CHAPTER 6

MONITORING PUBLIC PARTICIPATION IN MULTI-LATERAL INITIATIVES USING SOCIAL MEDIA INTELLIGENCE

6. 1 Introduction

Governments, multilateral agencies like the World Bank, United Nations and Development Banks are involved in a variety of developmental activities across the world. A lot of resources are spent to ensure proper consultations and post implementation verification of results. But this does not completely ensure whether the objectives are achieved. The new web technologies have provided methodologies and developed tools that allow users to pool resources on projects over the internet. Social media allows real-time feedback for citizens which helps in monitoring developmental initiatives for Governments and Multilateral agencies. The role of technology ensures that the on-going consultations and feedback can be captured, analysed and used in understating the stakeholder reaction to the project and its implementation. This helps in making necessary course

corrections avoiding costly mistakes and over runs. Here, we model a tool to monitor study and analyse popular feedback, using forums, Social Media, surveys and other crowdsourcing techniques. The feedback is gathered and analysed using both quantitative as well as qualitative methods, to understand the opinion of the crowd. The summation and visualisation of patterns are automated using text mining and sentiment analysis tools including text analysis, tagging/annotation. SENTIMATCH developed is discussed in detail in the previous chapter. These patterns provide insight into the popular feedback and sentiment effectively and accurately than the conventional method. The model is created by integrating such feedback channels. Data is collected, analysed and the results are presented using tools developed in Open Source platform.

6.2 Collaboration and Consultation Portal

Collaboration is the key aspect in every organization where opinions of their policies and programs (or products and services) are collected from a wide variety of custodians like the society (or customers), policy stakeholders (or employees), experts (business partners) external stakeholders (General public), etc. Organizations face challenges in terms of analysing and reporting contents coming from various mediums including but not limited to Feedbacks, Reviews, Comments, Social Media Interactions, Blogs and Surveys are hosted internally or on the cloud. Manual evaluation of these are impossible, and there is a need to automate the review process and bring

out the message in a format which stakeholders can understand and act upon appropriately. We live in the information age where the content gets generated from various sources dynamically. Sourcing all the unstructured information into a common platform and processing the information are the challenges faced by the Data Engineers. Named Entity Recognition (NER), Relationship Extraction and Sentiment Analysis (Polarity) are the key aspects of a Text Analysis Process. Many research institutions are contributing heavily into this platform to enable the machine learning from the human generated content.

Governments and Multilateral agencies work on a variety of projects which impact the society. A large majority of mankind is living in a conflict affected environment. The upkeep of peace and improving the living conditions include investment in destroyed and weak infrastructure. This is apart from the various challenges including lack of food and essential services. This involves expenditure of a large amount of money with multiple stakeholders. The beneficiaries of these projects are mostly communities of economically backward and less developed countries having limited access to communication and consultation. The project aims at benefiting the community, in terms of social and economic upliftment, better infrastructure, education, health, environment and sustainable practices. There is often a disconnect due to the long time frames, number of stakeholders, as also the distance and communication lag between the affected community, the project implementation team, and the funding agencies. In the current context, greater transparency and open

communications with all are very much essential for the involvement of stakeholders during all the stages of a project. Computer and communication technologies promise a greater role in the peace building efforts. The ubiquity and the seamless communication capability of social media channels help to tide over these problems of traditional media.

6.3 Social Media and Public Participation

Traditional print media and communication like radio were the standards in earlier peace building efforts. But the more transformational nature of the promising technologies enhances the current process. There are a variety of platforms currently available with the public, to express their experiences and opinions. These expressions further influence the building up of opinions of others. Dialogue is a critical component in this scenario. Social media can support and set the stage for broader participation. Tapping into these sentiments will make it easier to implement projects. Policy decisions can be shaped more easily. Monitoring of media has been in the process for a long time whether for public relations, for understanding public opinion or for general market intelligence. But with fragmenting media and changing public sentiments over long term projects, traditional monitoring methods like press clippings, field surveys or ad hoc research are highly insufficient. Thus, involving the participation of the public, and analysing public sentiment, much of which is expressed online becomes a necessary part of the public engagement.

