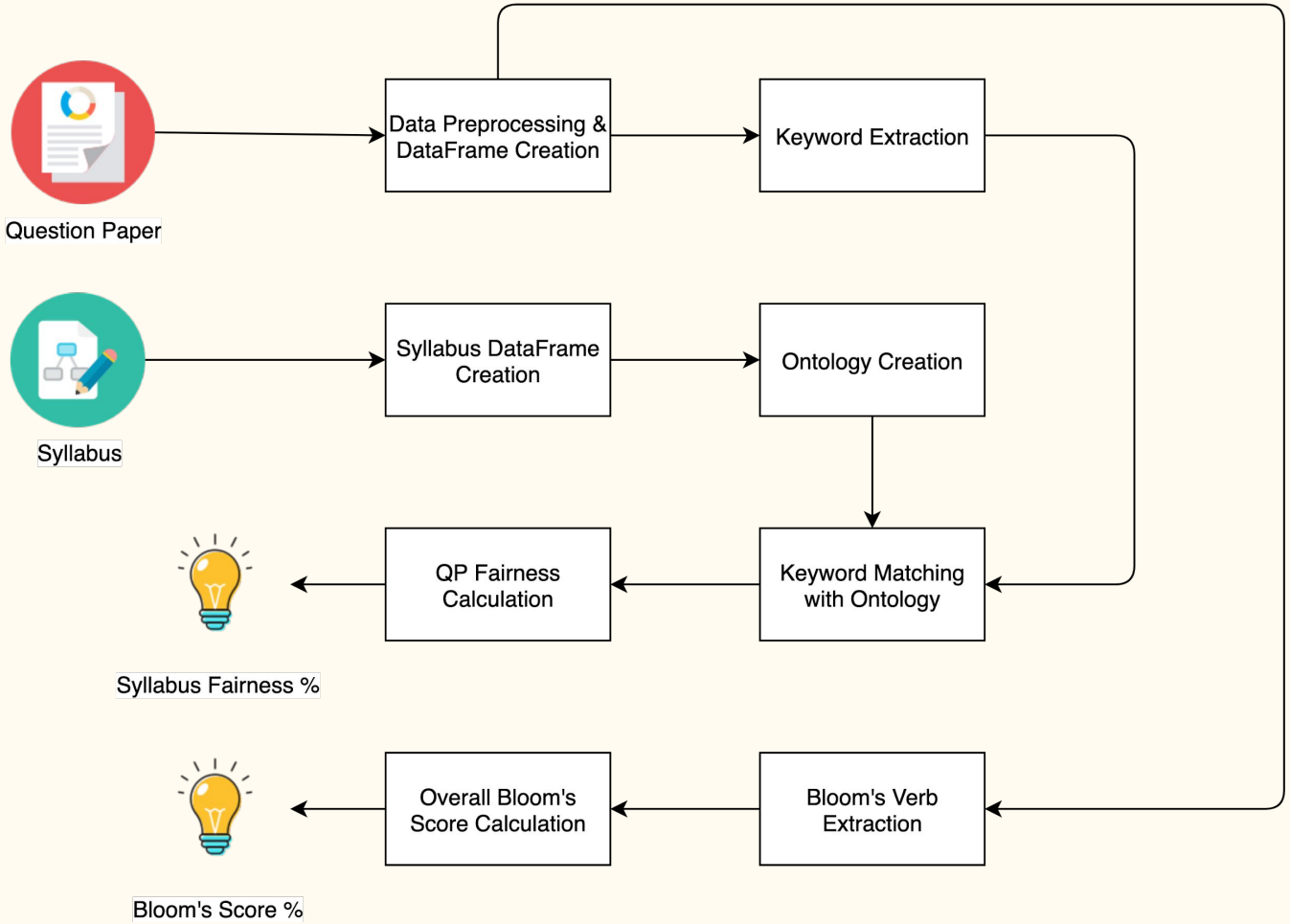
Exam Invigilator



Project Demo

