

# Implementation Paper on Visual Education using Data Mining and Innovative Visualization on Cloud

Suraj Gaikwad

IT Department

Dr. D.Y. Patil Institute of  
Engineering and Technology,  
Ambi, Pune

Kunal Pokale

IT Department

Dr. D.Y. Patil Institute of  
Engineering and Technology,  
Ambi, Pune

Anirban Dutta

IT Department

Dr. D.Y. Patil Institute of  
Engineering and Technology,  
Ambi, Pune

## ABSTRACT

Kinesthetic, Auditory, Reading / Writing - Preference Learning and Visual Learning are the 4 categories of learning types as per the widely used Fleming VAK / VARK model. Visual Learning represents concepts visually. By representing the information visually students are able to concentrate on meaning, acknowledge and group similar ideas easily making better use of visual memory. Visual learning improves student performance in the areas of Retention, Reading Comprehension, Student Achievement and Critical Learning skills. This system focuses on semi-automatically converting textual learning resources related to important world events into visual information. Few examples include Terrorist Activities in India, 26/11 Mumbai bomb blasts and Earth Quakes in India. This system scrapes the unstructured text from textual learning resources, applies natural language processing to extract named entities in the text, uses location api's and innovative visualizations (e.g. Google Maps) to visualize the world event information in an intuitive manner. Human intervention is needed in case of incomplete information in the learning resource and verification of generated information. This system makes it easy to identify human intervention need via data health reports.

## General Terms

Fleming VAK / VARK model, Visual Learning Data mining, NLP(Natural Language Processing), visualization, cloud.

## Keywords

Web scraping, rule engine, curate engine, NNP.

## 1. INTRODUCTION

Kinesthetic, Auditory, Reading / Writing - Preference Learning and Visual Learning are the 4 categories of learning types as per the widely used Fleming VAK / VARK model. Visual Learning represents concepts visually. By representing the information visually students are able to concentrate on meaning, acknowledge and group similar ideas easily making better use of visual memory. Visual learning improves student performance in the areas of Retention, Reading Comprehension, Student Achievement and Critical Learning skills.

Students learn about historical / current events using books / websites. Books and websites both contain information written on pages. Pages contain a bunch of text and dates. Visualization is relatively less used as compared to the text. Learning from such sources is tedious and time consuming. Events occur on maps and not on pages. Thus events are more intuitively visualized on maps. Visual learning provides a complete picture at one go.

This system focuses on semi-automatically converting textual learning resources related to important world events into visual information. Few examples include Terrorist Activities in India, 26/11 Mumbai bomb blasts and Earth Quakes in India. In this system scrapes the unstructured text from textual learning resources, applies natural language processing to extract named entities in the text, uses location api's and innovative visualizations (e.g. Google Maps) to visualize the world event information in an intuitive manner. Human intervention is needed in case of incomplete information in the learning resource and verification of generated information. This system makes it easy to identify human intervention need via data health reports. Visualizations generated by our system are live on web by hit this url <http://www.visualization113.appspot.com/>.

## 2. PROJECT PURPOSE

The purpose of our project is to semi automatically convert textual learning resources related to world events to Visual learning resources. E.g. 26/11 Mumbai attacks on Wikipedia to Visualization on Google Maps.

## 3. OUR SOLUTION

### 3.1 Existing Problem

Students learn about historical / current world events using books / web pages. Books contain information that is laid out using pages. Pages contain a bunch of text and dates. Figures are sparingly used compared to the text.

World events cannot be visualized on pages. They are intuitively better visualized on maps. Below are some examples of the same.

1. Chronology of an event e.g. Visualizing 26/11 terrorist attacks on Mumbai.
2. All bomb blasts that happened in India.

Dynamic searching of information is not possible on books ex., given a place find all events that happened in that place.

### 3.2 Solution

We use data mining to aggregate data related to an event in an automated way. We then curate this data. We use intuitive visualization techniques to visualize the historical event data.

1. History happened on maps. Hence we show chronology of the event using Google maps.
2. Animated images showing blasts, war etc.

### 3.3 Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into needed information. Data mining software is one of a number of analytical tools for classifying data. It allows users to analyze data from many different resources, separated it, and summarize the relationships identified.

In this system web scraping is used to mine data from web pages.

### 3.4 Web Scraping

Web Scraping refers to an application that processes the HTML of a Web Page to extract data for manipulation such as converting the Web page to another format (i.e. HTML to WML). Web Scraping scripts and applications will simulate a person viewing a Web site with a web browser. With these scripts you can interconnect to a Web page and request a web page, correctly as a browser would do. The Web server will send back the web page in which you can then manipulate or extract specific information.

### 3.5 Natural Language Processing (NLP)

Natural language processing (NLP) is a field of computer field, linguistics and artificial intelligence concerned with the interactions between computers and human languages. As such, natural language processing (NLP) is related to the part of human-computer interaction. Many challenges in NLP involve natural language understanding, that is, permissive computers to derive meaning from natural or human language input, and any others involve natural (human) language generation.

In this system we NLP is used to extract named entities from the unstructured text e.g. place where event occurred.