The collaboration engine supports six channels for data collection; Review /Feedbacks, Comments, Social Media Interactions (LinkedIn, Facebook, Twitter, and Google+), Blogs and Surveys. These channels act as data consumers and store those in data repositories for analysis. This also allows submissions of additional materials pertaining to any channels; which allows content extraction from documents (PDF, Excel, and Documents) as well. Consultation Hub, shown in figure 6.1 is the platform where all the individual consultation specific analysis gets added.

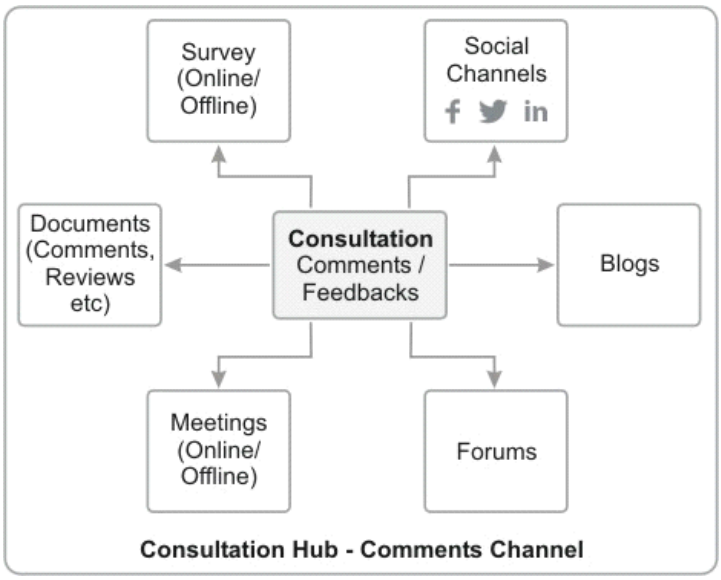

Fig.6.1.Consultation Hub

The solution enables the stakeholders to pick and choose the source of information (channels) depending on the interest or the need for the data collection practice. For example, someone wants to compare between

meeting comments and social media comments. Sometimes, the user wants to filter analysis based on the geographical or sector or topic or theme. The reason is that, when we compare the sentiments of two data sets and if both are from different Sector or Topic or Geography, those results may be logically different. Technically the data may be correct, but contextually it may be incorrect. The solution should enable the custodians to identify and define the logical relationships, depending on the context of the business function.

Organizations using this model can build their own vocabulary (Keywords, Entities and Concepts) for processing the data collected from various sources. The institutions follow specific vocabulary to avoid confusions or different interpretations of the results. A Collaboration Portal is developed for Consultation Hub which facilitates integration with industry proven data analysis platforms, like Alchemy and Open Calais along with the Analysis Engine. The custom solution benefits advanced optimization to narrow down the results and fine tune the process according to the business requirements.

6.4 Challenges of the Consultation Hub

The Consultation hub will provide the outcome of consultations. The challenges faced in the conventional method are :

- Effective feedback at the appropriate time
- Collecting and collating the feedback

- Intended beneficiaries and residents affected by these projects are distributed and vast
- Beneficiaries having no direct channels for conveying their opinions, and the extended logistics of such an exercise make it difficult to study and incorporate the feedback.
- With multiple channels of social media expression, there is an information overload and skewing tendency.
- With more data points, with all and sundry giving opinions and feedback on the multitude of media, collating, curating all this data, and making sense of it, to achieve the desired goals are challenging.

6.5 Related Technologies

6.5.1 Social Media Monitoring

Social media monitoring makes use of text mining and Natural Language Processing (NLP), where the user-generated contents are analysed to understand the awareness, mood and emotions in relation to a particular topic. Text Data Mining (or opinion mining) is the process of obtaining relevant and high quality information from text, typically by studying and coming up with patterns and trends using statistical pattern learning. By using known statistical patterns and keywords or taxonomies, it is possible to devise parsing techniques, which can highlight certain words, and ignore the others. By deriving patterns from the text data so structured, it is then

examined for relevance and categorized, keywords extracted and relationships between words or concepts analysed. Social media monitoring and analysis is a universal term used to accommodate 'brand monitoring,' 'buzz monitoring' and 'online anthropology,' to 'market influence analytics,' 'conversation mining' and 'online consumer intelligence'.

