

# *Platform That Provides Timeline Summary of All Tec- Development*

*Mrs. Indupriya V<sup>1</sup>, Neetu Sharma<sup>2</sup>, Gudiya Singh<sup>3</sup>, Preety Pandey<sup>4</sup>, Arvind Gupta<sup>5</sup>*

<sup>2345</sup>*Student, Department of Computer Science and Engineering, AMCEC, Bangalore- 083,  
Karnataka*

<sup>1</sup>*Faculty, Department of Computer Science and Engineering, AMCEC, Bangalore- 083,  
Karnataka*

## **Abstract**

Staying informed about the latest tech developments is crucial in today's fast-paced world. However, the abundance of information available makes it challenging to filter through and identify the most relevant updates. This project presents TechChronicle, a comprehensive platform designed to address this information overload by providing users with a concise timeline summary and article segregation of all tech developments. TechChronicle employs advanced natural language processing and machine learning techniques to scan and analyze a diverse range of reputable tech sources, including news articles, research papers, industry blogs, and expert forums. The platform's intelligent algorithms extract key information, categorize it, and generate a timeline summary that highlights significant milestones and breakthroughs across various technology domains.

**Keywords:** technology developments, timeline summary, article segregation, natural language processing, machine learning, news aggregation, semantic analysis, topic modeling, community engagement.

---

## **I. INTRODUCTION**

To address this challenge, we introduce a powerful platform that provides a comprehensive solution: a combination of timeline summary and article segregation of all tech developments. This platform aims to empower users by offering a concise overview of the chronological progression of tech advancements across various domains, while also facilitating easy access to in-depth articles for those seeking more detailed information. In addition to the timeline summary, the platform offers powerful article segregation features. Users can explore specific technology categories of interest, allowing them to focus on the domains that align with their professional or personal interests.

Through intelligent algorithms that employ semantic analysis and topic modeling, the platform accurately categorizes and tags articles, making it easy for users to

access relevant and in-depth information tailored to their preferences. Furthermore, the platform fosters community engagement and collaboration. Users can engage in discussions, share insights, and collaborate with fellow tech enthusiasts and professionals through comment sections and forums associated with each article. This collaborative environment encourages knowledge exchange, networking, and the generation of new ideas.

## **II. IMPLEMENTATION OF MODULES**

This stage is the underlying stage in moving from issue to the course of action space. Accordingly, starting with

what is obliged; diagram takes us to work towards how to full fill those requirements. System plot portrays all the critical data structure, record course of action, yield and genuine modules in the structure and their Specification is picked. This assumes an essential part on the grounds that as it will give the last yield on which it was being working. In our work we use four modules, these modules are listed below: -

### 1. Data Collection using Web Scrapping

Web scraping is the method of gathering data from the Web. Indeed replicating and gluing the verses of your favorite melody may be a shape of web scraping! In any case, the words “web scraping” as a rule allude to a prepare that includes computerization. A few websites don’t like it when programmed scrubbers accumulate their information, whereas others don’t intellect.

URLS

1. **The base URL** represents the path to the search functionality of the website. the base URL is <https://www.techtarget.com/>.

2. **The specific site location** that ends with .html is the path to the job description’s unique resource.

Any job posted on this website will use the same base URL. However, the unique resources’ location will be different depending on what specific topic posting we are viewing. URLs can hold more information than just the location of a file. Some websites use **query parameters** to encode values that you submit when performing a search.

### Language Translator

We have used google language translator to check the content’s language. Python module googletrans is used. If the language is other than English the application translate it into English and save. The final English contents are further processed to generate the summery.

### Text Summarization

In the event that you’re scratching a page deferentially for instructive purposes, at that point you’re impossible to have any issues. Still, it’s a great thought to do a few investigate on your possess and make beyond any doubt that you’re not damaging any Terms of Benefit some time recently you begin a large-scale extend.

give the last yield on which it was being working.

In our work we use four modules, these modules are listed below:-

#### 1. Pre-processing of input document:

The phase of pre-processing involves chopping the paragraph into words. This phase involves four stages.

1. Sentence segmentation
2. Tokenization
3. Stop word Removal
4. Stemming

In each stage the document undergoes different changes. The changes are explained below

##### 1.1 Sentence Segmentation of paragraph in the document

Sentence Segmentation is the process of breaking down/segmentation the given text document into sentences. In this system sentence is segmented by identifying the boundary of sentence which ends with period symbol (.), question mark (?), exclamatory mark (!) and the total number of sentences present in the document are also identified.

