Federal State Autonomous Educational Institution of Higher Education of St. Petersburg National Research University of Information Technologies, Mechanics and Optics

Chair of High Performance Computing

Proposal

Title: Visualization of conversation as a semantic network

St. Petersburg

2017

Table of Content

Keywords: Interactive interface, dialogue visualization, text interaction, text data networks, conversation browsing.

Summary:

This paper will present an interactive web interface for visual analysis of text conversation. Our web interface will analyze archival conversation and later tune the platform to adapt to live conversations using big data tools (Apache Spark). Our system will be designed with the help of natural language processing tools to summarize text conversations as a semantic network. The system will be designed to support both monologue and dialogue. Since our target population is the general public we design the system in such a way that people with the least education will be able to use it for their information needs.

Our focus is creating a visual representation that highlight relevant data and sentiment in conversations to help people make sense out of the social dialogue archive. We will take a huge chunk of data and visualize it so that the salient points become apparent. The next important thing we want to focus on is to develop a system that is able analyze and visualize any text conversation from any social media platform to derive sensitive information and recognize social patterns that may exit in the data in question.

The internet has predominantly piled up in volume and complexity of conversational data generated through our day to day communication, which has provided us with a rich source of private and public chunk of data left unattended. The next important thing we want to focus on is to develop a system that is able analyze and visualize any text conversation from any social media platform to derive sensitive information and recognize social patterns that that can serve any purpose in terms of security, market prediction, health and others.

Background:

In our daily lives we have conversations with each other, some lead to good actions whereas others bring terror to humanity. It is the burden of every government to intercept any act of terror before they happen. To achieve this goal governments institutions and researchers in the past had strive to develop systems that can quickly analyze text data to discover hidden information. However, there is no best approach to discover patterns in data than a graphical representation of the data.

This calls for a more robust approaches to data visualization. It’s clear that automatic summarization, visualization and the generation of semantic networks to find associations in a massive amount of text data will help uncover hidden information and quickly get a general overview of the message the data conveys.[Data cloning: Data visualisation, smoothing, confidentiality, and encryption, 2012]

Our main aim is visualization of semantic network from unrelated text conversation based on their time stamp. There are many researches on text visualization but ours will generate new conversational pattern that will help us identify the relationships in these isolated text data.[Belkaroui и др., 2016] One problem with text data is that, it becomes very hard to get a glimpse of the actual intent in the shortest possible time due to the fact that human conversation deviates from topic severally and this doesn't give the ability to get enough understanding in a shortest possible time until a thorough study is done.[Joseph Keshet David Grangier Samy Bengio, 2009]

Also According to J. Diesner and K. Carley, “semantic networks are a structured representation of knowledge that can be used for reasoning and making inferences” [Diesner, Carley, 2011]. Representing data as a semantic network requires nodes that represent concepts and edges between these concepts. The words in this concept serve as labels to the nodes and based on the relationship between two words, a connecting line called an edge is explicitly drawn to signify this relationship. This network created out of the text data is used to as a baseline to describe who did what to whom in the course of describing occurrences of events in the data, and who said what to whom in the conversational medium.[Knowledge extraction and visualization of digital design process, 2018] This can further be enhanced with metadata, such as spatial and temporal data, and attributes of nodes and links, etc. For collaborative analysis of conversations among individuals and groups, organization, management, and as far as utilizing collaborative knowledge base, by performing reasoning and inference on the network data, knowledge is gained from semantic networks. The knowledge stored and inferred from semantic networks does not have to be factually correct or logical. This is because semantic networks were designed to represent what is meant by a piece of information and it doesn’t the meaning being different from the truth, the content or the most likely interpretation of some information. Also, it depends on the data from which the network was constructed, a semantic network can range from being universal, culturally dependent, domain-specific and individual knowledge.[Knigge, Cope, 2006]

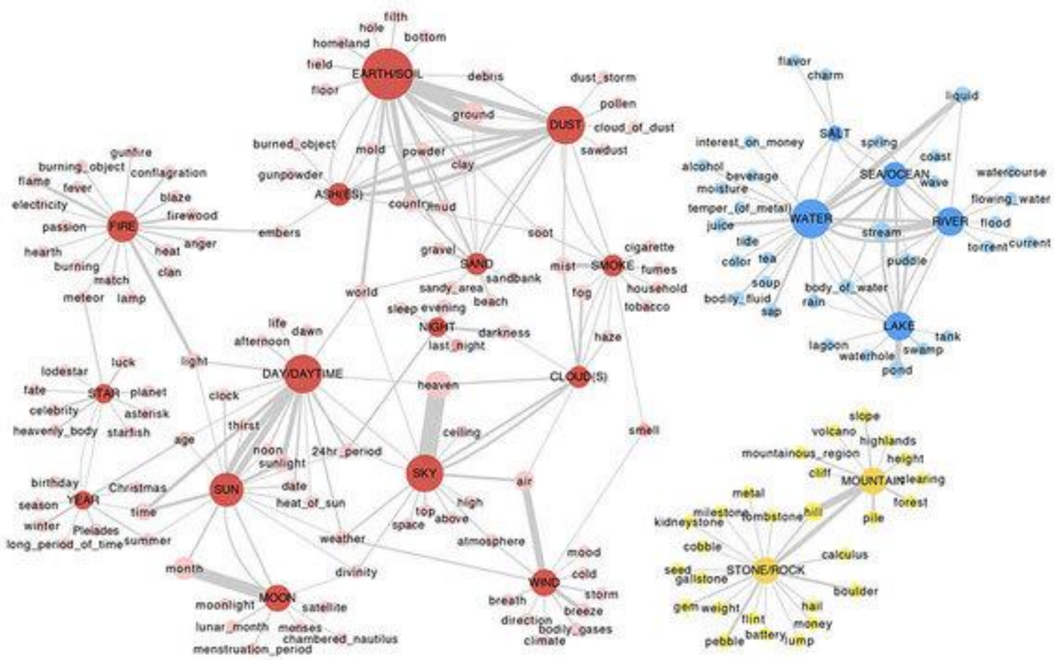


Figure 1: Retrieved from https://www.csmonitor.com/Science/2016/0203/Not-so-lost-in-translation-How-are-words-related. This figure displays a universal pattern in word meanings, revealing insight into how words change their meanings overtime.