Sentiment Analysis attempts to extract sentiments or opinions, associated with positive or negative feelings or polarities towards a specific subject within one or more documents, rather than just branding the entire document as positive or negative. So it attempts to identify how sentiments are expressed in texts, and whether those expressions are favourable or unfavourable to the subject.

Sentiment Analysis involves the identification of the following:

- Sentiment Expressions or the statement containing sentiments
- Polarity and strength of the expressions (negative or positive); and
- The relationship to the subject (the sentiment expressed about a car may not be relevant while considering opinions about a road construction).

6.5.2 Lexical and Quantitative Analysis

Lexical Analysis has traditionally been used to design compilers. A modified lexical analyser is used to study word distributions. Quantitative text analysis is used to extract semantic or grammatical relationships between words, in order to find meanings or patterns, as in profiling. A

variety of techniques support these models, including pattern recognition, tagging and annotation, link and association analysis, and predictive analysis. The recognition of specific related keywords (e.g. country names or abbreviations, names of political or social leaders or organizations) or patterns (email addresses, phone numbers, addresses, numeric or price data etc.), within the context of textual data, is also possible. While decoding such data, context becomes very important. For example "Ford" can refer to a former US president, a car model, a movie, or some other entity. Disambiguation will require an understanding of the context.

6.5.3 Text Analysis , Semantic tagging, and Analysis

Statistical models and tools, Lexicon tools and models for quantitative processing have been around for some time. But due to the complexity of the use of these models, the processing power required, and the lower reliability made this more of a tool for research or large organizations. Sentiment Analysis and Qualitative tools are computing intensive tools and hence were not very popular earlier, though algorithms were available. There was difficulty in creating the taxonomies that improve the quality of analysis. However, now with the advent of various commercial and open source tools, text analysis is available to a much larger user base. Governments, Multilateral agencies, as well as Non Profits and For Profits, all adopt these tools. The SENTIMATCH is incorporated into the Collaboration Portal. An example of analysis of the comments is illustrated.

The number of tools which work on semantic and text analysis is increasing.

A lot of Open Source and proprietary tools are available. But each one has its own limitations and cannot be used as such. For our needs, we have combined various tools as explained in detail in the last chapter and built part of our own taxonomies to suit the domain we have selected.

6.4 Other Analytic Techniques

Several other analytic techniques are combined to extract further information. Relationship, fact and event extraction involves identifying relationships between entities and other information in the text. Concept and Entity extraction and production of granular taxonomies are also important to the text mining tasks. The final objective is to understand the relationships between named entities, using Natural Language Processing and analytic methods to populate a database or search index with the information extracted.

6.5 Open Source and Commercial Tools Used

Some of the commonly available and useful tools used include:

Open Calais: Open Calais [67], by Reuters, provides a web service allowing you to submit text for analysis and receive results of the analysis. With a free API, a lot can be done. However, the service is chargeable for usage beyond certain limits. Also, the concern is that the data will be processed on their servers; where privacy issues are to be taken care of. The advantage is the usage of their extensive taxonomy and semantics library.

Text Razor: Another service with free and paid services. Text Razor [68]

uses Natural Language Processing and Artificial Intelligence, to provide content analysis. Other similar tools available provide free or paid services and include Alchemy [69] Silo [78], etc.

ATLAS.ti : Standalone tools like ATLAS.ti [79], nVivo [80] provide many of these features and more. These tools are collectively called CAQDAS (Computer Assisted Qualitative Data AnalysiS Tools).

Stanford NER: The Stanford Named Entity Recognizer (NER) [70] provides a set of libraries, which can be used, among other things for text analysis. This package has well-engineered feature extractors for Named Entity Recognition. The advantage is that it has many options for defining feature extractors also.