##### 1.2 Tokenization of segmented sentences

Tokenization is the process of breaking down the sentences into words. Tokenization is done by identifying the spaces ( ), comma (,) and special symbols between the words. In this process frequency of each word is

calculated and stored for further processing.

### 1.3 Stop Word Removal from the list of words

Stop words are the words that do not carry as important meaning as by keywords. These words are identified by supplying a list of words with less importance to the system. The system compares these stop words with the tokenized words obtained from the previous phase. These stop words are then disposed as they can interfere and influence the summary that will be generated at the end.

### 1.4 Stemming

A word can be found in different forms in the same document. These words have to be converted to their root form for simplicity. This process is known as Stemming. An algorithm is used to transform words to their root forms. In this system, Porter's stemmer method is used to turn a word into its root form using a predefined suffix list. Finally, the frequency of each word is calculated and retained for the next phase.

## 2. Sentence scoring

After module 1 the input document is segmented into a collection of words in which each word has its individual frequency. In module 2 the sentences are ranked based on seven important features:

1. Frequency
2. Sentence Position
3. Cue words
4. Similarity with the Title.
5. Sentence length.
6. Proper noun.
7. Sentence reduction.

### 2.1 Frequency

Frequency is the number of times a word occurs in a document. If a word's frequency in a document is high, then it can be said that this word has a significant effect on the content of the document. Salient sentences/words are those sentences/words that occur repeatedly. The frequently occurring word increases the score of sentences they are in. The most common measure widely used to calculate the word frequency is TF (Term frequency) IDF (Inverse document frequency). The total frequency value of a sentence is calculated by summing up the frequency of every word in the document.

### 2.2 Sentence Position Value

It depends on our requirement whether important sentences are located at certain positions in text or in a paragraph. Sentences in the beginning define the theme of the document whereas sentences in the end conclude or summarize the document.

The positional value of a sentence is calculated by assigning the highest score value to the first sentence and the last sentence of the document. The second highest score value is assigned to the second sentence from starting and the second last sentence of the document. Remaining sentences are assigned a score value of zero.

### 2.3 Cue Words

Cue words are the important words in a document. These cue words are given as input from the user. If a sentence contains these cue words then a score value of one is assigned to the sentence, otherwise the score value of the sentence will be zero.

### 2.4 Similarity with the Title

The words in the title and heading of a document that reappear in sentences are directly related to summarization. These words are considered for summarization as they have some extra weight in them. If a sentence contains words in the title and header then a score value of one is assigned to that sentence, otherwise the score value is zero for the sentence.

### **2.5 Sentence length**

The length of the sentence resembles the importance of sentence in summarization. Generally, sentences that are very long and very short are not suitable for summary. Sentences that are very long will have unnecessary information which is not useful for summarization of document. Whereas, sentences that are too short do not give much of information about the document.

### **2.6 Proper Noun**

Proper nouns play an important role in summarization. It gives information regarding, to whom or to what the author is referring. Roles played by individuals or locations description will be different more number of times in a document.

### **2.7 Sentence reduction**

Sentence reduction is the method of removing irrelevant phrases like prepositional phrases, clauses, to infinitives, or gerunds from sentences. The goal is to identify less important phrases in a sentence using reduction decisions. The reduction decisions are based on syntactic knowledge, context, and probabilities computed from corpus analysis.

The final score is a Linear Combination of frequency, Sentence positional value, Cue Words, Similarity with the title of the document, Sentence length and Proper noun.

### **3. Sentence Ranking**

After each sentence is scored they are arranged in descending order of their score value i.e. the sentence whose score value is highest is in top position and the sentence whose score value is lowest is in bottom position

### **4. Summary Extraction**

After ranking the sentences based on their total score the summary is produced selecting certain number of top ranked sentences where the number of sentences required is provided by the user. For the reader's convenience, the selected sentences in the summary are reordered according to their original positions in the document.

### **3. Testing:**

Testing is an important phase in the development life cycle of the product. This is the phase, where the remaining errors, if any, from all the phases are detected. Hence testing performs a very critical role for quality assurance and ensuring the reliability of the software. During the testing, the program to be tested was executed with a set of test cases and the output of the program for the test cases was evaluated to determine whether the program was performing as expected. Errors were found and corrected by using the below stated testing steps and correction was recorded for future references. Thus, a series of testing was performed on the system, before it was ready for implementation. It is the process used to help identify the correctness, completeness, security, and quality of developed computer software. Testing is a process of technical investigation, performed on behalf of stake holders, i.e. intended to reveal the quality-related information about the product with respect to context in which it is intended to operate. This includes, but is not limited to, the process of executing a program or application with the intent of finding errors. The quality is not an absolute; it is value to some person. With that in mind, testing can never completely establish the correctness of arbitrary computer software; Testing furnishes a 'criticism' or comparison that compares the state and behaviour of the product against specification. An important point is that software testing should be distinguished from the separate discipline of Software Quality Assurance (SQA), which encompasses all business process areas, not just testing.

