The 5W Structure for Sentiment Summarization-Visualization-Tracking

Amitava Das[1], Sivaji Bandyopadhyay[2], and Björn Gambäck[1]

[1] Department of Computer and Information Science (IDI)
Norwegian University of Science and Technology (NTNU), Trondheim, Norway
Sem Sælands Vei 7-9, NO - 7491, Trondheim, Norway
[2] Department of Computer and Engineering
Jadavpur University
Kolkata-700032, India
`amiava.santu@gmail.com, sivaji_cse_ju@yahoo.com,`
`gamback@idi.ntnu.no`

Abstract. In this paper we address the Sentiment Analysis problem from the end user's perspective. An end user might desire an automated at-a-glance presentation of the main points made in a single review or how opinion changes time to time over multiple documents. To meet the requirement we propose a relatively generic opinion 5Ws structurization, further used for *textual and visual summary and tracking*. The 5W task seeks to extract the semantic constituents in a natural language sentence by distilling it into the answers to the **5W** questions: **Who**, **What**, **When**, **Where** and **Why**. The visualization system facilitates users to generate sentiment tracking with textual summary and sentiment polarity wise graph based on any dimension or combination of dimensions as they want i.e. "*Who*" are the actors and "*What*" are their sentiment regarding any topic, changes in sentiment during "*When*" and "*Where*" and the reasons for change in sentiment as "*Why*".

Keywords: 5W Sentiment Structurization, Sentiment Summarization, Sentiment Visualization and Sentiment Tracking.

1 What Previous Studies Suggest, Opinion Summary: Topic-Wise, Polarity-Wise or other-Wise?

Aggregation of information is the necessity from the end user's perspective but it is nearly impossible to make consensus about the output format or how the data should be aggregated. Researchers tried with various types of output format like textual or visual summary or overall tracking with time dimension. The next key issue is "***How the data should be aggregated?***" and "***What is the End User's requirement?***". Dasgupta and Ng [1] throw an important question: "***Topic-wise, Sentiment-wise, or Otherwise?***" about the opinion summary generation techniques. Instead of digging for the answer of the unresolved debate we experimented with multiple outputs formats. At first we will look into the topic-wise, polarity-wise and other-wise summarization

systems proposed by various previous researchers and then will describe the systems developed by us.

Topic-Wise: There is clearly a tight connection between extraction of topic-based information from a single document and topic-based summarization of that document, since the information that is pulled out can serve as a summary; see [2] for a brief review (Section 5.1). Obviously, this connection between extraction and summarization holds in the case of sentiment-based summarization, as well. There are various topic-opinion [4], [5] summarization systems, proposed by the previous researchers. Leveraging existing topic-based technologies is the most common practice for sentiment summarization. One line of practice is to adapt existing topic-based multi-document summarization algorithms to the sentiment setting. Sometimes the adaptation consists simply of modifying [3], [6] the input to these pre-existing algorithms.

Polarity-Wise: Indeed the topic-opinion model is the most popular one but there could be a requirement at the end user's perspective that they might look into an at-a-glance presentation of opinion-oriented summary. For example: One market surveyor from company *A* might be interested in the root cause for why their product *X* (suppose camera) become less popular day by day. And for this particular case *A* may wants look into for the *negative* reviews only. Therefore opinion-oriented summary is the end user's requirement here. Relatively a few research efforts could be found on the polarity-wise summarization in the literature than the popular topic-opinion model. There are a few important related works [8], [9] which are significant in both the aspects: problem definition and solution architecture with best of our knowledge.

Visualization: To convey all the automatically extracted knowledge to the end user concisely the graphical or visualized output format is one of the trusted and well acceptable methods. Thus a numbers of researcher tried to leverage the existing or newly developed graphical visualization methods for the opinion summary presentation. Some noteworthy related previous works on opinion summary visualization techniques are by Gamon et al., [11], Yi and Niblack, [12], Carenini et al. [14][1] and [15].

Tracking: In many applications, analysts and other users are interested in tracking changes in sentiment about a product, political candidate, company or other issues over time. The tracking system could be a good measure to understand the people's sentiment changes or it could be helpful sociological survey also. In general sense tracking means plotting of sentiment values over time into a graphical visualization. Some significant research efforts on opinion tracking are Lydia[2] project (also called TextMap) [16], Ku et al., [17] Mishne and Rijke, [18] and Fukuhara et al., [19].

2 The Proposed 5W Rationalism

We mentioned a few (Due to space complexity) of noteworthy related works in this section. During the literature survey we realized that there is no consensus among the

---

[1] http://www.cs.ubc.ca/~carenini/storage/SEA/demo.html

[2] http://www.textmap.com/

researchers could be found on the output format of any sentiment summarization system.