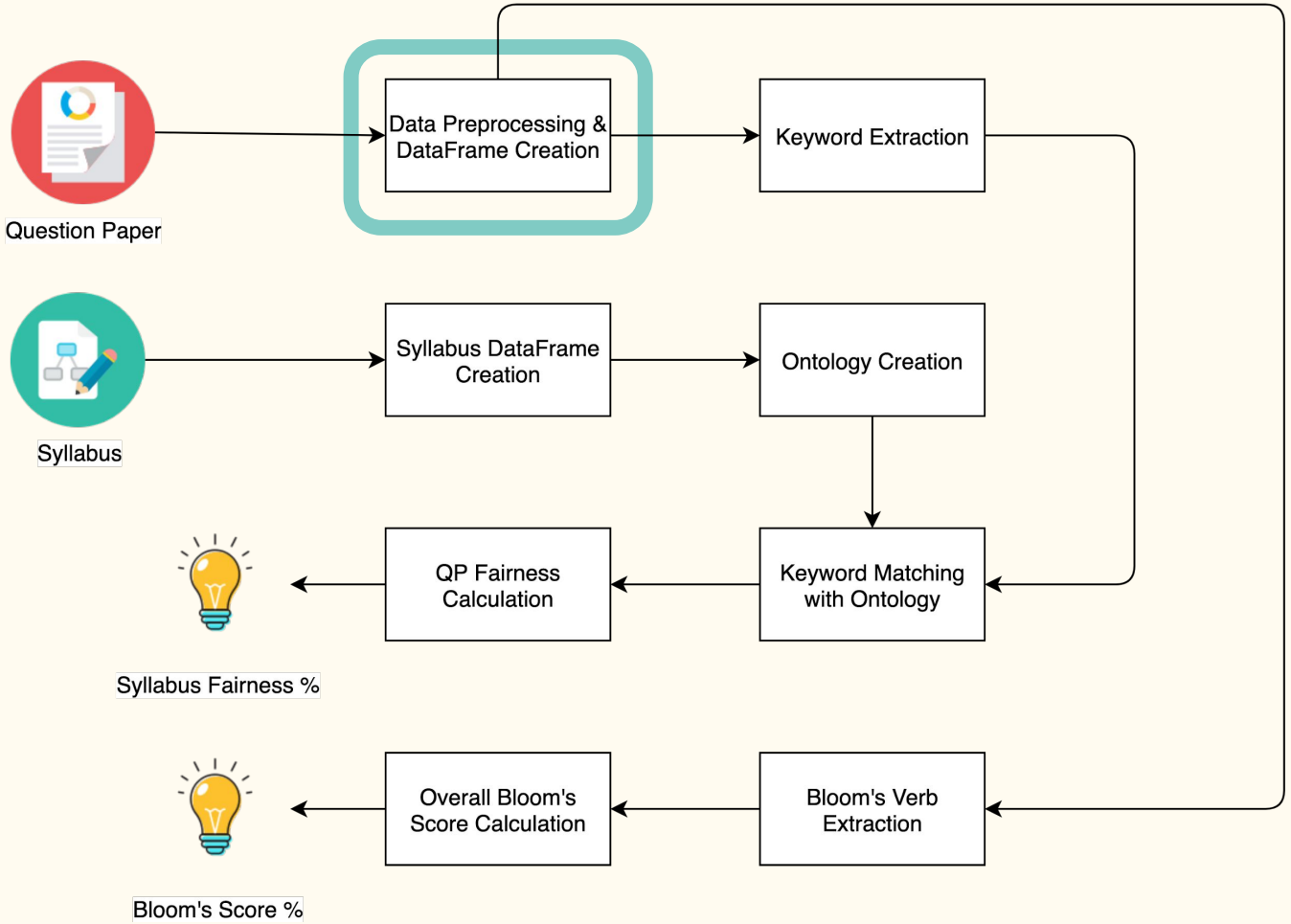
Watch a live **demonstration** of our project.