There are many approaches to software testing, but effective testing of complex products is essentially a process of investigation not merely a matter of creating and following routine procedure. Although most of the intellectual processes of testing are nearly identical to that of review or inspection, the word testing is connoted to mean the dynamic analysis of the product-putting the product through its paces. Some of the common quality attributes include capability, reliability, efficiency, portability, maintainability, compatibility and usability. A good test is sometimes described as one, which reveals an error; however, more recent thinking suggest that a good test is one which reveals information of interest to someone who matters within the project community.

### 3.1 Types of Testing

#### 3.1.1 Unit Testing

Individual component are tested to ensure that they operate correctly. Each component is tested independently, without other system component. This system was tested with the set of proper test data for each module and the results were checked with the expected output. Unit testing focuses on verification effort on the smallest unit of the software design module. This is also known as **MODULE TESTING**. This testing is carried out during phases, each module is found to be working satisfactory as regards to the expected output from the module.

#### 3.1.2 Integration Testing

Integration testing is another aspect of testing that is generally done in order to uncover errors associated with flow of data across interfaces. The unit-tested modules are grouped together and tested in small segment, which make it easier to isolate and correct errors. This approach is continued unit I have integrated all modules to form the system as a whole.

#### 3.1.3 System Testing

System testing is actually a series of different tests whose primary purpose is to fully exercise the computer-based system. System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration testing. System testing is based on process description and flows, emphasizing pre-driver process and integration points.

#### 3.1.4 Performance Testing

The performance testing ensure that the output being produced within the time limits and time taken for the system compiling, giving response to the users and request being send to the system in order to retrieve the results.

#### 3.1.5 Validation Testing

The validation testing can be defined in many ways, but a simple definition is that. Validation succeeds when the software functions in a manner that can be reasonably expected by the end user.

##### **Black Box testing**

Black box testing is done to find the following

- ✓ Incorrect or missing functions
- ✓ Interface errors
- ✓ Errors on external database access
- ✓ Performance error
- ✓ Initialization and termination error

##### **White Box Testing**

This allows the tests to

- ✓ Check whether all independent paths within a module have been exercised at least once
- ✓ Exercise all logical decisions on their false sides
- ✓ Execute all loops and their boundaries and within their boundaries
- ✓ Exercise the internal data structure to ensure their validity
- ✓ Ensure whether all possible validity checks and validity lookups have been provided to validate data entry.

#### 3.1.6 Acceptance Testing

This is the final stage of testing process before the system is accepted for operational use.

The system is tested within the data supplied from the system procurer rather than simulated data

### III. ARCHITECTURE DESIGN

The architectural configuration procedure is concerned with building up a fundamental basic system for a framework. It includes recognizing the real parts of the framework and interchanges between these segments. The beginning configuration procedure of recognizing these subsystems and building up a structure for subsystem control and correspondence is called construction modeling outline and the yield of this outline procedure is a portrayal of the product structural planning. The proposed architecture for this system is given below. It shows the way this system is designed and brief working of the system.



### IV. CONCLUSION

In conclusion, the platform we introduce offers a comprehensive solution for staying informed about the latest tech developments. By providing a timeline summary that captures the major milestones and breakthroughs, as well as offering article segregation and in-depth summaries, users can navigate the vast technology landscape with ease. With community engagement features, the platform also facilitates collaboration and knowledge sharing among tech enthusiasts and professionals

### REFERENCES

- I. Visualizing literature review theme evolution on timeline maps: comparison across disciplines Chunlei Tang, Yangyong Zhu, Suzanne V Blackley, Joseph M Plasek, Meihan Wan, Li Zhou, Jing Ma, David W Bates *IEEE Access* 7, 90597-90607, 2019
- II. 2021 quantum threat timeline report Michele Mosca, Marco Piani *Global Risk Institute, Toronto, ON, 2022*
- III. TimeLine: visualizing integrated patient records Alex AT Bui, Denise R Aberle, Hooshang Kangarloo *IEEE Transactions on Information Technology in Biomedicine* 11 (4), 462-473, 2007
- IV. On summarization and timeline generation for evolutionary tweet streams Zhenhua Wang, Lidan Shou, Ke Chen, Gang Chen, Sharad Mehrotra *IEEE Transactions on Knowledge and Data Engineering* 27 (5), 1301-1315, 2014