Instead of digging for the answer of the unresolved debate we experimented with multiple output formats: multi-document topic-opinion textual summary but realizing the end user's requirement and to less their effort and to present an *at-a-glance* representation we devise a 5W constituent based textual summarization-visualization-tracking system. The 5W constituent based summarization system is a multi-genre system. The system facilitates users to generate sentiment tracking with textual summary and sentiment polarity wise graph based on any dimension or combination of dimensions as they want i.e. "*Who*" are the actors and "*What*" are their sentiment regarding any topic, changes in sentiment during "*When*" and "*Where*" and the reasons for change in sentiment as "*Why*". During the related work discussion we categorize the previous systems in "*Topic-Wise*", "*Polarity-Wise*" or "*Other-Wise*" genres. In the "*Other-Wise*" genre we described the necessity of the visualization and tracking systems. As par our understanding the 5W constituent based summarization system fall into every genre and the supportive argumentations from our side are as follows:

Topic-Wise: The 5W system facilitates users to generate sentiment summary based on any customized topic like Who, What, When, Where and Why and based on any dimension or combination of dimensions as they want.

Polarity-Wise: The system produces an overall gnat chart, could be treated as an overall polarity wise summary. An interested user can still look into the summary text to find out more details.

Visualization and Tracking: The visualization facilitates users to generate visual sentiment tracking with polarity wise graph based on any dimension or combination of dimensions as they want i.e. "Who" are the actors and "What" are their sentiment regarding any topic, changes in sentiment during "When" and "Where" and the reasons for change in sentiment as "Why". The final graph for tracking is been generated with a timeline.

There are very few research attempts where 5W structurization have been attempted. The ideas of 5Ws have been used successfully for a machine translation evaluation methodology [20]. The methodology addresses the cross-lingual 5W task: given a source language sentence and the corresponding target language sentence, it evaluates whether the 5Ws in the source have been comprehensibly translated into the target language. In addition we previously tried the 5W extraction task from Bengali [21].

From the next section we describe the development process of our 5W constituent based textual and visual summarization and tracking system.

3   Corpus Collections and Annotation

The present system has been developed for the Bengali language. Resource acquisition is one of the most challenging obstacles to work with resource-constrained languages like Bengali. Bengali is the fifth popular language[3] in the World, second in India and the national language in Bangladesh.

---

[3] http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers

The details of corpus development could be found in [22] for Bengali. We obtained the corpus from the authors. For the present task a portion of the corpus from the editorial pages, i.e., Reader's opinion section or Letters to the Editor Section containing 28K word-forms have been manually annotated with sentence level opinion constituents. The detail statistics about the corpus is reported in Table 1.

Table 1. Bengali News Corpus Statistics

Statistics	NEWS
Total number of documents	100
Total number of sentences	2234
Avgerage number of sentences in a document	22
Total number of wordforms	28807
Avgerage number of wordforms in a document	288
Total number of distinct wordforms	17176

Annotators were asked to annotate 5Ws in Bengali sentences in terms of Bengali noun chunks. Instructions have been given to annotators to find out the principle opinionated verb in a sentence and successively extract 5W components by asking 5W questions to the principle verb.

Table 2. Agreement of annotators at each 5W level

Tag	Annotators X and Y Agree percentage
Who	88.45%
What	64.66%
When	76.45%
Where	75.23%
Why	56.23%

Table 3. Agreement of annotators at sentence level

Annotators	X vs. Y	X Vs. Z	Y Vs. Z	Avg.
Percentage	73.87%	69.06%	60.44%	67.8%
All Agree	58.66%			

The agreement of annotations between two annotators (Mr. X and Mr. Y) has been evaluated. The agreements of tag values at each 5W level are listed in Tables 2. For the evaluation of the extractive summarization system gold standard data has been prepared and three annotators took part. The inter-annotator agreement for the identification of subjective sentences for opinion summary is reported in Table 3.

It has been observed that in the present task the inter-annotator agreement is better for Who, When and Where level annotation rather than What and Why level though a small number of documents have been considered.