System Architecture



Question Paper DF

pandas



Output:

	Num	Question	Marks
0	2a	Write a program to implement Circular Linked L...	10
1	4b	Explain different cases for deletion of a node...	10
2	5a	Write a program in 'C' to implement Stack usin...	10
3	5b	Explain Depth First search (DFS) Traversal wit...	10
4	6a	Application of Linked-List –Polynomial addition	10
5	6d	Topological Sorting	10

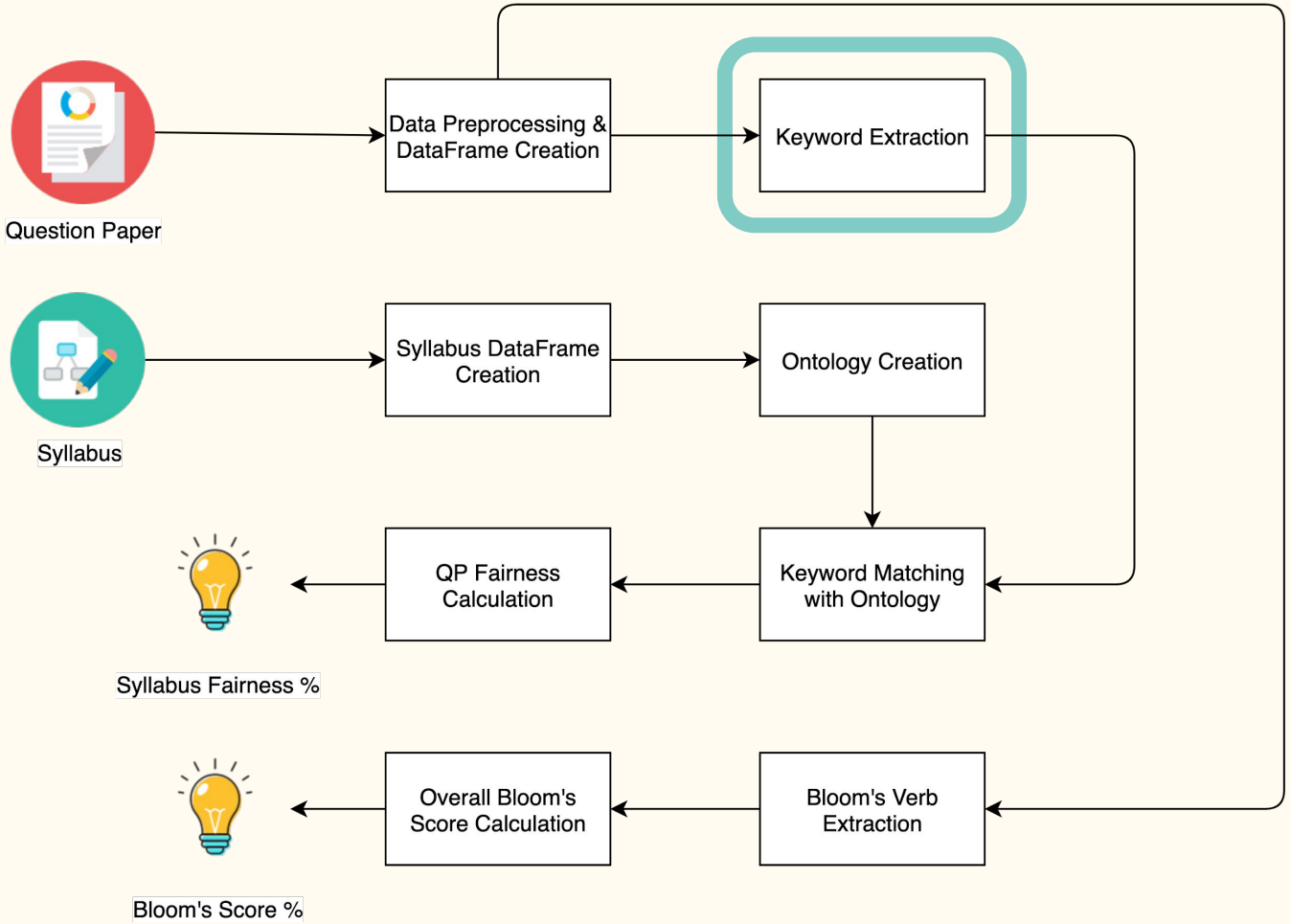
Keyword Extraction

pandas

Python RegEx

nltk wordnet

nltk stopwords



Output:

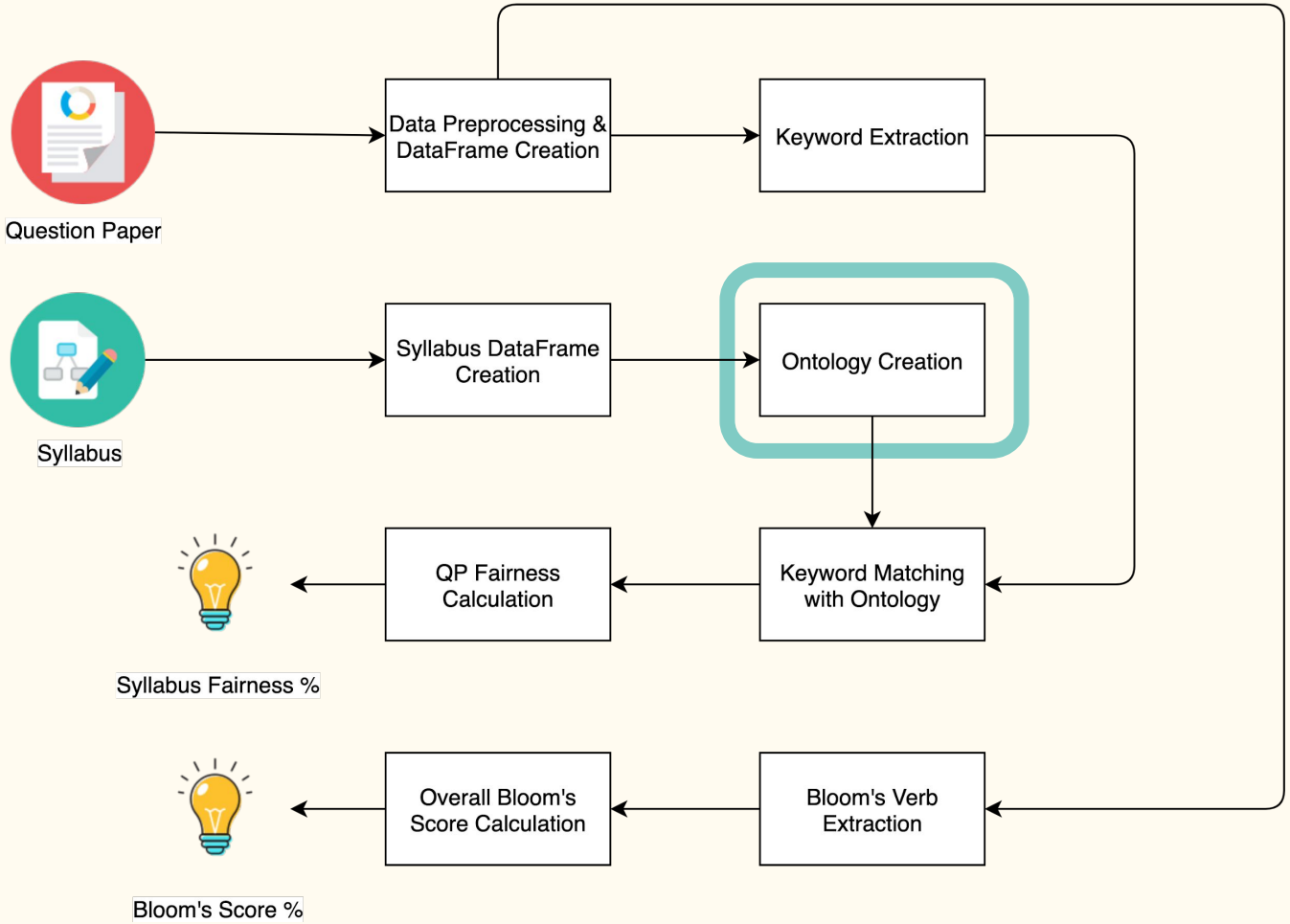
	Num	Question	Marks	Extracted Keywords
0	2a	Write a program to implement Circular Linked L...	10	circular linked
1	4b	Explain different cases for deletion of a node...	10	case deletion node binary search tree function
2	5a	Write a program in 'C' to implement Stack usin...	10	c stack linked
3	5b	Explain Depth First search (DFS) Traversal wit...	10	depth first search dfs traversal recursive fun...
4	6a	Application of Linked-List –Polynomial addition	10	application linked polynomial addition
5	6d	Topological Sorting	10	topological sorting