Apache UIMA and TIKA: Apache UIMA or Unstructured Information Management Applications and TIKA [71] is used for text analysis and parsing. It extracts metadata from structured text content and unstructured content like audio and video. It is easy to integrate with Lucerne/SOLR for indexing and searching.

6.6 Research Methodology

Most public policy, planning, and projects implementation address the inter-relationships between different stake holders including the public in the region, the regional governments, civil society, the implementing agency and the funding agencies. Stakeholders can have differing views on the

project or policy, based on their understanding and the effect of the project on their daily life.

The project seeks inputs from all stakeholders during all phases of the project. The levels of public participation help in ensuring transparency, by bringing about public accountability on issues and decisions that affect the community.

The consultation hub and the collaboration channel created help to:

1) Promote public participation into an engagement

2) Bringing forth the public opinion and tracking it

3) Building tools to quantitatively and qualitatively assess interactions between people.

4) Building a feedback system to respond to the initiatives.

6.7 Case study

Discussion Forums, Surveys, Rating/Polls, to simpler forms like Thumbs up/down or Five Star Ratings are all popular methods, used for getting feedback from sites. Increasingly Facebook, Twitter and other Social Media sites are also used to track user feedback. Since all these platforms are now commonly used and prevalent, the model includes the discussions from all the channels. The framework is specifically built to collect, aggregate and analyse multiple social media streams. This is a portal with media

monitoring tool which does real-time keyword extraction, text analytics and topic visualization. This helps in directly connecting the expression with real time response platforms. This helps in monitoring the conversations and thereby reducing the elements of any negative feeling and respond to them if at all through the response platform.

Collaboration Portal with tracking tool is built for Multilateral Development Agencies for management and monitoring of multiple consultations. All consultations of the agency and partner agencies across the world are managed through consultation hub. When a consultation starts, an entry is made, demarcating geography involved, stakeholders whether public or private and key themes or areas of consultation. These key areas can be the environment, health, power etc. Once these parameters are decided, it starts collecting the inputs.

6.8 Architecture of the Automated Monitoring and Evaluation for Multilateral Development Agency

The Collaboration portal of Consultation Hub helps anyone to find, share and participate in consultations that interest him. Users can be involved at any stage. Anyone including public can be a part of the consultation and can generate content. This creates a large volume of data. The tool will exploit opinion mining and promote public participation for better decision making. This helps the participating organizations to improve the delivery mode of services and resource optimization. The expressions are analysed with

opinion miners understanding the natural language using knowledge bases. A machine readable ontology is defined to provide a unified schema for interpretation.

The crawler collects data from multiple data automatically. Inputs are collected from all channels including blogs, forums, Facebook posts, etc. Semantic tagging is done using the services including Apache TIKA and UIMA and Text Razor. Tags are populated to SOLR indexing

Natural Language Processing algorithms are applied to every post, interpreting emotions as positive and negative. The machine learning algorithms run across the data, identifies hidden characteristics, and bring out previously unnoticed patterns.

Automatic identification of the topic of every post gives insights into their authorship. Semantic Context is understood from Ontology and Taxonomy. The knowledge base is continuously updated.

Visualization tools help in understanding and acting on complex information, thus providing intelligence from unstructured data.

6.9 Solution Architecture

Figure 6.2 illustrates a sample implementation for the usage of one of these mechanisms – comments or Forum. This is further taken up for analysis. This is a text based comment in the feedback system. Apart from the simple textual data, other methods like star ratings, polls, and on-line survey

queries are collected regularly at data collection points. Compared with text data, these are easier to analyse and collate.

This solution helps organisation to build their own vocabulary for processing data collected through various sources. Apache UIMA (Unstructured Information Management Applications) an add-on to Solar for indexing unstructured content is used. It is an API which connects easily with Solar and allows connectivity with various open text analysis engines (Alchemy and Open Calais). We were able to gather Keywords, Concepts, and Entities out of the text passed to the analysis engine.