This research is about designing graphical representation for dialogue between a simulated interactive software called Pythia which mimic the behavior of the ancient Greek high priestess (called Pythia) in the temple of Apollo (an ancient Greek god). Pythia was highly regarded for it was believed that she channeled prophecies directly from Apollo to tell people about their destiny in this world. Pythia was not just an oracle but was regarded as both a political and spiritual figure in Greek history. It was believed that prophecies from Delphi helped shape modern civilization and defined the course of history. She was also known as the Oracle of Delphi and was consulted in all major decisions before action was taken. Many rich and influential people pay a fortune to consult the oracle and it was believed that even the gods also sought advice from the oracle.

Although few people remember her and still believes her legendary powers, her wisdom continues to inspire us and teaches us that through knowing yourself, you're able to penetrate the mysteries of the past and the future.

Knowing all these inspire us and prove beyond all reasonable doubt that these kind conversation requires all time and attention for in depth research to understand people's thoughts and believes in oracles in the 21st century and also how this kind of oracle would have been perceived in the presence of modern technology and the world wide web where answers to questions are just "google click away from us".

We will aim at designing an effective symbiosis between human perceptual and cognitive skills and the computer ability to mine and summarize text to design our interface by following these three steps: Firstly, we will adopt a simplistic strategy to base the interface component in common metaphors so that the interface can be used by a large user population. Secondly, in anticipation of user differences in preference and expertise in a large population of audience which we are expecting to accommodate, we will satisfy their information needs in multiple ways, by performing different sequences of visual and interactive actions and finally, Although we are basing our case study on the Pythia dataset, we will expand our algorithm information extraction, analysis and visualization to be applicable to other works on different conversational modalities such as emails, blogs, etc.

We will test our interface on a formative user study, in which participants will use the interface to analyze long and complex conversations, in order to answer questions addressing their specific information needs. This evaluation will help us create an ideal interface that is intuitive, easy to use and provides the tools necessary for the task of a visual interface for analyzing text conversations.

Finally, we wish to discover new phrases can evolve out of the tokenized words derived from the already existing phrases.

Literature review:

Visual interface for text analysis

Our approach is not to just create charts and maps from the text as seen by many visualization tools but rather preprocess the data to draw a semantic graph which gives a different direction to the data. In "A Visual Interface for Analyzing Text Conversations" paper written by Rashid and Carenini [Rashid, Giuseppe, Ng, 2013], they presented a web interface to help users analyze long and complex conversations. Their interface design aimed at effectively combining the power of human perceptual and cognitive skills using Natural Language Processing (NLP) techniques to mine and summarize text conversations. After the evaluation, it was revealed that the interface was interactive, easy to use and provides the tools necessary for the task. Those who took part in the experiment also found components of the interface useful with the main problem from inaccuracies in the information extraction process, as well as from deficiencies in the generated summaries. In a nutshell, it seems that the choice of offering redundant functionalities was beneficial. The logged interaction behaviors reveal that different users actually selected very different strategies, even independently from their performance. And this was in line with the high variability in the user preferences for the different interface components. On the other hand, it seems that the Natural Language Processing techniques used to generate the information to the interface seemed unsatisfactory.[Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application, 2017] To be specific, half of the users were at least halfway disturbed by the inaccuracies in the classifiers that map utterances into the dialog acts they express. Furthermore, most users did not find abstractive summaries particularly useful. And abstractive summaries essentially represent the most sophisticated Natural Language Processing component of their framework.

Text analysis as a semantic Network

Another group of researchers also developed a socio-semantic network named Witology platform [Drutsa, Yavorskiy,] to visualize articles of more than 500 member but their network visualization contains more than 200 major networks members. Their visualization software can display monocot graph and bipartitegraphs visualizations. They could generate a general view as well as individual view of the graph and also visualizations of user actions like" content creation, content evaluation, content commenting, dendogram visualizations, adjacency matrix visualizations, histograms and densities, valued graph clusterization, N-gram and word extracting from user content. Despite all these functions performed by their system, they couldn't fine tune it to detect collusion and "mark up" groups of participants for many different subgraphs of the platform and also address the problem of which thresholds and for which the parameters of nodes and edges should be set[Apanovich, 2007]. They attributed this to the fact that their company is young and will try to improve their system as time goes by.