Further discussion with annotators reveals that the psychology of annotators is to grasp all 5Ws in every sentence, whereas in general all 5Ws are not present in every sentence. But the same groups of annotators are more cautious during sentence identification for summary as they are very conscious to find out the most concise set of sentences that best describe the opinionated snapshot of any document. The

annotators were working independent of each other and they were not trained linguists. As observed, the most ambiguous tag to identify is "Why". The overall annotation has been done on 2234 sentences as mentioned in the Table 1. Generally each W type presents in a sentence only once but sometime it may twice. For example in the following sentence there are two "Who" tags. A post statistical analysis revealed that only in 3-5% cases each W tag repeats in a sentence and the percentage vary tag wise. Another important observation is every Ws are not present in every sentence. To better understand the distribution pattern of 5Ws in a corpus we gather a statistics for each 5W tag level as listed in Table 4

Table 4. Sentence wise co-occurrence pattern of 5Ws

Tags	Percentage						Total No. of Occurrence Of Each Ws in the Corpus
	Who	What	When	Where	Why	Overall	
Who	-	58.56%	73.34%	78.01%	28.33%	73.50%	1642
What	58.56%	-	62.89%	70.63%	64.91%	64.23%	1435
When	73.34%	62.89%	-	48.63%	23.66%	57.23%	1278
Where	78.0%	70.63%	48.63%	-	12.02%	68.65%	1533
Why	28.33%	64.91%	23.66%	12.02%	-	32.00%	714

```
The Gopal Krishna Gandhiy/Who, expressed his grief for the
rail accident and Smt. Mamata Banerjee/Who followed the
same line of act to express her own feelings.
```

Sentiment tagging is always very ambiguous because it differs from the writer to reader's perspective [23]. Therefore it is very hard to achieve high agreement score in sentiment data.

Another important observation is that 5W annotation task takes very little time for annotation. Annotation is a vital tedious task for any new experiment, but 5W annotation task is easy to adopt for any new language.

4   The 5W Extraction

The 5Ws semantic role labeling task demands and addressing various NLP issues such as: predicate identification, argument extraction, attachment disambiguation, location and time expression recognition. To solve these issues the present system architecture relies on Machine Learning technique and rule-based methodologies simultaneously.

One of the most important milestones in SRL literature is CoNLL-2005 Shared Task[4] on Semantic Role Labeling. All most all SRL research group participated in the shared task. System reports of those participated systems eminently prove that

---

[4] http://www.lsi.upc.es/~srlconll/st05/st05.html

Maximum Entropy[5] (ME) based models work well in this problem domain as 8 among 19 systems used ME as the solution architecture. The second best performing system [24] uses ME model uses only syntactic information without using any pre or post processing. For the present system we did a number of experiments and finally choose an n-gram (where n=4) window with the best-identified features to train the classifier.

Table 4 presents the distribution pattern of 5Ws in overall corpus. It is very clear that 5Ws are not very regular jointly in the corpus. Hence sequence labeling with 5Ws tags using ME will lead a label biased problem (as we reported in Section 7) and may not be an acceptable solution for present problem definition as concluded in [24] (although in a different SRL task).

We apply both rule-based and statistical techniques jointly to the final system. The rules are being captured by acquired statistics on training set and linguistic analysis of standard Bengali grammar. The features used in the present system are reported in the following section.

4.1   The Feature Organization for MEMM

The features to be found most effective are chosen experimentally. Bengali is an electronically resource scarce language, thus our aim was to find out the less number of features but the features should be effective. Involving more number of features will demand more linguistic tools, which are not readily available for the language. All the features that have been used to develop the present system are categorized as Lexical, Morphological and Syntactic features. These are listed in the Table 5 below and have been described in the subsequent subsections.

Table 5. Features

Types	Features	
Lexical	POS	
	Root Word	
Morphological	**Noun**	Gender
		Number
		Person
		Case
	Verb	Voice
		Modality
Syntactic	Head Noun	
	Chunk Type	
	Dependency Relations	

Part of Speech (POS): POS of any word cannot be treated as direct clue of its semantic but it definitely helps to identify it. Finding out the POS of any word can reduce the search space for semantic meaning. It has been shown by [25], [26] etc.

---

[5] http://maxent.sourceforge.net/

that the part of speech of any word in sentences is a vital clue to identify semantic role of that word.

Root Word: Root word is a good feature to identify word level semantic role especially for those types of 5Ws where dictionaries have been made like "When", "Where" and "Why". There are various conjuncts and postpositions, which directly indicate the type of predicate present in any sentence. As example জন্য, হেতু give clue that the next predicate is causative ("Why").

Gender: Gender information is essential to relate any chunk to the principle verb modality. In the case of "What"/"Whom" ambiguities gender information help significantly. For inanimate objects it will be null and for animates it has definitely a value. Bengali is not a gender sensitive language hence this feature is not such significant linguistically rather number and person features. But the statistical co-occurrence of gender information with the number and person information is significant.

Number: Number information helps to identify specially for "Who"/"What" ambiguities. As we reported in inter-annotator agreement section "Who" has been identified first by matching modality information of principle verb with corresponding number information of noun chunks.

Person: Person information is as important as number information. It helps to relate any head of noun chunks to principle verb in any sentence.