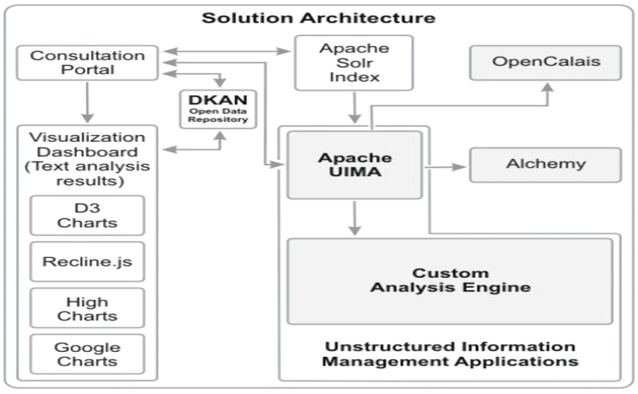

Fig. 6.2 Solution Architecture

The critical aspect of the whole solution is that how best we portray the processed information in a way it connects to all sorts of people easily and effectively; through strong self-depicting visualizations. Data visualization tools help to consume those results and portray those effectively. It has enabled the result visualization effectively, irrespective of any devices or

software solutions.

In this case study, we have taken text data, not concentrating on categorical data or numerical where results are easily obtainable for the purpose of analysing feedback and measuring the popular sentiments.

6.10 Technical Architecture

Collaboration portal has exciting Information Extraction features. These features are extracted through the integration of open source libraries like Apache TIKA toolkit, Open NLP, and Apache UIMA project. Metadata is extracted from the source document.

6.11 Working of Collaboration Portal with Consultation Hub

The first step is to properly identify the semantic relationships between the sentiment expressions and the subject. By applying semantic analysis with syntactic parser and sentiment taxonomies, sentiments are found. It has been possible to get 80% or more precision in finding sentiments within multiple documents. The concept can be further explained with the help of an example of what is involved. World Bank is a multinational, multilateral development agency engaged with developing countries in reducing extreme poverty. Partnering with Government, they give financial and technical assistance for their projects. This needs to be constantly monitored. Banks' effort to review and update safeguard policies is to be welcomed. But linking up clearly with other reforms within the bank and

elsewhere will be critical. Specific consultation Hub for the activity is marked with time period in the collaboration portal for monitoring.

6. 12 Text Analysis

Public conversations are tracked from different channels to understand what the public is talking about. The sentiments are tracked, based on the written message. NLP offers intelligence. The text is tagged as positive, negative or neutral. Pipe the result into the tool, which assigns a sentiment score. NLP analyses the language patterns. This analysis helps much more than sentiment analysis. It can also monitor the malicious acts like attacks.