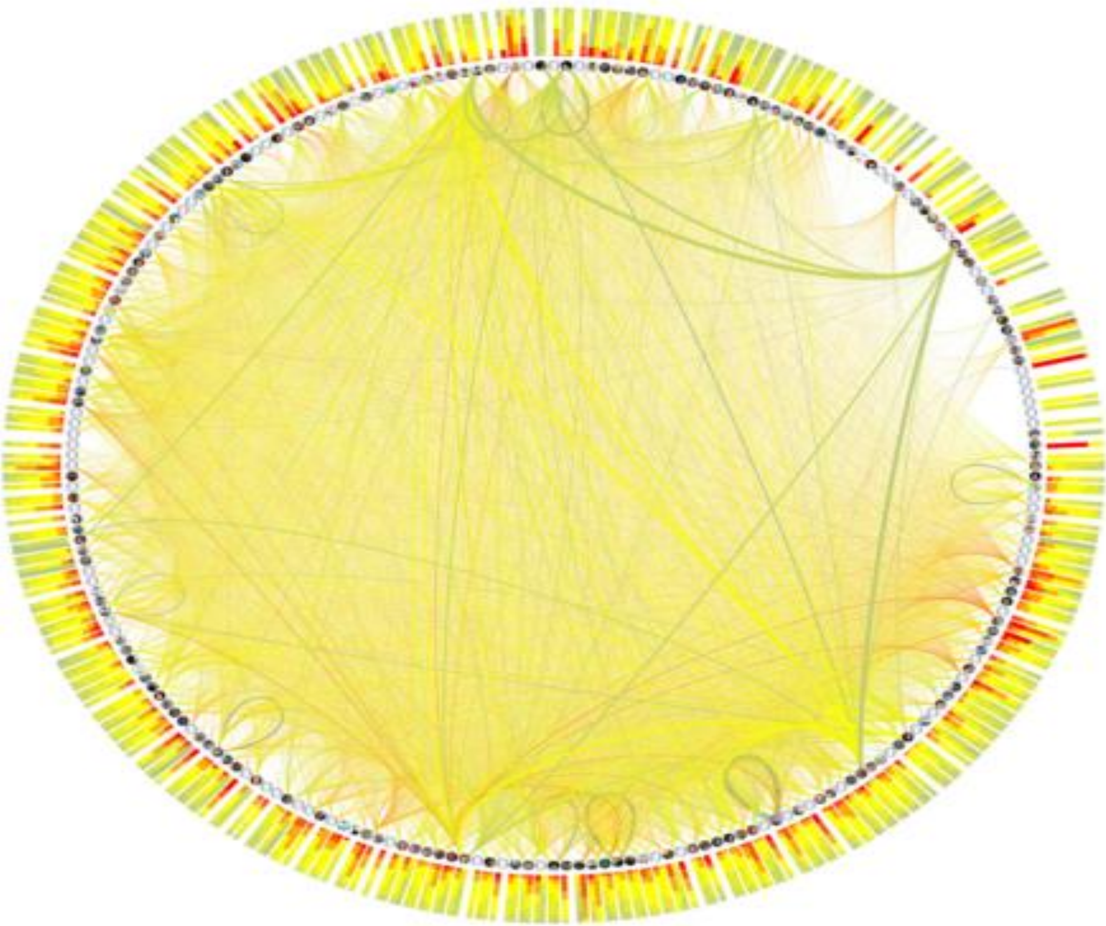


Figure 3:visualization of a local user neighborhood. Retrieved from [Drutsa, Yavorskiy,].

In semantic network, we also reviewed this paper which present practical application of semantic network visualization in the narrow context of financial indicators analysis in the light of both rational and behavioral approach[Dudycz, 2017].They assessed graphical presentations that can be used to visualize knowledge.They also designed an ontological framework to represent financial indicators analysis. Also this paper [Diesner, 2013] gave hands on tutorial on Extraction and Analysis of Semantic Network Data from Text Data. It further constructs one-mode & multi-mode semantic networks from unstructured, natural language text data, identify salient concepts from single documents and corpora and also identified several natural language processing and text mining techniques.However their analysis was based on few lines of data and presented a holistic charts and graphs from it. The bizarre and chaos that big data presents and ways to curb these problems were not addressed by this document. Although, they still solve lots of problems that the previous papers couldn't tackle, their analysis was skewed to analyzing text data in general and omitted the value of the peculiarity of bidirectional and mono-directional forms of conversation.

Implicated Meanings in text conversations

Gijeong Jang and co also took a new turn in text conversation visualization in their paper[Jang и др., 2013]. They implemented data visualization tools to look into the implicated meaning of text conversation instead of the face value meaning we all consider when analyzing text data. They stated that, implicated meaning from a conversation is important for a successful comprehension of a speaker's utterance. Under their investigation on the derivation of implicated meaning of conversations, they selected 23 participants, took them under fMRI test with a series of conversational pairs consisting of questions and answers. The expected answers from these questions were divided into three compartments namely: explicit answers, moderately implicit answers, and highly implicit answers. Participants were asked to provide answers in a yes or no format and Longer reaction time was allowed for the highly implicit answers than for the moderately implicit answers without affecting the accuracy. They displayed their results by visualizing the brain activities in the fMRI scan and also plotted the answers relating the brain activations on the question pairs (explicit answers, moderately implicit answers, and highly implicit answers.) on a line graph mapping percentage in signal change with respect to time. They evaluated Percent signal changes in brain regions such as the left anterior temporal lobe (lATL: Montreal Neurological Institute coordinate −50/8/−24), right anterior temporal lobe (rATL: 54/0/−20), left inferior frontal gyrus (LIFG: −50/26/0), and left medial prefrontal cortex (mPFC: −10/66/10). Their results and visualization presents a great deal of understanding to the conversational data and employed both quantitative and qualitative research tools. However, their choice of conversations to visualize was too myopic because yes/no conversation is not realistic. It doesn't fulfill the actual pattern of real world because only one party gets the opportunity to ask questions and the other answers given only two options.

Making sense of complex data through visualization

Diane J. Janvrin, a researcher at Iowa state university in the department of accounting also highlighted the significance of interactive visualization of text data in his journal "Making sense of complex data using interactive data visualization" [Diane J. Janvrin, Dilla, 2014]. He made an emphasis on how interactive data visualization will help accountants to assist senior management in the analysis of complex dataset. It is important that accountant become familiar with Interactive data visualization tools to help them in their day to day activities. Access to information visualization will help in the exploration of data, selection of relevant information for decision making, select presentation format for relevant data which will intern promote quick decision making based on the visualized data. He made an important point that, designing a complex visualization environment with numerous choices among variables, parameters and display format will unnecessarily overwhelm the user with too much detail to make him skip the actual intent of the visualization. In the end, simplicity is the key to a meaningful data visualization. On the contrary, depending on the nature of the data with its target audience complexity is highly subjective.

Text analysis using Artificial Neural Networks

In the works of MohammadrezaEbrahimi and friends[Mohammadreza Ebrahimia Ching Y. Suenb Olga Ormandjieva, 2016]. They utilized deep learning technologies in the analysis of text conversational data. They used Convolutional Neural Network (CNN) to find predatory conversations on social media to helps the law enforcement agencies act proactively through early detection of predatory acts in cyberspace. Although they didn't use any visualization tools to identify anomalies in text conversations on the social media, it's worth looking into. In their study, they didn't just selected CNN but rather made a comparison on the strength and weakness of other classification algorithms such as Support Vector Machine, Recurrent Neural Network the traditional Multilayer perceptron, etc. compared to the other methods, CNN was selected because of the innovative state-of-the-art results in image processing tasks and have recently gained attention in text mining[Joseph Keshet David Grangier Samy Bengio, 2009][Jürgen Schmidhuber, 2015]. On the other hand, they pointed out in their future works that, Recurrent Neural Networks and LSTMs has shown to be the most efficient neural network model in text analysis and speech recognition but due to the difficulty of their applicability during training the model, it makes it less practical when dealing with large documents. If that problem is solved, it will be the best model of choice in such kinds of task. Without any visualization tools, their system was able to analyze text conversation to discover hidden pattern to raise red-flags on them. On the contrary, this kind of system can do just one thing and nothing more. It cannot display a social network of patterns. It can only identify outliers in data which is doesn't form a holistic interpretation of conversation. It can only serve it purpose as an emergency response agent to catastrophes.