Syllabus Ontology Creation

pandas

NetworkX

gpickle



Output:

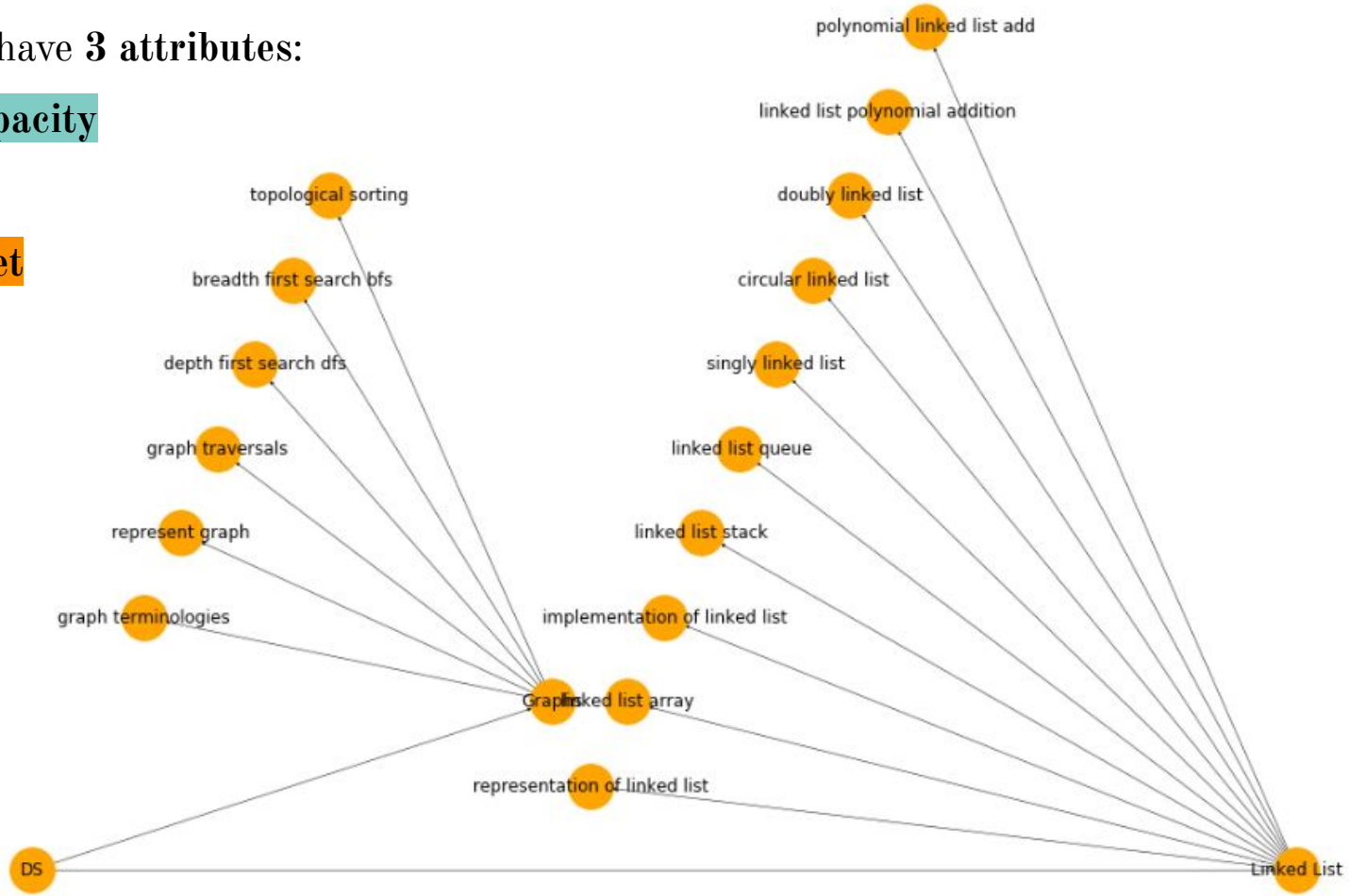
	Module	Topic	Weightage
0	Introduction to Data Structures	[[types data structures], [linear and nonlinea...	7.5
1	Stack and Queues	[[adt of stack], [operations on stack], [array...	25.0
2	Linked List	[[representation of linked list], [linked list...	25.0
3	Trees	[[tree terminologies], [tree traversal], [bina...	30.0
4	Graphs	[[graph terminologies], [represent graph], [gr...	15.0
5	Sorting and Searching	[[bubble sort], [insertion sort], [merge sort]...	17.5

