

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

Incremental Short Text Summarization

Switee Zambre¹, Ankita Gavali¹, Devika Kulkarni¹, Anusree Manamkottu¹

B.E Student, Department of Computer Engineering, Zeal College of Engineering and Research, Pune, India¹

ABSTRACT: Now-a-Days, the popularity of social networking services has increased tremendously, so the quantity of comments can increase at a high rate immediately after a social message is published. Thus the system focuses on the problem of short text summarization on the comment stream of a particular message from social network services. The users of the social sites always desire to get a brief understanding of a comment stream without reading the whole comment list. So this system attempts to group comments with similar content together and generate a concise opinion summary for the message. Various type of users can request the summary at any moment, current clustering methods cannot be directly applied because they cannot meet the real-time need of such application. So this comment stream summarization problem is modeled as incremental clustering problem. This approach can incrementally update clustering results with latest incoming comments in real time. As a result visualization interface is generated that help users to rapidly get an overview summary.

KEYWORDS: Comment streams, Comment summarization, Incremental clustering, Real-time short text summarization, Social network services.

I. INTRODUCTION

The social network provides a set of methods for studying the structure of whole social items as well as a variety of theories explaining the patterns observed. Due to the popularity and convenience of these platforms, celebrities, corporations, and organizations also set up social page: to interact with their fans and the public. For each message, users are able to give their opinions by forwarding, liking, and leaving comments on it. Not only the quantity of comments is huge, but also the generation rate is remarkably high.

Users unnecessarily and almost impossibly go over the whole comment list.

Users unnecessarily and almost impossibly go over the whole comment list of each message. However, we may still desire to know what are they talking about and what are the opinions of these discussion participants. Moreover, celebrities and corporations will have high interest to understand how their fans and customers reacting to certain topics and content.

With this as inspiration, we are inspired to develop an advanced summarization technique targeting at comment streams in social network services. Numerous studies and systems have proposed techniques to generate different types of summaries on comment streams. One major type aims to segregate comments from messy discussion.

Our proposed system does not focus on traditional comment streams that usually express more complete information. such as the discussion on products or movies. We consider the comment streams in SNS that are it short text style with casual language usage. Our main motive is to cluster comments with similar content together and generate a precise summary for this message. We want to discover how many different group opinions exist and provide an overview of each group to make users easily and rapidly understand.

For instance, when lady Gaga uploads a photo to SNS, there are hundreds and thousands of comments given by her fans during a short period of time. Some of them may say that she is very beautiful and another group of fans may think that the outfit is too weird. Even more, some may particularly discuss the hair style of this photo. Our goal is developing an efficient and effective technique to identify the clusters of these comments. Note that this problem is clearly different from existing research and possesses numerous unique characteristics and challenges.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

II. RELATED WORK

Sr no	Author, Title and Journal Name	Advantages	Disadvantage
1	H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in Proc. 3rd ACM Int. Conf. Web Search Data Mining, 2010, pp. 291–300.	<ol style="list-style-type: none"> 1. The advantages of this modification to the ensemble similarity learning technique include improved efficiency via the use of centroids, providing for a more direct similarity metric computation. 2. To produce high quality clustering results. 3. The similarity metric learning techniques yield better performance than the baselines 	<ol style="list-style-type: none"> 1. The work focuses on event documents only not non documents. 2. It does not summarize event contents.
2	H. Becker, M. Naaman, and L. Gravano, "Selecting quality twitter content for events," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 442–445.	<ol style="list-style-type: none"> 1. To communicate Twitter content effectively. 2. The re-ranking techniques that boost the centrality score of tweets with potentially useful features (e.g., URLs, tags). 2. Automatically select event messages that are both relevant and useful. 	<ol style="list-style-type: none"> 1. The sub-event content and topic analysis, such that multiple views or temporal variations not represent in the data.
3	M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: Interactive topic-based browsing of social status streams," in Proc. 23rd Annu. ACM Symp. User Interface Softw. Technol., 2010, pp. 303–312.	<ol style="list-style-type: none"> 1. A simple, novel topic detection algorithm that uses noun-phrase detection and a search engine as an external knowledge base. 2. Eddi is more enjoyable and more efficient to browse than the traditional chronological Twitter interface. 	<ol style="list-style-type: none"> 1. Users had access to our clients for a limited time, making it difficult to extrapolate conclusions on how the tool might be used longitudinally. 2. Users used to view the history of their feed rather than tweets they had never seen before, making our task slightly less realistic.
4	J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 450–453.	<ol style="list-style-type: none"> 1. The events in the social, political, cultural and economical sphere do have an important, immediate and very specific effect on the various dimensions of public mood. 2. Long-term changes seem to have a more gradual and cumulative effect. 3. To speculate that the social network of Twitter may highly 	<ol style="list-style-type: none"> 1. Minute texts such as microblogs may pose particular challenges for this approach.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

		affect the dynamics of public sentiment. 4. Sentiment analysis techniques rooted in machine learning yield accurate classification results.	
5	D. Chakrabarti and K. Punera, "Event summarization using tweets," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 66–73.	1. The advantage of leveraging existing query matching technologies and for simple one-shot events . 2. The HMM is able to learn differences in language models of sub-events completely.	1. The disadvantage that SUMMHMM has to account for tweet words that only occur in some of the events, but not in others.

III. EXISTING SYSTEM

In social network consist of problems during when a user like Lady Gaga uploads a photo to social network services; there are hundreds and thousands of comments given by her fans during a short period of time. Some of them may say that she is very beautiful, and another group of fans may think that the outfit is too weird. Even more, some may particularly discuss the hair style of this photo. Therefore, our goal is developing an efficient and effective technique to identify the clusters of these comments.