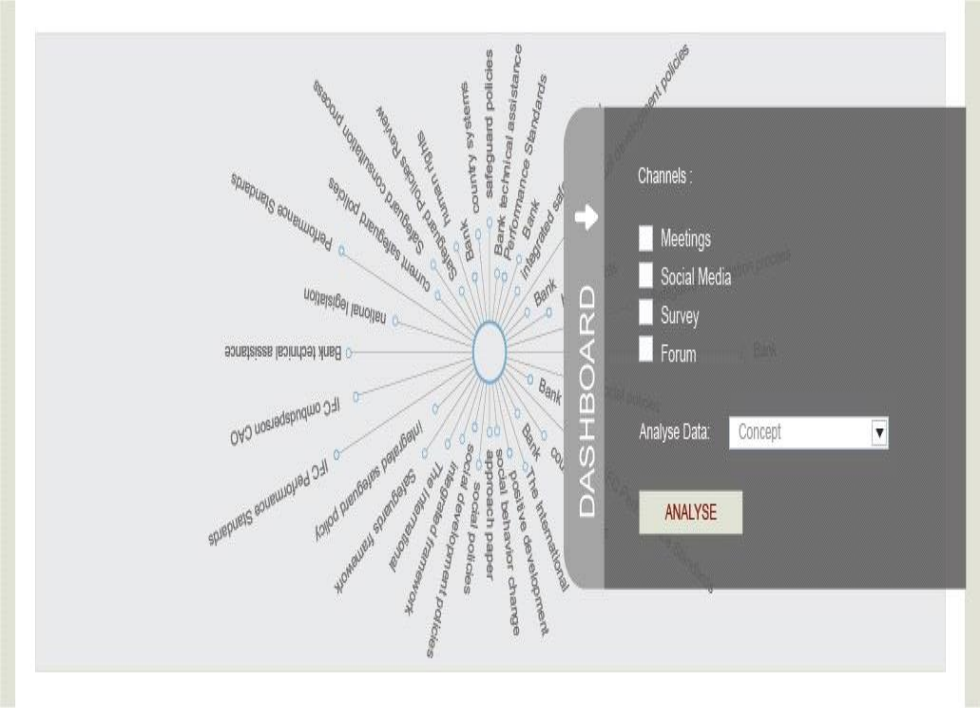

Fig 6.3 UI for channel Monitoring

A user interface (UI) is provided for channel monitoring as shown in fig 6.3. It listens to what people are talking about . Two different views of crowd sourcing are shown in Fig 6.4 and Fig 6.5

Fig. 6.4 Screenshot of Crowd Sourcing

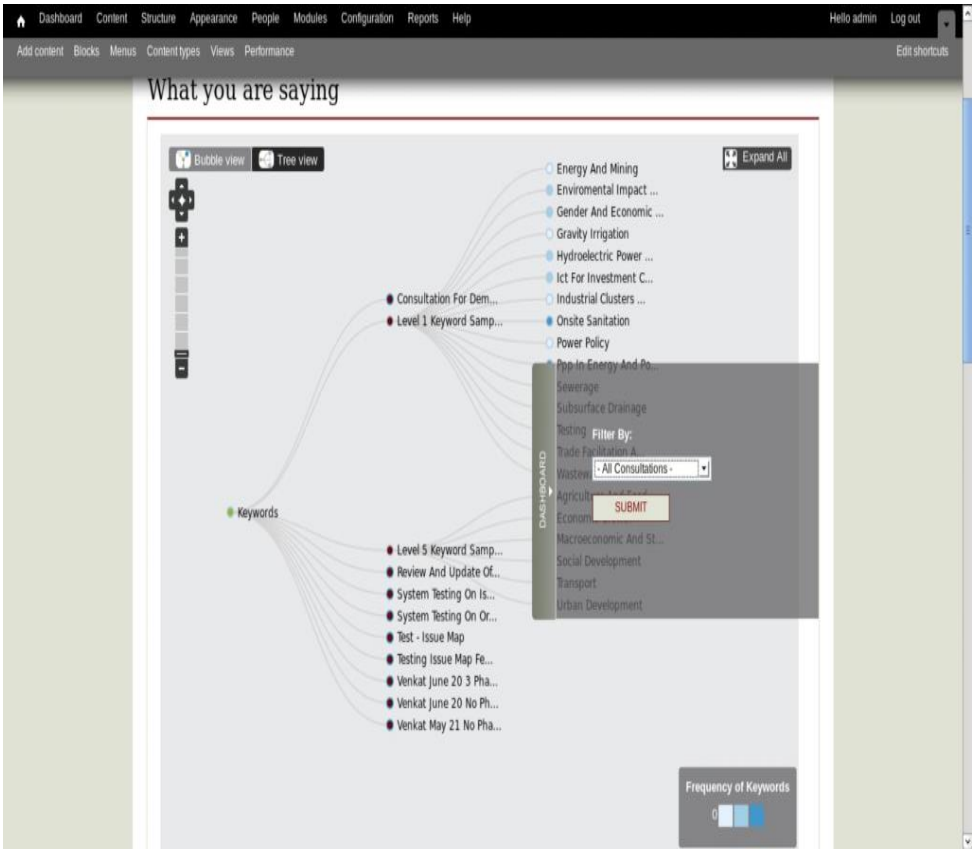

Fig.6. 5 Crowd Sourcing in Tree Structure

The core ideology was to reuse the solution stack as much possible and remain in the open space to avoid additional financial overheads. UIMA (Unstructured Information Management Applications) solution is used. The API connects easily with Solar and allows connectivity with various text analysis engines. Keywords, Concepts, and Entities derived out of the text are gathered.