Each node will have **3 attributes**:

Max Marks Capacity

Legal Bucket

Overflow Bucket

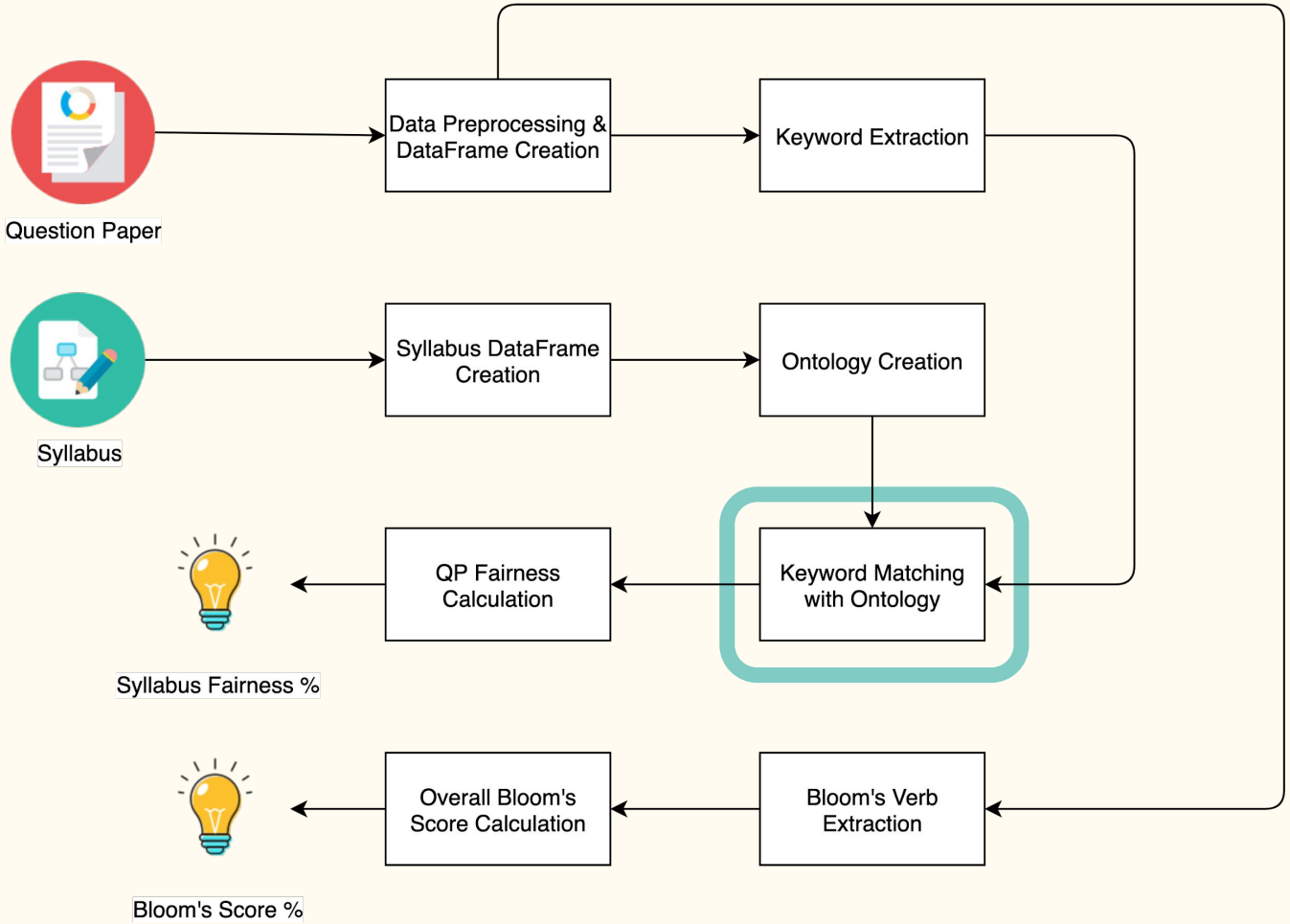


Keyword Matching

NetworkX

SequenceMatcher

FuzzyWuzzy



(circular linked list, 10)

	Num	Question	Marks	Extracted Keywords	Identified Node
0	2a	Write a program to implement Circular Linked L...	10	circular linked	circular linked list
1	4b	Explain different cases for deletion of a node...	10	case deletion node binary search tree function	None
2	5a	Write a program in 'C' to implement Stack usin...	10	c stack linked	linked list stack
3	5b	Explain Depth First search (DFS) Traversal wit...	10	depth first search dfs traversal recursive fun...	depth first search dfs
4	6a	Application of Linked-List –Polynomial addition	10	application linked polynomial addition	linked list polynomial addition
5	6d	Topological Sorting	10	topological sorting	topological sorting

From Question:

(circular linked list, 10)

Node Attributes:

Max Marks Capacity = 5

Legal Bucket = 5

Overflow Bucket = 10 - 5 = 5

$$\frac{TotalLegalMarks}{TotalMarks} * 100$$

$$= \frac{TotalMarksInNodeBucketofSubjectNode}{TotalMarksOfQuestionPaper} * 100$$