Case: Case markers are generally described as karaka relations of any noun chunks with main verb. It has been described that semantically karaka is the ancestor of all semantic role interpretations. Case markers are categorized as Nominative, Accusative, Genitive and Locative. Case markers are very helpful for almost in every 5W semantic role identification task.

Voice: The distinction between active and passive verbs plays an important role in the connection between semantic role and grammatical function, since direct objects of active verbs often correspond in semantic role to subjects of passive verbs as suggested by various researchers [24]. A set of hand-written rules helps to identify the voice of any verb chunk. The rules rely on presence auxiliary verbs like হয়েছে ,হোক etc indicate that the main verb in that particular chunk is in passive form.

Modality: Honorific markers are very distinctly used in Bengali and it directly reflects by the modality marker of any verb. As example the honorific variation করা/do are as কর (used with তুই: 2nd person either of same age or younger), করো (used with তুমি: 2nd person either of same age or slightly elder) and করুন (used with আপনি: 2nd persond generally for aged or honorable person). Verb Modality information helps to identify especially the "Who" tag. "Who" is identified first by matching modality information of principle verb with corresponding number information of noun chunks.

Head Noun: The present SRL system identifies chunk level semantic roles. Therefore morphological features of chunk head is only important rather other chunk members. Head words of noun phrases can be used to express selectional restrictions on the semantic role types of the noun chunks. For example, in a communication frame, noun phrases headed by Ram, brother, or he are more likely to be the SPEAKER

(Who), while those headed by proposal, story, or question are more likely to be the TOPIC (What).

Chunk Type: Present SRL system identifies noun chunk level semantic roles. Hence chunk level information is effectively used as a feature in supervised classifier and in rule-based post processor.

Dependency Relations: It has been profoundly established that dependency phrase-structures are most crucial to understand semantic contribution of every syntactic nods in a sentence [25], [26]. A statistical dependency parser has been used for Bengali as described in [27]. Shallow parsers[6] for Indian languages developed under a Government of India funded consortium project named Indian Language to Indian Language Machine Translation System (IL-ILMT) are now publicly available.

4.2   Rule-Based Post-processing

As described earlier post-processing is necessary in this setup. The rules developed here are either based on syntactic grammar, manually augmented dictionary or corpus heuristic.

In order to apply rule-based post-processor for "When" tag we developed a manually augmented list with pre defined categories as described in Table 6. Similar to "When", we categorized "Where" and "Why" as general and relative as listed in Table 7.

Table 6. Time Expressions

	Bengali		English Gloss
General	সকাল/সন্ধ্যে/রাত...		Morning/evening/night
	টার সময়/ঘটিকায়/মিনিট		O clock/hour/minute
	সোমবার/মঙ্গলবার		Monday/Tuesday
Relative	আগে/পরে...		Before/After…
	সামনে/পেছনে...		Upcoming/
	Special Cases	উঠলে/থামলে	When rise/When stop

Table 7. Locative Expressions

Type	Locative	
General	**Bengali**	**English Gloss**
	মাঠে/ঘাটে/রাস্তায়	Morning/evening/night/dawn
Relative	আগে/পরে...	Before/After…
	সামনে/পেছনে...	Front/Behind
Causative		
General	জন্য/কারনে/হেতু...	Hence/Reason/Reason
Relative	যদি_তবে	If_else
	যদিও_তবুও	If_else

---

[6] http://ltrc.iiit.ac.in/showfile.php?filename=downloads/shallow_parser.php

5   Performance of the 5Ws Extraction

The performance result of ML (1) technique has been reported in Table 9. After using rule-based postprocessor the system (2) performance increases as listed in the following Table 8.

It is noticeable that the performance of the MEMM-based model differs tag-wise. For such heterogeneous problem nature we propose a hybrid system as rule-based post processor followed by Machine Learning. The rule-based post processor can identify those cases missed by ML method and can reduce false hits generated by statistical system.

Table 8. Performance of 5Ws Opinion Constituents by MEMM + Rule Based-Post Processing

Tag	Precision (%)		Recall (%)		F-measure (%)		Avg. F-Measure (%)	
	1	2	1	2	1	2	1	2
Who	76.2	79.6	64.3	72.6	69.8	75.9		
What	61.2	65.5	51.3	59.6	55.9	62.4		
When	69.2	73.4	58.6	66.0	63.4	69.5	62.2	68.1
Where	70.0	77.7	60.0	69.7	64.6	73.4		
Why	76.2	63.5	53.9	55.6	57.4	59.2		

6   The Summarization Methodologies

The present system is a multi-document extractive opinion summarization system for Bengali. Documents are preprocessed with the subjectivity identifier (as described in [28]) followed by the polarity classifier (as described in [29]). All the 5W constituents extracted from each sentence and clustered depending upon common constituents present at document level. The document clusters are then formed as tightly coupled network. The node of the network is the extracted sentiment constituent and the edges represent the relationship among them.

