

EIONS: System for Detecting, Tracking & Visualization

Smriti Sharma

Lecturer, Computer science & engineering, BSAITM, Haryana, India.

Abstract--This paper presents a simple extraction system for extraction & visualization of a range of content. EIONS will present specific information for future reference too. This paper aims to 'mine' (or extract) crime information from online newspaper articles and makes this information available to the public. Paper presents various algorithms for detecting & tracking of news events. Information overload is a main challenge in world of extraction these days. EIONS apply techniques from text mining, automated machine learning & statistical analysis can help to reduce this overload of information & than system will visualize it. This paper presents a more detailed view for working of EIONS, which presents concise & compact information & save it for future reference. This system has advantages over other systems developed before. A very simple 5W1H approach is used for detecting & tracking of news events & than save it for future reference. This paper also presents a good way to find intelligence by adding one more step & so concludes relevant information.

This paper is presented at International Conference on Recent Trends in Computer and information Technology Research on 25th & 26th September (2015) conducted by B. S. Anangpuria Institute of Technology & Management, Village-Alampur, Ballabgarh-Sohna Road, Faridabad.

1. INTRODUCTION

EIONS presents a lightweight approach for extraction & visualization of news articles. Previous paper we describe & using approach of 5W1H [12]. This approach comprises a key event identification step, an event semantic elements extraction step and an event ontology population step. We first use a machine learning method to identify the key events from news stories. Then algorithms are applied to extract content. Visualization means mapping the articles on a timeline and location map. It is difficult to follow the entirety of the global of news sources, and the events happening every day. If an analyst in her area has to follow and map all these according to the timeline they happen, the task quickly becomes overwhelming. As we see in cluster centric approach where NEXUX system is used news can be mapped by live event tracking.

We presented a tool which attempts to ease the task of finding all news articles about an event, and mapping them without using Google earth.[7]. This paper presents the system for work done in [7] used for providing information & storing it for further references.

2 RELATED WORK

This paper is detailed description of paper presented [4]. Paper fits in the areas of data visualization [3] and in particular in the areas of text visualization [6] and more recently, developments of semantic web and visualization of ontologies and other knowledge structures [1]. Approach used for detecting & tracking of news events we used News Event Extraction Using 5W1H Approach & Its Analysis [12]. Most prominent is the overview publication from MITRE team [7] giving good overview over the approaches for visualization of different document types, including news stories. Their goals are similar to the work presented here, but the actual approach is quite different. Their publication appeared also at [8] together with some other interesting approaches for document visualization. Another approach for visualizing trends in news documents is the system ThemeRiver [9] developed at Pacific Northwest National Laboratory together with many other interesting approaches for information text visualization [10]. ThemeRiver in particular is specialized for analyzing and visualizing trends in news stories over time, enabling efficient detection of trends in the vocabulary used in the texts. Among others, we would also like to mention work of the authors on visualization of large text corpora [11] presented at "TELRI -Information in Corpora" workshop which directly precedes this work. Main problem with older methods were their complexity. ACE by using RSS [2] will lead to information loss. We presented EIONS [7] system so that it can overcome all drawbacks.

3. EIONS : EXTRACTING INTELLIGENCE FROM ONLINE NEWS SOURCES

EIONS represents an approach of extracting news events along with its visualization. Intelligence means "detecting,

tracking & visualization". So objective is not only extracting the news events occurred but to visualize it.

System architecture for EIONS is given in figure consists of different parts. Collecting news stories from online sources, applying our event extraction mechanism, collecting facts according to their nature of existence, finding extracted information in a database & then performing its visualization. These all steps make it possible to detect track & visualize our extraction process. System is explained in four steps as shown in fig :

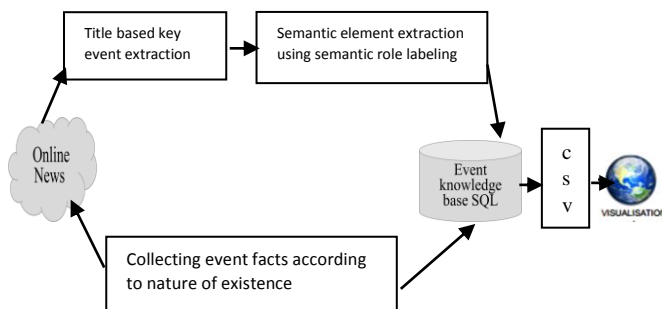


Fig 1: System architecture

3.1 Title based key event extraction

A key event identification step which finds topic sentences that contain key event by measuring a sentence's importance. The main idea of it is to use surface and semantic characteristics of a news story to identify a most important sentence which has the highest possibility to describe the key event of the story.