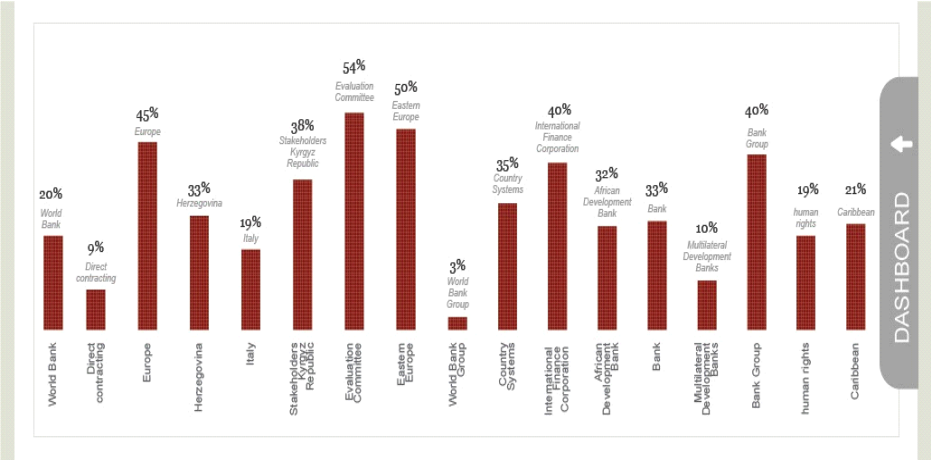

Fig.6. 6 Graphical View of Dashboard

6.13 Deployment of the Collaboration Portal

Collaboration portal is deployed for monitoring and evaluation of public opinion. SENTIMATCH is integrated into this portal. Expert views for better decision making is implemented through public consultation portal. This provides collaboration between various funding agencies and government, civil society organizations and the public. The consultation portal is made for a particular project for a particular region. The idea is the enhancement of public consultation and citizen engagement across the regulatory cycle for a wider outreach to stakeholders. This will make sure that affected parties are heard, will enhance transparency and will facilitate inclusiveness and trust in overall decision-making processes, leading to

more acceptability for the project. The stakeholders include public, NGOs, Government, academics, civil society and more.

There are multiple media of information channels. The primary medium is the live consultation meetings held publicly with the stakeholders. Online discussion forums including social media channels are also there. The other important channel is the informal forum including newspapers and other news media, social media like FaceBook, blogs, tweets etc. which are owned or controlled by third parties. SENTIMATCH listens to all social media conversations for related keywords and does an on-going sentiment analysis to know the feel of the swings in opinions of the populace and other stakeholders.

The consultation hub portal with SENTIMATCH helps in monitoring the civil sentiments in societies around the world. A typical example of this implementation is The Social Cohesion and Reconciliation (SCORE) Index. SCORE examines two main components of peace – cohesion and reconciliation. Social cohesion refers to the intricate relationship. Reconciliation is the harmonious coexistence between groups that were previously engaged in an event of dispute or conflict. The quality of coexistence between people within their own group and with the institutions that surround them is measured. Apart from that, the culturally-specific components of peace that vary across different contexts is also considered to improve the analysis. SCORE was developed in Cyprus. This adaptable tool is suited for multi-ethnic societies that was in conflict and is in the peace-building and state-building challenges.

6.14 Findings

Input for the analytics are taken from different input sources like Meetings, Social Media, Surveys and Forums. These are configured by the administrator and forms the input for harvesting and mining live streaming data. The appropriate ontologies, taxonomies and dictionaries are applied.

A sample screenshot is given in figure 6.3. We see that two statements are being made here. The first statement indicates a favourable sentiment, while the second one shows a negative opinion. Thus, to do a proper analysis, we would need to identify the individual statements and present results accordingly.