Automatic Platform for text and sentiment Analysis

According to Ahmed K. [Kamal, 2015], opinion mining and visualization has mass adoption in business intelligence, product recommendation, targeted marketing, etc. It success in these fields has fascinated many researchers and there are always lots of researches ongoing to continue to improve its impact. They pointed out that, "although opinion mining and sentiment analysis is somehow automated with most of them attempting to perform document-level sentiment analysis, classifying a review document as *positive*, *negative*, or *neutral"*. Moreover, Such

document-level opinion mining approaches fail to provide insight about users' sentiment on individual features of a product or service. So they recommended that if the review could be in a fine-grained manner where users individual opinions presented in a visual summarized. They presented a unified opinion mining and sentiment analysis framework which uses machine learning technologies to preprocess the data and modern visualization tools to visually present it for decision making.

However, their idea of automated opinion summarization and sentiment analysis system was just a "black board project". There was no actual practical work done by them to see its feasibility. As it usually goes that, is easier said than done. They didn't complete a single prototype to actually confirm the possibility of such as system than a mere talk.

Pattern of conversation in social media

In this 21st century, majority of our day to day interactions and conversation happens on social media and it is prudent to analyze the information flow on social media to ascertain facts that will help our daily lives. Social media also represent a radical revolution through which companies refine the manner in which they conduct business [Pace, Buzzanca, Fratocchi, 2016]. According to George Kuk, “The typical form of online interaction is that of conversations through which actors create knowledge and value” [Kuk, 2006]. So there is the need to understand the structure of these information to improve communication strategies and user experience. This paper argued that conversations in social media takes two forms: dialogic which in the form of horizontal interactions among peers and dialectic, that exclusively takes vertical interaction with the source of an input, such as the company’s comment initiating the thread and individuals expressing their opinions on the stated topic. For instance, a company can publish a prototype of their new product on Facebook or twitter to request for the views of the general public. The aim of this paper is to expose how organizations can effectively and efficiently exploit social media conversations by monitoring the structure of online conversations. They focused their study on the network structure of these conversations, while less is known about the user preferences, identifying the underlying network will help us understand the twists and turns of the conversational structure. They opined that, by analyzing both the network structure of a conversation and the perception of online users, they can contribute to advance the knowledge on this gap by studying the preference of online users towards a dialogic vs. a dialectic structure of conversations. They employed empirical research methods for their analysis and discovered that, dialogic conversations are richer than dialectic conversations because they engage users in interactions among themselves rather than simple interactions with the initial content. They added that, most of dialectic conversations are in the form of likes, retweets or claps which doesn’t reflect the true nature of user’s intents and feelings. On the other hand, it is not always true that the format of conversation by companies are dialectic, companies nowadays engage in active one-on-one conversations with their clients to seek their concerns and address them accordingly. Companies have also created support platforms on social media just to cater for the needs of their customers but examining the conversational structure will help understand the routine needs of clients to provide answers even before they ask. Shailendra R. also took the same approach to study text conversation through visualization and semantic networks but the aim was to address the security concerns that social media poses individuals and companies [Rathore, 2017]. He described the threads into multimedia content threats, traditional threats and social threads. He further conducted analysis of the possible and existing schemes for protecting social media network users and presented some easy-to-apply response techniques that can be easily followed by users of this media to better protect themselves against various security threats.

Another school of thought[Weiler, Grossniklaus, Scholl, 2014] also leveraged on the benefit of visualization to observe situation in cities. It's emphatically true that the best place to find societal events is the social media. So they created a visualization platform to monitor real-time situations in societies via twitter feeds. They decided to monitor live tweets in the various states in USA. They identified that, users are not able to follow all topics due to high volume of data flow and summarized visualization platform will help keep users up to date with all that happens in the social media. Their aim was to create a platform to monitor abnormal behaviors in a city and its surroundings filtered by geographical location and situation keywords with live feeds from twitter. They also used sentiment analysis to further analyze the emotions in the tweets with red and green colors signifying negative and positive sentiments on the graph in a particular geographical area. The selected Boston as their use case and concluded that the live observation visualization can support local people and news reporters in getting up-to-date information about the state of emotion and topics in and about a city.

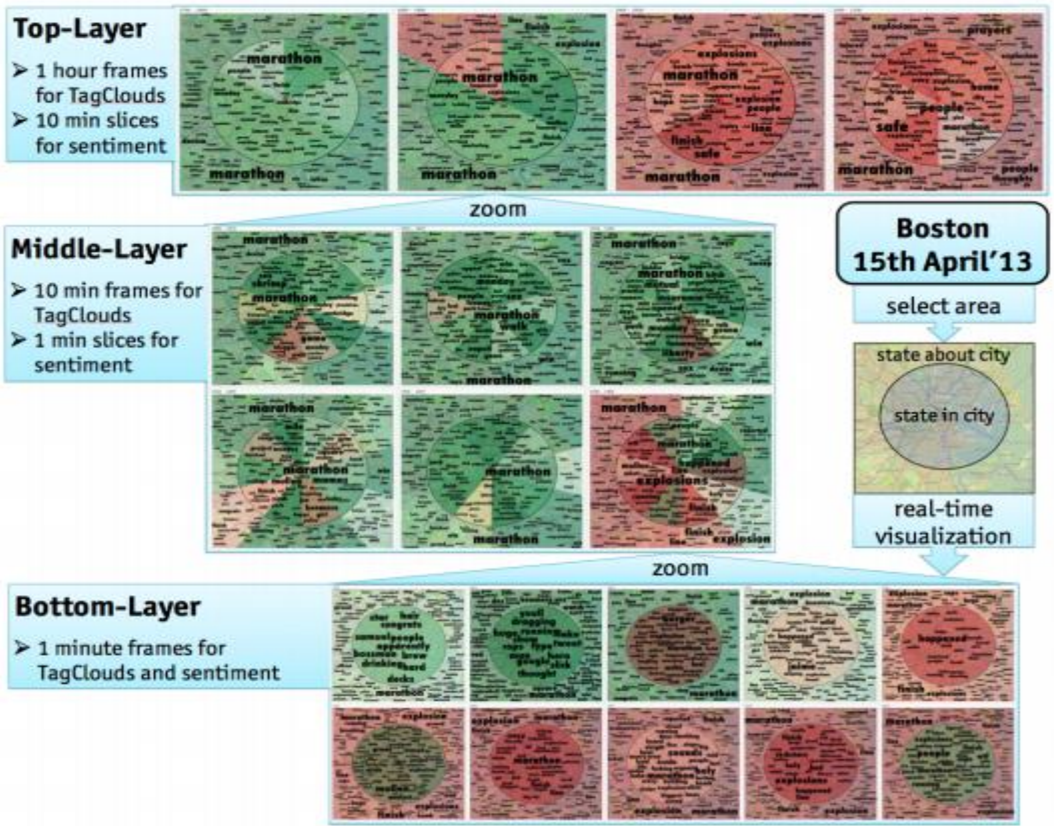


Figure 4: retrieved from [Weiler, Grossniklaus, Scholl, 2014]. *Live tweeter feed in the city of Boston on 15th April 2013, from 5PM to 9PM (UTC).*

Although it seems there were no flaws in their work, there were some negligence they portrayed which can have a negative effect on the results. They didn't take into account of false alarms and also didn't give any in-depth explanation on how their model was trained on classifying messages for positive and negative sentiments.