Steps

- 1) NLP processing.
- 2) Classify title as informative & non-informative.
- 3) Extract topic words to be searched. Make them triggers.
- 4) Check similarity between triggers & title for which information is to be extracted.
- 5) For each sentence do
 - Check how much frequently any trigger is used.
 - Position of trigger in sentence.
 - How much of sentence length is used.
 - Apply named entity recognizer.
 - Sentence & title similarity.

3.2. Event Semantic Element Extraction

The second step is an event semantic elements extraction step which extracts 5W1H elements from the identified topic

sentences using SRL (Semantic Role Labeling) and NER (Name Entity Recognition).steps are as follows:

Begin

For each topic sentence do

- 1) Named Entity Recognition.
- 2) POS tagging.
- 3) Event identification and classification; //WHAT
- 4) Syntactic-semantic rules-based <Subject, Predicate, Object> recognition; //WHO did WHAT to WHOM
- 5) Time & date identification; //WHEN
- 6) Location identification; //WHERE
- 7) Topic sentences as short summarization; //HOW

End do

End

3.3 . Collecting Event Facts:

This step defines concepts of entities (time, person, location, organization etc.), events and relationships to capture temporal, spatial, information, experiential, structural and causal aspects of events.

3.4. Visualization

Visualization means mapping the articles on a time-line and location map. Text visualization is an area having the main goal to present textual contents of one or many documents in a visual form. The intention of producing visualization of the textual contents is mainly to create graphical form of content summary on different levels of abstraction. The documents are preprocessed in two different ways. First, the text is cleaned and bag- of- words representation is created, and next, the name - entities are extracted. All the documents are stored in three different representations (as already described: plain text, bag -of -words and name - entities) in the database which is then used by the client software using efficient graphical user interface described in the following sections.

We find a database table of output of event extracted [pr2]. This system is yet to be implemented so Here we used a very simple approach of mapping events using ArcGIS explorer. Database table extracted after news event extraction using 5W1H approach which should be in Microsoft Excel CSV format, which is comma separated values file. . ArcGIS Explorer Online open a map, add other content to it, navigate around it, ask questions the map can answer, and present and share the map with others.

3. EXPERIMENTS AND EVALUATION

EIONS system proposed is an efficient easy to implement news extraction system. There system is divided in major two

parts to see how it will work. First part will take input as a sentence, headline or an article of crime information classified. This input while processed using 5W1H approach for semantic element extraction will lead to information that we needed. Output of semantic element extraction will be stored in database further for visualization of information on map [4].

Example:

- ♦ Car bomb kills 50 in Iraq
HindustanTimes Wednesday, June 18, 2008 5:07:00 AM CEST
A car bomb blast in northern Baghdad left more than 50 people dead and 80 wounded on Tuesday, a police source said...
- ♦ Biggest blast in months leaves at least 50 dead in Iraq
reliefWeb Wednesday, June 18, 2008 5:05:00 AM CEST
A car bomb blast in northern Baghdad, the largest in months, left more than 50 people dead and 80 wounded on Tuesday, a police source said...

For a given specific event, there can be lots & lots of articles on the web. Each & every article will contribute in the information for the event .So information extracted for a particular event given above will be as follows:

• WHEN : 18 June 2008
• WHERE: Baghdad, Iraq
• WHAT: terrorist attack
• WHOM: Number killed-50 ,Number wounded-80,Number kidnapped- 0
• WHY: not reported

TABLE-1: EXTRACTED EVENTS WITH SPECIFIC INFORMATION

Date	Address	Incident	Fatalities	Injured	Status of case
------	---------	----------	------------	---------	----------------

7/7/1987	haryana	1987 Punjab killings	34		
15/6/1991	Ludhiana district, Punjab	1991 Punjab killings	88		
12/3/1993	Mumbai, Maharashtra	1993 Bombay bombings	257	713	verdict given
30/12/1996	Western Assam	Brahmaputra Mail train bombing	33		
14/2/1998	Coimbatore, Tamil Nadu	1998 Coimbatore bombings	58	200+	verdict given
22/12/2000	CP, Delhi	2000 terrorist attack on Red Fort			verdict given
24/9/2002	Gandhinagar, Gujarat	Terrorists attack the Akshardham temple in Gujarat	31		

Geographical Information System is among today's fast developing technologies and is being integrated with various other computer applications. This table is obtained after extraction process .Now we use ArcGIS maps. ArcGIS Explorer Online is an online application that lets you explore and present maps within an efficient and well-structured environment. Maps show you where things are, they tell you what they are and help you understand why they are that way.

ArcGIS Explorer Online, helps you to show that algorithms given by you can extract the content and "Can that content be visualized on the map?". When a more detailed work is not done on visualization part, it lets you open a map, add other content to it, navigate around it, ask questions the map can answer, and present and share the map with others. The

visualization of news events extracted in table 1 is shown in fig-7.