The next major step is to extract relevant sentences from each constituent cluster that reflects the contextual concise content of the current constituent cluster. Our summarization system is a dynamic one and the output depends on user's dimension choices. To adopt this kind of special need we used Information Retrieval (IR) based technique to identify the most "**informed**" sentences from the constituents cluster and it can be termed as IR based cluster center for that particular cluster. With the adaptation of ideas from page rank algorithms [30], it can be easily observed that a text fragment (sentence) in a document is relevant if it is highly related to many relevant text fragments of other documents in the same cluster. The basic idea is to cover all the constituents' node in the network by the shortest path algorithm as given by user. The adaptive page rank algorithm helps to find out the shortest distance, which covers all the desired constituents' node and maximizes the accumulated edge scores among them. Accordingly sentences are chosen based on the presence of those particular constituents. The detail description could be found in the following subsection.

6.1   Constituent Based Document Clustering

Constituent clustering algorithms (*K-Means*) partition a set of documents into finite number of groups or clusters in terms of 5W opinion constituents. Documents are represented as a vector of 5W constituents present in the opinionated sentences within the document into various subjective sentences.

The similarity between vectors is calculated by assigning numerical weights to 5W opinion constituents and then using the cosine similarity measure as specified in the following equation.

$$s\left(\vec{d}_k, \vec{d}_j\right) = \vec{d}_k . \vec{d}_j = \sum_{i=0}^{N} w_{i,k} \times w_{i,j}$$

where $\vec{d}_k$ and $\vec{d}_j$ are the document vectors. *N* is the total number of unique 5Ws that exist in the document set $\vec{d}_k$ and $\vec{d}_j$. The $w_{i,k}$ and $w_{i,k}$ are the 5W opinion constituents that exist in the documents $\vec{d}_k$ and $\vec{d}_j$ respectively. An example of inter-document theme cluster has been reported in Table 9. The numeric scores are the similarity association value assigned by the clustering technique. A threshold value of greater than 0.5 has been chosen experimentally to construct the inter-document theme relational graph in the next level.

Table 9. Theme Clusters by 5W Dimensions

Generated Clusters						
5Ws	**Constituents**	**Doc1**	**Doc2**	**Doc3**	**Doc4**	**Doc5**
Who	Mamata Banerjee	**0.63**	0.01	**0.55**	**0.93**	0.02
	West Bengal CM	0.00	0.12	**0.37**	0.10	0.17
What	Gyaneswari Express	**0.98**	**0.79**	**0.58**	0.47	0.36
	Derailment	**0.98**	**0.76**	0.35	0.23	0.15
When	24/05/2010	**0.94**	0.01	0.01	0.01	0.01
	Midnight	**0.68**	**0.78**	0.01	0.01	0.01
Where	Jhargram	**0.76**	0.25	0.01	0.13	**0.76**
	Khemasoli	**0.87**	0.01	0.01	0.01	0.01
Why	Maoist	**0.78**	**0.89**	0.06	0.10	0.14
	Bomb Blast	0.13	**0.78**	0.01	0.01	**0.78**

To better aid our understanding of the automatically determined category relationships we visualized this network using the Fruchterman-Reingold force directed graph layout algorithm [31] and the NodeXL network analysis tool [32][7] as shown in Fig. 1. In the following graphical representation one color depict one cluster.

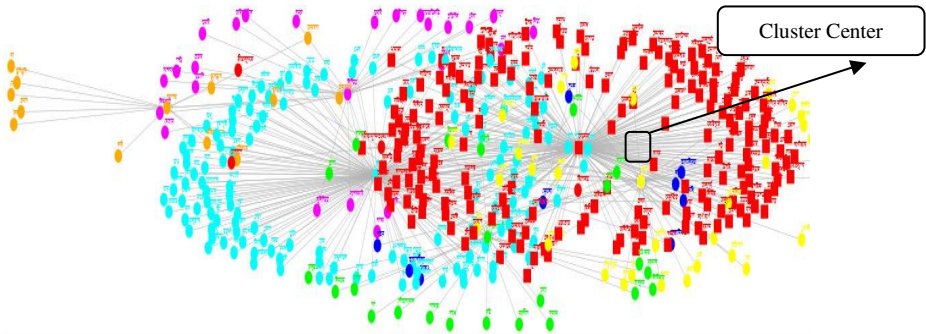[7] Available from http://www.codeplex.com/NodeXL