Methodology:

The sources of literature for this project was selected from science direct database. This database is well known for high end repository of peer reviewed literature for scientific and medical research. The data was gathered from an interactive installation called Pythia through google speech API. [http://www.sonicartist.me/wp/pythia/] It is an interactive software that analyze audio speech into a prophetic message. This was designed to mimic the ancient Greece Goddess of Apollo. By the use of microphone, people's voice was recorded as they talk to the "oracle". The data collected covers a specific time frame from Jan 2015 to December 2015. It is a text data with time stamps but due to noise and other interferences we might have other data element which does not match up to the intended data output. The dataset was recorded and saved in a csv file format with 7 fields (Number of Records (bin); Number of Records; Client; Id; Lang; Text; Time) and 161,533 records.

We analyzed the frequency variations in the phrases and attained a mean value 2.7 and standard deviation of 16.6. the p value was less than 0.01 which signifies that the data is not normally distributed.

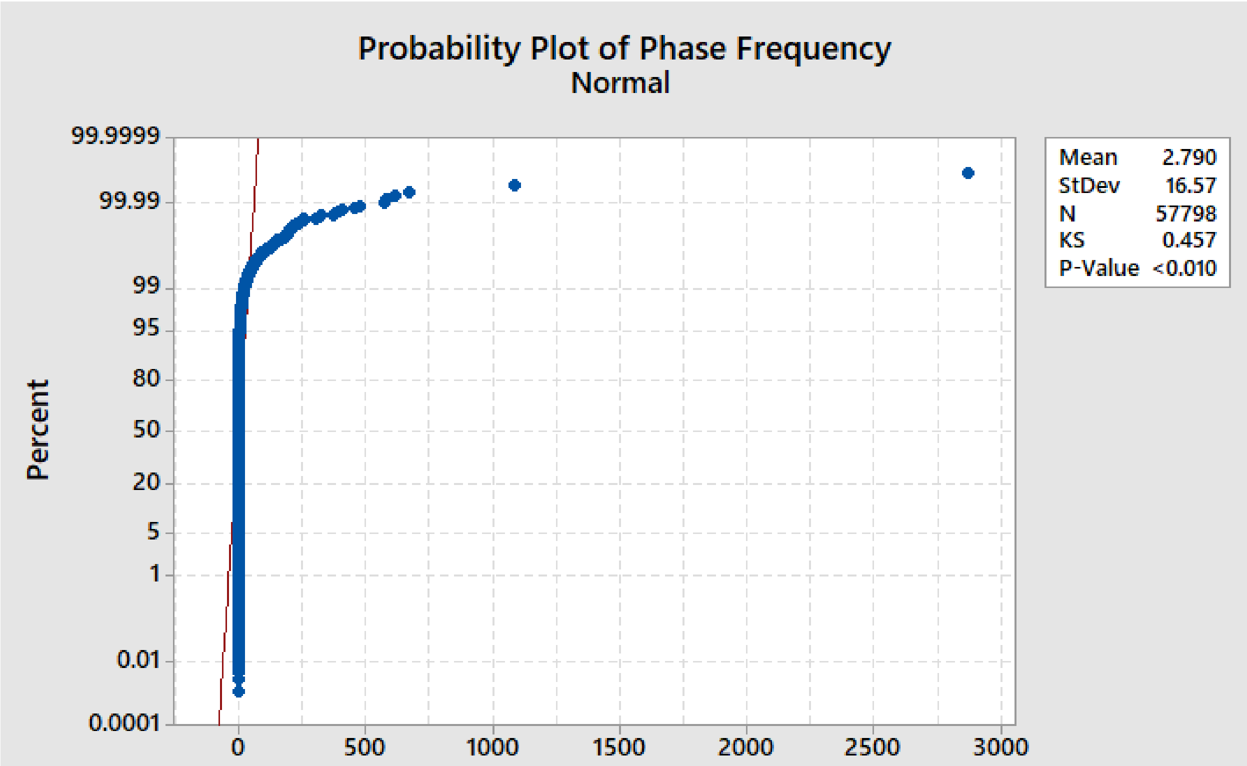


Figure 5: frequency distribution of phrase in the dataset.

The figure above demonstrates that about 99% of the have appearance between 1 and 20 leaving only 1% of the phrase having greater representation. An outlier was discovered with one phrase appearing almost 3000 times in the dataset.