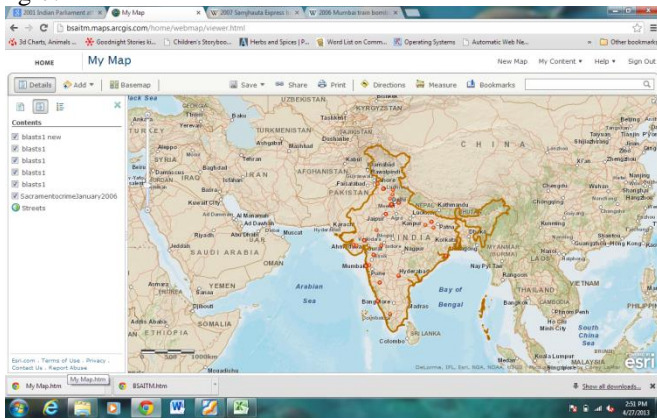


Fig-8: Blasts information extracted in table-1 on map

Here we can specifically see the information on map of nay event. There is no need to see output table of extraction. All information will be visualized.

4. CONCLUSION

Paper presents a more detailed view of EIONS [4]. This paper presents an approach of event tracking n visualization. Algorithms propose a fine extraction of articles and content and then visualization is to be done. a more detailed work should be done on visualization part of system presented.

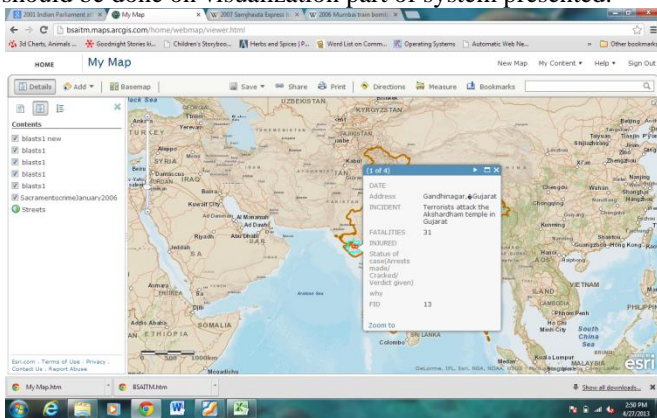


Fig-8: Gandhinagar, Gujrat blast information on map

Vagueness common in spatial descriptions, such as north of the city or near the border, is a problem that must be addressed in geographic information systems in order to successfully represent text-based events. future work this vagueness should be investigated.

REFERENCES:

- [1] Geroimenko, V., Chen, C. (ed): Visualizing the Semantic Web. Springer Verlag, (2003).
- [2] Hao Han, Tomoya Noro, Takehiro tokuda.: An Automatic Web News Article Contents Extraction System Based on RSS Feeds.(2009)
- [3] Fayyad, U., Grinstein, G., Wierse, A.: Information Visualization in Data Mining and Knowledge Discovery. Morgan Kaufmann (2001).
- [4] Smriti Sharma, Rajesh Singh, Pawan Bhadana: Extracting Intelligence from Online News Sources (2013).
- [5] Jurafsky, D., Martin. J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall (2000).
- [6] Chen, C.: Visualization of Knowledge Structures, In "Handbook of Software Engineering and Knowledge Engineering", World Scientific Publishing (2002).
- [7] Chase, P., D'Amore, R., Gershon, N., Holland, R., Hyland, R., Mani, I., Maybury, M., Merlino, A., Rayson J. :Semantic Visualization. ACL- COLING Workshop on Content Visualization and Intermedia Representation.
- [8] Content Visualization and Intermedia Representations (CVIR'98), <http://acl.ldc.upenn.edu/W/W98/>
- [9] Havre, S., Hetzler, E., Whitney, P., Nowell, L.: ThemeRiver: Visualizing Thematic Changes in Large Document Collections. IEEE Transactions on Visualization and Comp. Graphics, V8, No.1, 2002.
- [10] Pacific Northwest National Laboratory, Information Visualization, <http://www.pnl.gov/infoviz/>
- [11] Grobelnik, M., Mladenic, D.: Efficient visualization of large text corpora. 7th TELRI, Info.in Corpora (2002)
- [12] Smriti Sharma, Rajesh Kumar, Pawan Bhadana, News Event Extraction Using 5W1H Approach & Its Analysis, IJRET.

Authors



Ms Smriti Sharma is B.Tech.(CSE), MBA (HRM), & M.Tech(CSE). She has 7 yrs of experience in field of education. 11 research papers have already been published out of 6 papers are published in various international journals and 5 are presented and published in national conferences.