Fig. 1. Document Level Theme Relational Graph by NodeXL

6.2   Constituent Relevance Calculation

In the generated constituent network all the lexicons are connected with weighted vertex either directly or indirectly. Semantic lexicon inference could be identified by network distance of any two constituent nodes by calculating the distance in terms of weighted vertex. We computed the relevance of semantic lexicon nodes by summing up the edge scores of those edges connecting the node with other nodes in the same cluster. As cluster centers are also interconnected with weighted vertex so inter-cluster relations could be also calculated in terms of weighted network distance between two nodes within two separate clusters. As an example: suppose we have the following two clusters: A and B. A has m numbers of nodes while B consists of n numbers of nodes. a_x and b_y are the clusters centers of A and B.

$$A = \left\{ a_1, a_2, a_3, a_4 \ldots \ldots, a_x, \ldots \ldots a_m \right\} \quad B = \left\{ b_1, b_2, b_3, b_4 \ldots \ldots, b_y, \ldots \ldots b_n \right\}$$

The lexicon semantic affinity inference between a_x and b_y could be calculated as follows:

$$S_d(a_x, b_y) = \frac{\sum_{k=0}^{n} v_k}{k} \quad \text{----(3)} \quad \text{or} \quad \sum_{c=0}^{m} \frac{\sum_{k=0}^{n} v_k}{k} \times \prod_{c=0}^{m} l_c \quad \text{----(4)}$$

where $S_d(a_x, b_y) = $ semantic affinity distance between two constituent a_x and b_y. Equation (3) and (4) are for intra-cluster and inter-cluster semantic distance measure respectively. k=number of weighted vertex between two constituent a_x and b_y. v_k is the weighted vertex between two lexicons. m=number of cluster centers between two lexicons. l_c is the distance between cluster centers between two lexicons.

6.3   Dimension Wise Opinion Summary-Visualization-Tracking

The working principle of the present system is as follows.

• The system identifies all the desired nodes in the developed semantic constituent network as given by user in the form of 5W.

- Inter-constituents distances have been calculated from the developed semantic constituent network. For example, suppose user gave the following input. Therefore the calculated inter-constituents distances may look like the Table 10.\

Input:

Who	**What**	**When**	**Where**	**Why**
মমতা ব্যানাজী	ভানেশ্বরী এক্সপ্রেস	মধ্যরাত	ঝাড়গ্রাম	মাওবাদী
(Mamata Banerjee)	(Gyaneswari Express)	(Midnight)	(Jhargram)	(Maoist)

Table 10. Calculated inter-constituents distances

Type	Inter-Constituents Distances				
	Who	What	When	Where	Why
Who	-	0.86	0.02	0.34	0.74
What	0.86	-	0.80	0.89	0.67
When	0.02	0.80	-	0.58	0.23
Where	0.34	0.89	0.58	-	0.20
Why	0.74	0.67	0.23	0.20	-

- All the sentences consist of at least one of the user-defined constituents are extracted from all the documents.
- Extracted sentences are then ranked with the adaptive Page-Rank algorithm based on the constituent present in that sentence. In the first iteration the standard IR based Page-Rank algorithm assign a score to each sentence based on keyword (constituents are treated as keyword in this stage) presence. In the second iteration the calculated rank by the Page-Rank algorithm are multiplied with the inter-constituents distances for those sentences where more than one constituent present. For example: in the next sentence two Ws: "Who" and "What" are present jointly as constituent. Suppose the assigned rank for the following sentence by the basic Page-Rank algorithm is n. Then in the next iteration the modified score will be $n*0.86$, because the inter-constituents distances for "Who" (মমতা বন্দ্যোপাধ্যায়) and "What" (ভানেশ্বরী এক্সপ্রেস) is 0.86.

মমতা_বন্দ্যোপাধ্যায়/**Who** ভানেশ্বরী_এক্সপ্রেস_ঘটনাকে/**What** রাজনৈতিক চক্রান্ত বলে মন্তব্য করেন।

  English   Gloss: Mamta_Bandyopadhyay/**Who**   commented   that   the
  Gyaneshwari_Express_incident/**What** is a political conspiracy.

- The ranked sentences are then sorted by descending order and top-ranked 30% sentences (from all retrieved sentences) are shown as a summary.

Ordering of sentences is very important is very important in case of summarization. We prefer the temporal order of sentences as they occurred in original document, when it published.