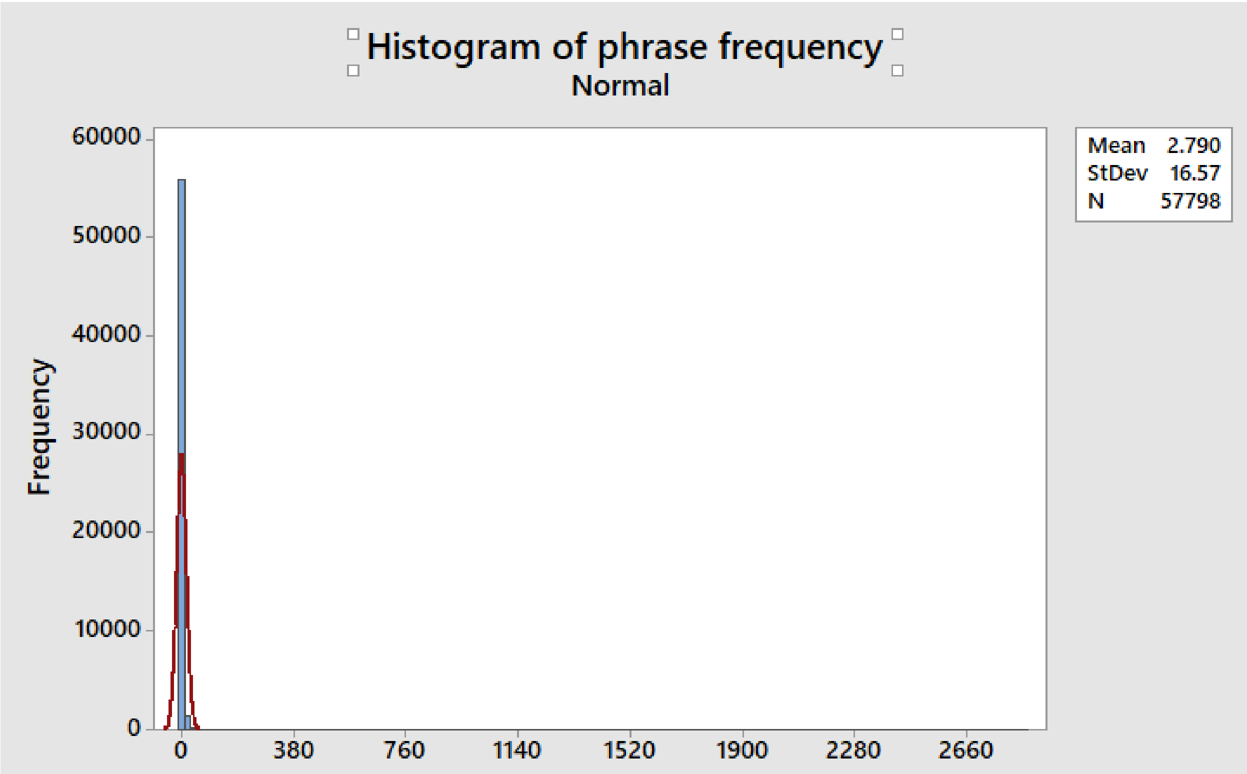


Figure 6: Histogram of Phrase frequency distribution.

Since the histogram displays a sharp left skewness, we can confidently say that there is a high amount of unique phrases in the dataset

Our target aim allows us to select only the "text" and "Time" fields from the dataset. Jupyter Notebook, Matplotlib ("Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, IPython shell, the jupyter notebook and web application servers"), Numpy (" is the fundamental package for scientific computing with Python") will be used in conjunction with pandas("*pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.")' library in python for data preprocessing task to remove invalid fields and erroneous records from the dataset. We will also remove words repeating words in the records and fields of the dataset. E.g. (привет, привет, привет, привет, привет). Also, integers and special characters will be considered as invalid and removed from the dataset. After the cleaning of the data, the data will be clusterized using sklearn clustering cluster algorithms. Finally, on the data preprocessing we will group the phrases into word pairs. In the first place, with the help of natural language toolkit word tokenizer, we will tokenize each phrase on their respective rows and carefully taking into consideration the order in which the texts were recorded. Since our main task is to just add connecting arrows to the already existing words that make up each phrase and further observe the branches to discover the interconnectivity between one phrase to another and new phrases that will emerge out of this.

We will take the whole dataset as input into a function and iterate over each row with a sliding window of two. The pairs are then stored into a dictionary and a default weight 0.1 is add to each pair. If a pair is repeated in the dataset, the weight is increased by 0.1. Rows which contain single word will be skipped since we cannot form pairs with them. The cluster will help us color the nodes of the network based on the group they find themselves.

we will use ajax lazy loading to load our data into the d3 JavaScript library ("D3.js is a JavaScript library for manipulating documents based on data. **D3** helps you bring data to life using HTML, SVG, and CSS"). We will use Node.js as a backend server("Node.js is a platform for building fast and scalable server applications using JavaScript. Node.js is the runtime and NPM is the Package Manager for Node.js modules.") and Express("is a web application framework for Node.js, released as free and open-source software under the MIT License. It is designed for building web applications and APIs") with Body Parser middleware(Basically it's a middleware for parsing JSON, plain text, or just returning a raw Buffer object for you to deal with as you require.) to serve files from the server.

After exploring all these technologies, we will design an interactive website that support user queries, allows for testing and criticisms from our target group through user feedback questionnaires and base on that to fine-tune the system.

Expected Results:

We expect to create an ontological view that provides a systematic way for the users to explore the relevant concepts in the conversation and their relations. It will contain tree hierarchy with core nodes as the most frequent words down to the least frequent words. The root node in the ontological tree will represent all the sentences in the conversation while the other nodes will be subset or subclass of those sentences that satisfy a particular property. Example is shown in the figure below:

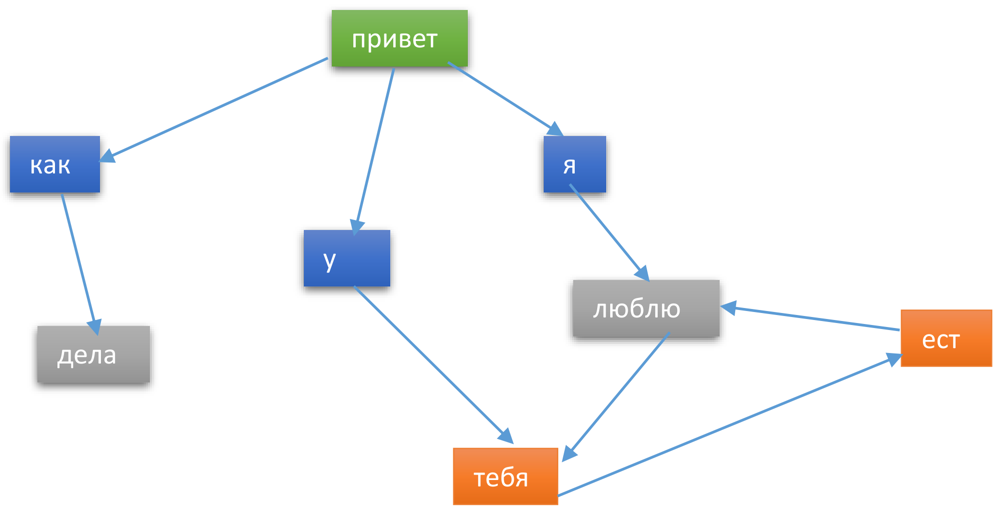


Figure 7: Sample of ontological view.

Secondly, an entity view will be created as a textual collage of a list of entities mentioned in the conversation represented as a frequency term Matrix with actual counts in brackets. This view will also pose as a lookup table to the actual phrase or conversation and also a quick overview of the content of the whole conversation. The user can search portions of the conversations or sentences mentioning any subset of the displayed entities by simply selecting that subset.