Syllabus Fairness Calculation

83.33%

Syllabus Fairness

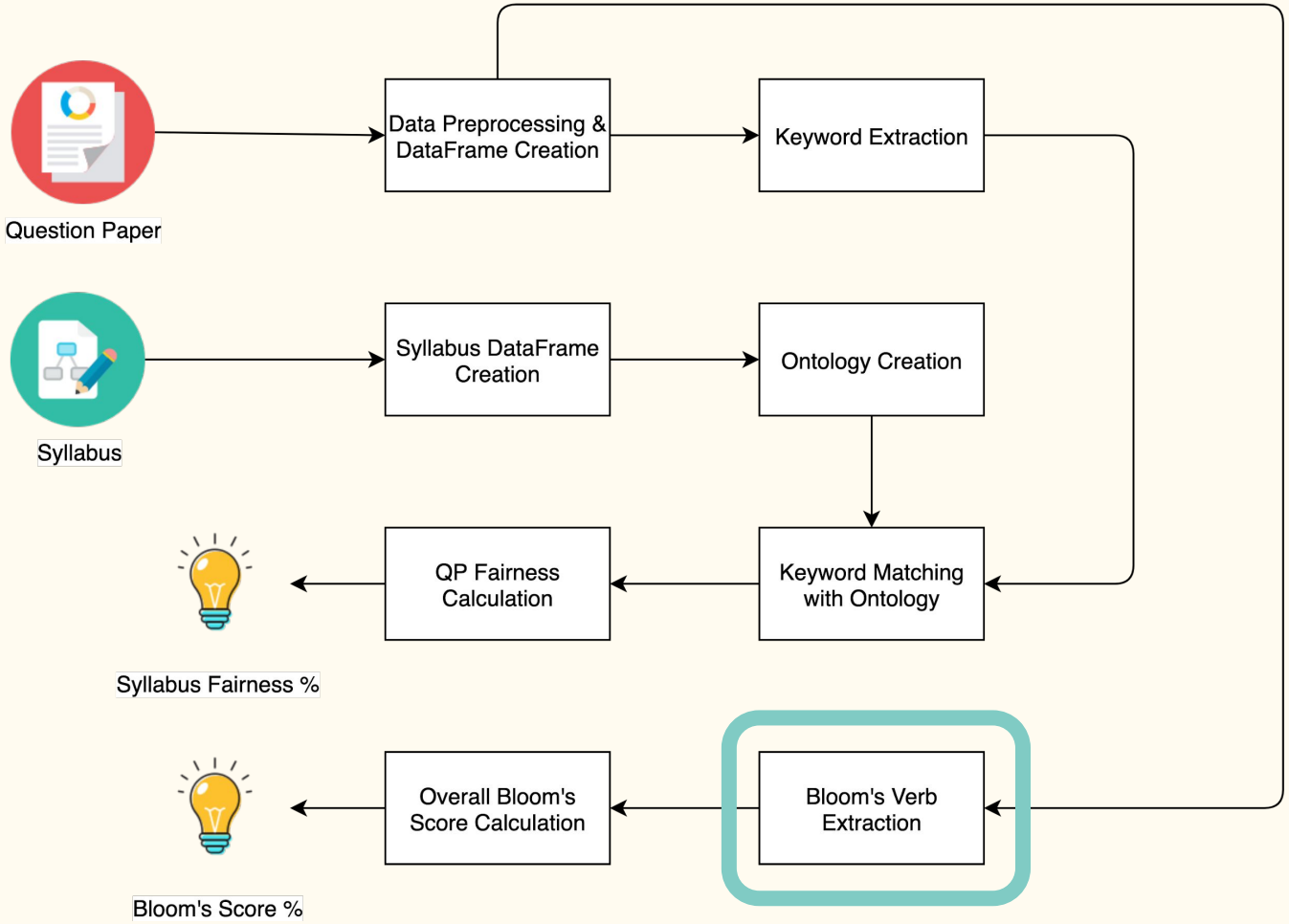
Bloom's verbs extraction

pandas

Python RegEx

nltk tokenizer

nltk tagger



	Num	Question	Marks	Extracted Keywords	Identified Node	Bloom's Verbs
0	2a	Write a program to implement Circular Linked L...	10	circular linked	circular linked list	[(write, remember), (program, create), (list, ...
1	4b	Explain different cases for deletion of a node...	10	case deletion node binary search tree function	None	[(explain, create), (write, remember)]
2	5a	Write a program in 'C' to implement Stack usin...	10	c stack linked	linked list stack	[(write, remember), (program, create), (list, ...
3	5b	Explain Depth First search (DFS) Traversal wit...	10	depth first search dfs traversal recursive fun...	depth first search dfs	[(explain, create), (example, understand), (wr...
4	6a	Application of Linked-List –Polynomial addition	10	application linked polynomial addition	linked list polynomial addition	[(list, remember)]
5	6d	Topological Sorting	10	topological sorting	topological sorting	[]

```
score = 0
For each bloom verb in question: (if no
verb : score += 1/21)

if level == 'remember':
    score += 1/21
if level == 'understand':