The visual tracking system consists of five drop down boxes. The drop down boxes give options for individual 5W dimension of each unique Ws that exist in the corpus. The present visual tracking system facilitates users to generate opinion polarity wise graph based visualization and summary on any 5W dimension and combination of 5W dimensions, as they want. (Shown in Fig. 2).
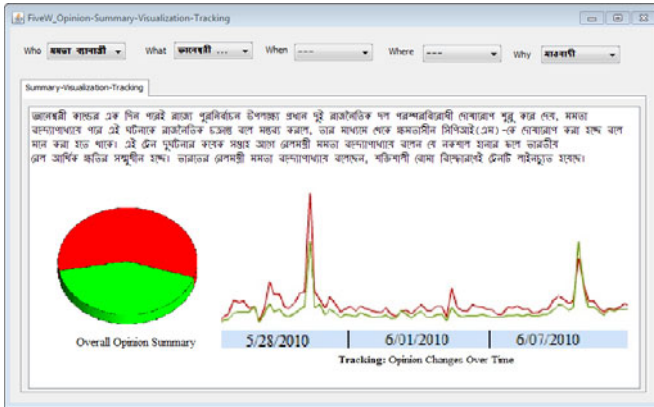
Fig. 2. A Snapshot of the Present Summarization System

7   Experimental Result

To evaluate the present system we follow a two-fold evaluation mechanism. The first-fold evaluation is to understand the system performance to detect relative sentences prior to generate final summary (as mentioned in the third step of the Summary process). For this evaluation we check system-identified sentences with every human annotator's gold standard sentences and finally we calculated the overall accuracy of the system as reported in Table 11.

Table 11. Final Results subjective sentence identification for summary

Metrics	X	Y	Z	Avg.
Precision	77.65%	67.22%	71.57%	72.15%
Recall	68.76%	64.53%	68.68%	67.32%
F-Score	72.94%	65.85%	70.10%	69.65%

Table 12. Human Evaluation on 5W Dimension Specific Summaries

Tags	Average Scores				
	Who	What	When	Where	Why
Who	-	3.20	3.30	3.30	2.50
What	3.20	-	3.33	3.80	2.6
When	3.30	3.33	-	2.0	2.5
Where	3.30	3.80	2.0	-	2.0
Why	2.50	2.6	2.5	2.0	-
Overall	3.08	3.23	3.00	2.77	2.40

It was a challenge to evaluate the accuracy of the dimension specific summaries. It is hardly possible to make a human extracted gold summary set for every dimension combinations; therefore we propose a direct human evaluation technique. Two evaluators have been involved in the present task and they are asked to give evaluative score to each system-generated summaries. We use a 1-5 scoring technique whereas 1 denotes very poor, 2 denotes poor, 3 denotes acceptable, 4 denotes good and 5 denotes excellent. The final evaluation result of the dimension specific summarization system is reported in Table 12.

8   Conclusion

The present paper started with a very basic question "*What is the End User's Requirement?*". To answer this question we do believe that our proposed 5W Summarization –Visualization-Tracking system could be treated as a qualitative and acceptable solution. To compare our suggestion we presented a vivid description of the previous works. Another self-contributory remark should be mentioned that according to best of our knowledge this is the first attempt on opinion summarization or visual tracking for the language Bengali. Moreover the 5W structurization is new to the community and proposed by us.

References

1. Dasgupta, S., Ng, V.: Topic-wise, Sentiment wise, or Otherwise? Identifying the Hidden Dimension for Unsupervised Text Classification. In: EMNLP 2009 (2009)
2. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval (2008)
3. Seki, Y., Eguchi, K., Kando, N.: Analysis of multi-document viewpoint summarization using multi-dimensional genres, pp. 142–145. AAAI (2004)
4. Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In: Proceedings of ACL (2004)
5. Ku, L.-W., Li, L.-Y., Wu, T.-H., Chen, H.-H.: Major topic detection and its application to opinion summarization. In: Proceedings of the SIGIR, pp. 627–628 (2005)
6. Liang, Z., Eduard, H.: On the summarization of dynamically introduced information: Online discussions and blogs. In: AAAI-CAAW, pp. 237–242 (2006)
7. Kawai, Y., Kumamoto, T., Tanaka, K.: Fair News Reader: Recommending News Articles with Different Sentiments Based on User Preference. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part I. LNCS (LNAI), vol. 4692, pp. 612–622. Springer, Heidelberg (2007)

8. Hu, M., Liu, B.: Mining and summarizing-customer reviews. In: Proc. of the 10th ACM-SIGKDD Conf., pp. 168–177. ACM Press, New York (2004)

9. Zhuang, L., Jing, F., Zhu, X., Zhang, L.: Movie review mining and summarization. In: ACM-SIGIR-(CIKM) (2006)

10. Das, S.R., Chen, M.Y.: Yahoo! for Amazon: Sentiment extraction from small talk on the Web. Management Science 53(9), 1375–1388 (2007)