A third view called transcript view will display original phrase in the network with specific color (e.g.: blue) when clicked and other evolving phrases that has been generated due to their interconnectivity in the network (e.g.: yellow). If entities are selected, the main node and its tributaries will be highlighted with their corresponding color described earlier in this paragraph.

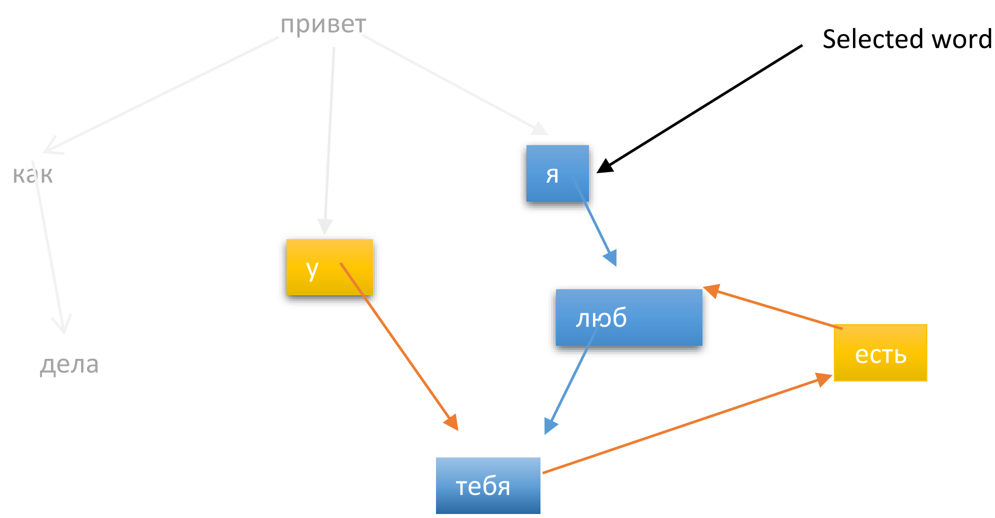
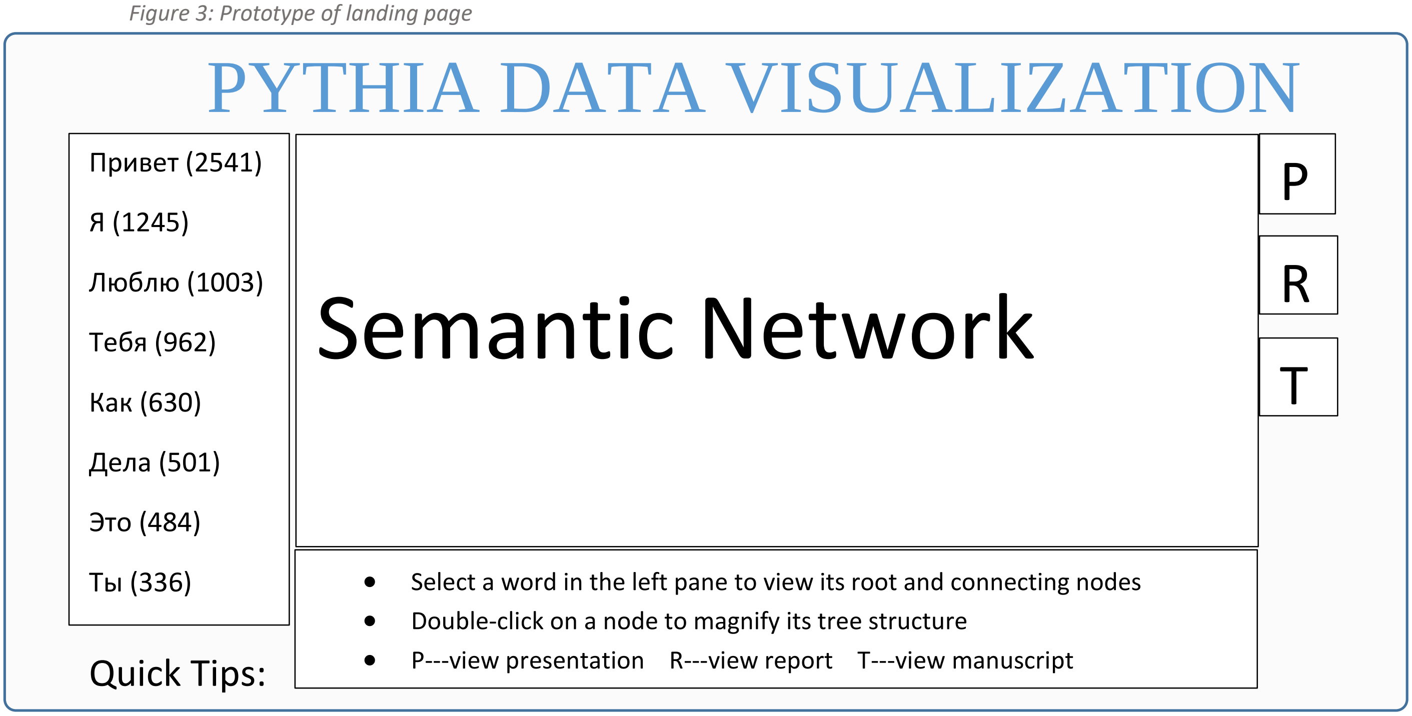