    score += 2/21
if level == 'apply':
    score += 3/21
if level == 'analyze':
    score += 4/21
if level == 'evaluate':
    score += 5/21
if level == 'create':
    score += 6/21
```

Each Question will get a score between 0 and 1.

Bloom's Taxonomy Score Calculation for each question

AVG(Bloom’s Scores of all Questions) *100

Final Bloom’s Score for the whole Question Paper

26.98%

Bloom's Taxonomy Score

References

1. Ontology Driven Framework for Assessing the Syllabus Fairness of a Question Paper by Rekha R, A. Angadi, A. Pathak, A. Kapur,
   H. Gosar, Ramanathan M., V. Thatte and M. Sasikumar

2. Keyword Extraction using NLP:
   https://medium.com/analytics-vidhya/automated-keyword-extraction-from-articles-using-nlp-bfd864f41b34

3. NetworkX: https://networkx.github.io/documentation/stable/

4. Python RegEx: https://docs.python.org/3/howto/regex.html

5. Wordnet Corpus: https://www.nltk.org/howto/wordnet.html

6. NLTK: https://pythonprogramming.net/stop-words-nltk-tutorial/

7. Difflib Sequence Matcher: https://docs.python.org/3/library/difflib.html

8. FuzzyWuzzy: https://pypi.org/project/fuzzywuzzy/

9. Bloom's Taxonomy: https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Questions