11. Gamon, M., Aue, A., Corston-Oliver, S., Ringger, E.: Pulse: Mining Customer Opinions from Free Text. In: Famili, A.F., Kok, J.N., Peña, J.M., Siebes, A., Feelders, A. (eds.) IDA 2005. LNCS, vol. 3646, pp. 121–132. Springer, Heidelberg (2005)

12. Yi, J., Niblack, W.: Sentiment mining in WebFountain. In: Proceedings of the International Conference on Data Engineering, ICDE (2005)

13. Gruhl, D., Chavet, L., Gibson, D., Meyer, J., Pattanayak, P., Tomkins, A., Zien, J.: How to build a Webfountain: architecture for very large-scale text analytics. IBM Systems Journal 43(1), 64–77 (2004)

14. Carenini, G., Ng, R.T., Pauls, A.: Interactive multimedia summaries of evaluative text. In: Proceedings of Intelligent User Interfaces (IUI), pp. 124–131. ACM Press (2006)

15. Gregory, M.L., Chinchor, N., Whitney, P., Carter, R., Hetzler, E., Turner, A.: User-directed sentiment analysis: Visualizing the affective content of documents. In: Proceedings of the Workshop on Sentiment and Subjectivity in Text, pp. 23–30. ACL (2006)

16. Lloyd, L., Kechagias, D., Skiena, S.S.: Lydia: A System for Large-Scale News Analysis. In: Consens, M.P., Navarro, G. (eds.) SPIRE 2005. LNCS, vol. 3772, pp. 161–166. Springer, Heidelberg (2005)

17. Ku, L.-W., Liang, Y.-T., Chen, H.-H.: Opinion extraction, summarization and tracking in news and blog corpora. In: AAAI-CAAW, pp. 100–107 (2006)

18. Mishne, G., de Rijke, M.: Moodviews: Tools for blog mood analysis. In: AAAI-CAAW, pp. 153–154 (2006)

19. Fukuhara, T., Nakagawa, H., Nishida, T.: Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. In: ICWSM (2007)

20. Parton, K., McKeown, K., Coyne, B., Diab, M., Grishman, R., Hakkani-Tür, D., Harper, M., Ji, H., Wei, Y.M., Meyers, A., Stolbach, S., Sun, A., Tur, G., Wei, X., Sibel, Y.: Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task. In: The Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 423–431 (2009)

21. Das, A., Ghosh, A., Bandyopadhyay, S.: Semantic Role Labeling for Bengali Noun using 5Ws: Who, What, When, Where and Why. In: The Proceeding of the International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLPKE2010), Beijing, China, pp. 1–8 (2010)

22. Ekbal, A., Bandyopadhyay, S.: A Web-based Bengali News Corpus for Named Entity Recognition. LRE Journal 42(2), 173–182 (2008)

23. Tang, Y.-J., Chen, H.-H.: Emotion Modeling from Writer/Reader Perspectives Using a Microblog Dataset. In: Proceeding of the Workshop Sentiment Analysis Where AI Meets Psychology (2011)

24. Haghighi, A., Toutanova, K., Manning, C.D.: A Joint Model for Semantic Role Labeling. In: CoNLL-2005 Shared Task (2005)

25. Gildea, D., Jurafsky, D.: Automatic Labeling of Semantic Roles. In: Association for Computational Linguistics (2002)

26. Palmer, M., Gildea, D., Kingsbury, P.: The Proposition Bank: A Corpus Annotated with Semantic Roles. Computational Linguistics Journal 31(1) (2005)
27. Ghosh, A., Das, A., Bhaskar, P., Bandyopadhyay, S.: Dependency Parser for Bengali: the JU System at ICON 2009. In: NLP Tool Contest ICON 2009 (2009)
28. Das, A., Bandyopadhyay, S.: Subjectivity Detection using Genetic Algorithm. In: WASSA 2010, Lisbon, Portugal, August 16-20 (2010)
29. Das, A., Bandyopadhyay, S.: Phrase-level Polarity Identification for Bengali. IJCLA 1(1-2), 169–182 (2010) ISSN 0976-0962s
30. Page, L.: PageRank: Bringing Order to the Web. Stanford Digital Library Project (1997)
31. Fruchterman Thomas, M.J., Reingold Edward, M.: Graph drawing by force-directed placement. Software: Practice and Experience 21(11), 1129–1164 (1991)
32. Marc, S., Shneiderman, B., Milic-Frayling, N., Rodrigues, E.M., Barash, V., Dunne, C., Capone, T., Perer, A., Gleave, E.: Analyzing (social media) networks with NodeXL. In: C&T 2009: Proc. Fourth International Conference on Communities and Technologies (2009)