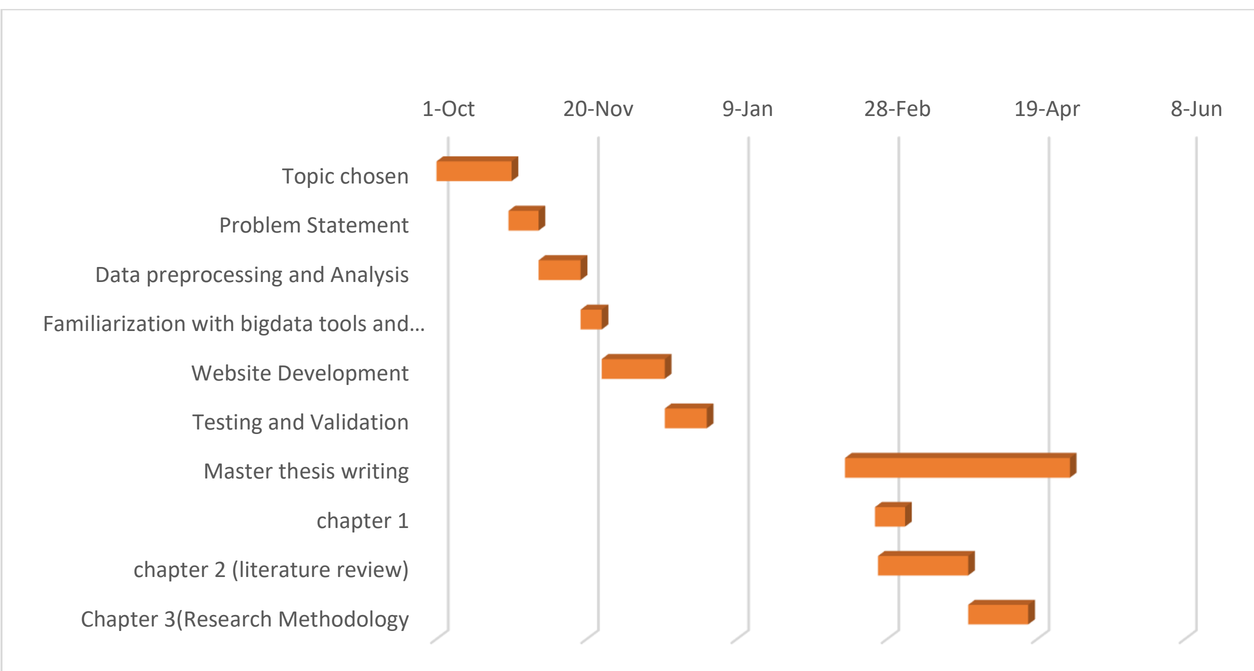
Figure 8: Sample of Transcript view.

Finally, there will be a summary view which will consist of the complete semantic network of disjointed words connected with arrows showing the direction and order of precedencies. This network of related words will be interactive and support user queries. We expect the interactive web app to be a single page app with links to all other views displaying in a popup window. Upon first entry into the page, the user will be welcomed with the summary page and the transcript view as a side navigation pane. We will also provide tooltips, legends and keys to help users navigate their way through our interactive web visualization platform for his information needs.

Figure 3: Prototype of landing page



Task	start date	duration
Topic chosen	1-Oct	25
Problem Statement	25-Oct	10
Data preprocessing and Analysis	4-Nov	14
Familiarization with big data tools and technologies	18-Nov	7
Website Development	25-Nov	21
Testing and Validation	16-Dec	14
Master thesis writing	14-Feb-18	75
chapter 1	24-Feb-18	10
chapter 2 (literature review)	25-Feb-18	30
Chapter 3(Research Methodology	27-Mar-18	20
Chapter 4(discussions)	16-Apr-18	8
Chapter 5(recommendations)	25-Apr-18	7



Bibliography

1. Apanovich Z. V. FROM GRAPH DRAWING TOWARDS INFORMATION VISUALIZATION // 2007.

2. Belkaroui R. и др. Conversation Analysis on Social Networking Sites To cite this version : Conversation Analysis on Social Networking Sites // 2016.

3. Diane J. Janvrin R.L.R., Dilla W.N. Making sense of complex data using interactive data visualization // J. Account. Educ. 2014. T. 32. № 4. C. 31–48.

4. Diesner J. Hands-onTutorial: From Words to Networks: Extraction and Analysis of Semantic Network Data from Text data // 2013. № May.

5. Drutsa A., Yavorskiy K. Socio-semantic network data visualization C. 1–7.

6. Dudycz H. Application of semantic network visualization as a managerial support instrument in financial analyses Rational vs . Behavioral Approach to Decision Making // 2017. T. 5. № 1. C. 112–128.

7. Jang G. и др. Everyday conversation requires cognitive inference: Neural bases of comprehending implicated meanings in conversations // Neuroimage. 2013. T. 81. C. 61–72.

8. Joseph Keshet David Grangier Samy Bengio. Discriminative keyword spotting // Speech Commun. 2009. T. 51. № 4. C. 317–329.

9. Jürgen Schmidhuber. Deep learning in neural networks: An overview // Neural Networks. 2015. T. 61. C. 85–117.

10. Kamal A. Review Mining for Feature Based Opinion Summarization and Visualizationzation // arXiv1504.03068 [cs]. 2015. C. 1–6.

11. Knigge L., Cope M. Grounded Visualization: Integrating the Analysis of Qualitative and Quantitative Data through Grounded Theory and Visualization // Environ. Plan. A. 2006. T. 38. № 11. C. 2021–2037.

12. Kuk G. Strategic interaction and knowledge sharing in open source software communities // Manage. Sci. 2006. T. 52. C. 1031–1042.

13. Mohammadreza Ebrahimia Ching Y. Suenb Olga Ormandjieva. Detecting predatory conversations in social media by deep Convolutional Neural Networks // Digit. Investig. 2016. T. 18. C. 33–49.

14. Pace S., Buzzanca S., Fratocchi L. The structure of conversations on social networks: Between dialogic and dialectic threads // Int. J. Inf. Manage. 2016. T. 36. № 6partA. C. 1144–1151.

15. Rashid S., Giuseppe C., Ng R. A Visual Interface for Analyzing Text Conversations // Enabling Real-Time Bus. Intell. 2013. T. LNBIP 154. C. 93–108.

16. Rathore S. Social Network Security : Issues , Challenges , Threats , and Solutions // 2017.

17. Weiler D.A., Grossniklaus M., Scholl M.H. SiCi explorer: Situation monitoring of cities in social media streaming data // CEUR Workshop Proc. 2014. T. 1133. C. 369–370.

18. Data cloning: Data visualisation, smoothing, confidentiality, and encryption // J. Stat. Plan. Inference. 2012. T. 142. № 2. C. 410–422.

19. Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application // Technol. Forecast. Soc. Change. 2017. T. 115. C. 210–220.

20. Knowledge extraction and visualization of digital design process // Expert Syst. Appl. 2018. T. 92. C. 206–215.

21. Kaidi Z. Data visualization C. 1–15.

22. Paper B. Data visualisation for researchers and scholars C. 1–5.

23. SearchBusinessAnalytics. (2017). *What is data visualization? - Definition from WhatIs.com*. [online] Available at: http://searchbusinessanalytics.techtarget.com/definition/data-visualization [Accessed 15 Aug. 2017].

24. UW Interactive Data Lab | Papers. (2017). Idl.cs.washington.edu. [online] Available at: https://idl.cs.washington.edu/papers/reactive-vega

25. Python Data Analysis Library — pandas: Python Data Analysis Library. Pandas.pydata.org [online]. Available at: http://pandas.pydata.org/