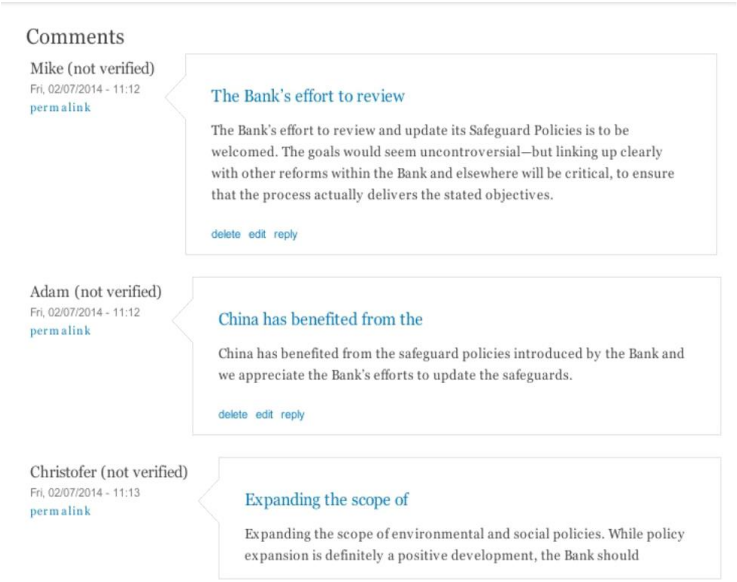

Fig. 6.7 Comments on a consultation

The analysis of the comments using relevant taxonomy is described. The screen in figure 6 4 shows the analysis being started

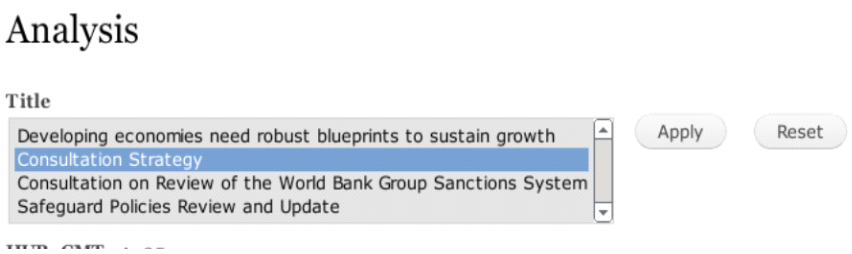

Fig. 6.8 Analysis of Comments

Result Analysis is done as follows:

Keywords

World Bank , world bank group , World Bank officials , social safeguard principles , Parliamentarians.\nThe best indicator , so-called \Development projects.\ , \Financial Intermediary Lending\ , best test , current exercise , genuine effort ,

Concept

Concept	Relevance
World Bank Group	96.75 %
World Bank	71.89 %

Entities

Entity	Count
World Bank	1
World Bank Group	2
IFC	2
India	1
World Bank	7
Principle of Intergenerational Equity	1
FPIC	1

Fig. 6.9 Result Analysis

Correlation between specific words are as shown in Table 6.1

World Bank	20%
Direct Contracting	9%
Europe	45%
Evaluation Committee	54%
African Development Bank	34%
Human Rights	19%
International Finance Corporation	40%

Table 6.1 Co-Relation between Specific words

Sociologists, Anthropologists, Peace workers and statisticians are collaborated to create this platform. They identify the key dimensions and metrics which could be the predictor of internal discontent, or social tensions; as also those which promote peace and harmony. The actionable insights from social media conversations which was previously hidden is the social intelligence output. By implementing this kind of platform and by monitoring and taking preventive steps, as well as proactively promoting Social Cohesion and Reconciliation, United Nations is able to monitor and reduce incidences of violence in key tension spots. We are addressing the seventh research question here.

6.15 Conclusion

Content Analysis, as well as Text Analysis, is just coming into its own. With huge amounts of data, an exciting range of tools and usages, we will see more of analysis. Digital dashboards showing project performance, collection and analysis of stakeholder feedback and comments are crucial along with Social Media, Open Data and Open Governance in the case of consultation hubs. This will help usher a more engaging and democratic policy and help Governments and other agencies to improve delivery capability, meet expectations, and increase transparency and accountability. The world is moving towards more and more transparency in actions with multilateral development agencies in the front to lead the world towards that goal. They are looking forward to publish the citizen centric data online and encourage the civil society to participate in the policy building or decision making. This activate the movement of transparent governance and collective decision making.

.....&&.....