### 3.6 Data Enhancement

This system enhances available data to retrieve new information. The extracted place named entity is used to extract location data. Google Places API is used for the same.

### 3.7 Data Curate

Source unstructured text might have some missing information e.g. date of the event is not available, information about the event may not be complete. In such cases This system automatically detects anomalies in the data that is mined. This system then prompts the admin to fix the data manually.

## 4. VISUALIZATION ON GOOGLE MAPS

- Automated data mining, natural language processing working along with manual curation of data.
- Events containing large number of sub-events can be seen in animated way on google maps.
- Animated images showing blasts, earthquakes etc.
- Example:



Fig 1: Snapshot of system

## 5. ALGORITHM ANALYSIS

- Read unstructured text from url.
- Break data into sentences.
- Break each sentence into noun, verb, adjectives.
- Selected required sentence.
- Convert unstructured sentence into structured data and store in database.
- Manually curate the data.
- Visualize the data.

## 6. MATHEMATICAL MODEL

In This system can be modeled mathematically using set theory as follows:

Let S be a system.

$$S = \{I, F, O\}$$

where I is a set of inputs, F is a set of functions and O is set of Outputs

**Input:**

$$I = \{Q_i\}$$

Q is a set of web pages containing unstructured / semi structured text

**Functions:**

$$F = \{F_1, F_2, F_3\}$$

F<sub>1</sub> = Web scrape text

F<sub>2</sub> = Extract named entities

F<sub>3</sub> = Get location information

**Output:**

O = Structured data visualized using innovative charts e.g. Google Maps.

## 7. PROPOSED SYSTEM

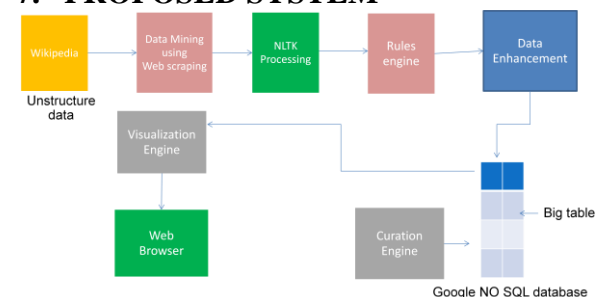


Fig 2: Architecture of system

## 7.1 Description

- i. Web scraping engine extracts unstructured / semi structured text from Wikipedia.
- ii. NLP engine extracts named entities from the text. (e.g. NNP)
- iii. Rules engine ignores unimportant words.
- iv. Data enhancement engine gets location information for the named entity.
- v. Curate engine identifies anomalies in the data mined and notifies the admin to manually correct the data.
- vi. Visualization is done using innovative charts e.g. Google maps.

## 8. RESULTS

This system successfully extracts unstructured / semi structured text from web pages, extracts named entities from the text, enhances data for the named entities and visualizes the same on innovative charts e.g. google maps. In case information is missing in the source text this system automatically identifies anomalies and reports them to the admin.

## 9. ADVANTAGES

- Visual learning.
- Any world event visualize on google map.
- Visualize the world event information in an intuitive manner.
- This systems scrapes the unstructured text from textual learning resources.

## 10. ACKNOWLEDGEMENT

We take this opportunity to thank our project guide Prof. Anirban Dutta and Head of the Department Prof. Amol Jadhav for their valuable guidance and for providing all the necessary facilities, which were intrinsic in the completion of this project paper. We are also thankful to all the staff members of the Department of Information Technology of Dr. D. Y. Patil Institute of Engineering and Technology for their estimable time, comments, support, suggestions and inveiglement. We would also like to thank the institute for

providing the required facilities, Internet surfing and important books/resources.

## 11. CONCLUSION AND FUTURE WORK

- I. This system will help viewers to better educate themselves on the events that happened in the world.
- II. This system will help in better recollection of events.
- III. This system could further be used adequately and independently for different purpose .

### 11.1 Future Work

This system can be easily used to extract data for more world events to increase the scope of visual learning in our application.

In the current implementation NLP processing has been done on smaller textual data. In future this work can be extended to parse complex textual information to get relationships in them e.g. with the use of Stanford Parsers.

## 12. REFERENCE

- [1] Building Domain Ontologies from Text for Educational Purposes Amal Zouaq and Roger Nkambou, Member, IEEE
- [2] Visualizing Dynamic Data with Maps Daisuke Mashima, Stephen G. Kobourov, and Yifan Hu
- [3] <http://www2.research.att.com/yifanhu/TrendMap/>, 2012. Reconstructing the Structure of the World-Wide Music Scene [withlast.fm. http://sixdegrees.hu/last.fm/index.html](http://sixdegrees.hu/last.fm/index.html), 2012.
- [4] U. Brandes and S.R. Corman, "Visual Unrolling of Network Evolution and the Analysis of Dynamic Discourse," Proc. IEEE Symp. Information Visualization (INFOVIS '02), pp. 145-151,2002
- [5] Y. Agichtein and S. Gravano, "Snowball: Extracting Relations from Large Plain-Text Collections,"
- [6] H. Alani and C. Brewster, "Metrics for Ranking Ontologies,"