The quantity of comments may increase at a high rate right after a social message is published. Moreover, distinct users will request the summary result at any moment. So, to immediately get a summary based on the existing comments, an incremental approach is preferred.

The comments in SNS are usually short, and users widely make use of informal and unstructured texts that contain acronyms, shortening words, etc. This phenomenon increases the difficulty of determining the similarity between comments. On the other hand, it is worth mentioning that instead of emphasizing on the quality of clustering, the most crucial point of this task is to produce a general summary so that users can easily get the overview of a comment stream. It can be perceived that this problem is able to be modeled as a clustering task. However, traditional clustering methods have several inherent restrictions that cannot be directly applied here.

First, the computational complexity of existing methods is high, and they cannot straightforwardly adapt to satisfy the incremental need. Moreover, in this problem, defining the number of desired clusters in advance is unreasonable, which is required in many clustering algorithms.

Disadvantages:

1. The difficulty of determining the similarity between comments is more.
2. Cluster identification is difficult one.
3. Computational complexity is high.
4. Due to sparse information contained in each comment, these approaches are not suitable for short text summarization especially when the number of comments is not large.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

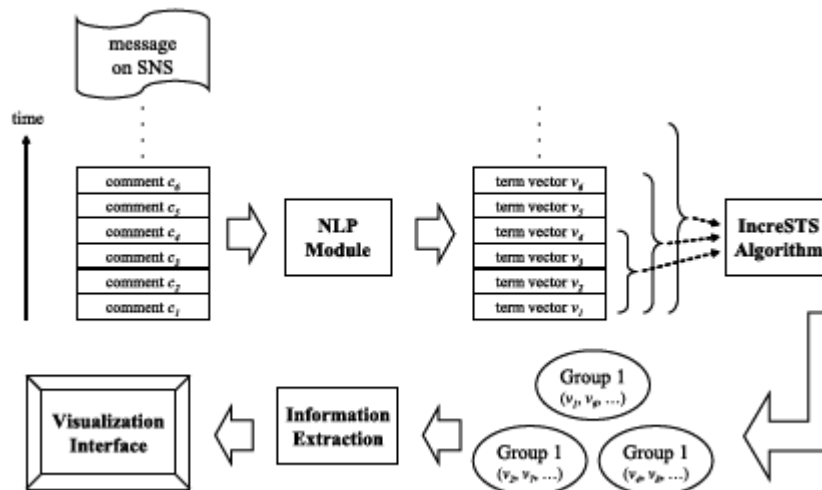


Fig.1 System model of the existing framework.

IV. PROPOSED SYSTEM

In this paper, explore the problem of incremental short text summarization on comment streams from social network services. To model this problem as an incremental clustering task and propose the IncreSTS (standing for Incremental Short Text Summarization) algorithm to discover the top-k clusters including different groups of opinions towards one social message. For every comment cluster, important and same terms will be extracted to construct a key-term cloud. This key-term cloud provides an overview of presentation that users can easily and quickly understand the main points of similar comments.

Representative comments in each group will also be identified. Our objective is to generate an informative, concise, and impressive interface that can help users get an overview understanding without reading all comments. Our proposed system depends on a fully incremental algorithm that is almost parameter-free and can handle the outlier problem. The most significant advantage of our algorithm is its high efficiency, indicating that it can generate clustering results with latest incoming comments in real time. These capabilities certainly meet the need of comment stream summarization on SNS.

To verify the effectiveness of IncreSTS algorithm, to collect real comment streams from Facebook and conduct extensive experiments with comparative methods to show the strength and superiority of our approach. Overall, the contributions of this paper can be summarized as follows.

- To model a novel incremental clustering problem based on the requirements of comment stream summarization on SNS.
- To propose IncreSTS algorithm that can incrementally update clustering results with latest incoming comments in real time.
- To design a representation which is precise and informative help users easily and quickly get an overview understanding of a comment stream.

Advantages:

- I. High efficiency indicating that it can generate clustering results with latest incoming comments in real time.
- II. High scalability and better handling outliers.
- III. Justifies the practicability of Incremental Short Text Summarization on the target problem.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

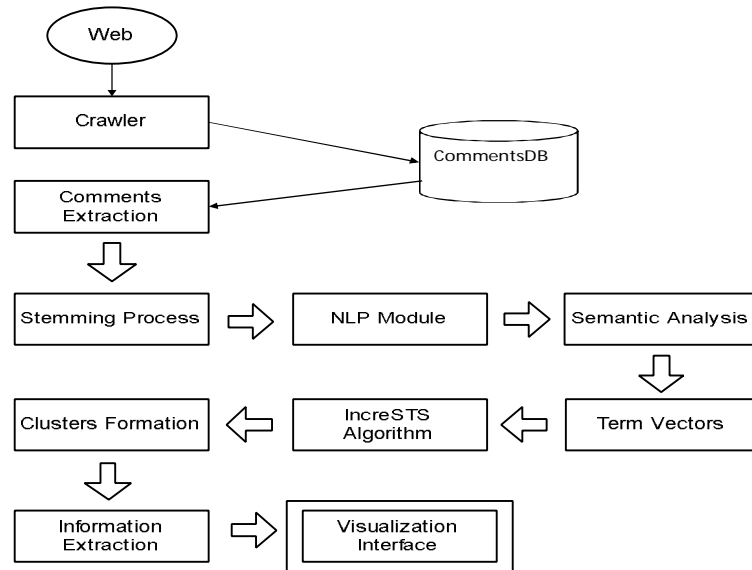
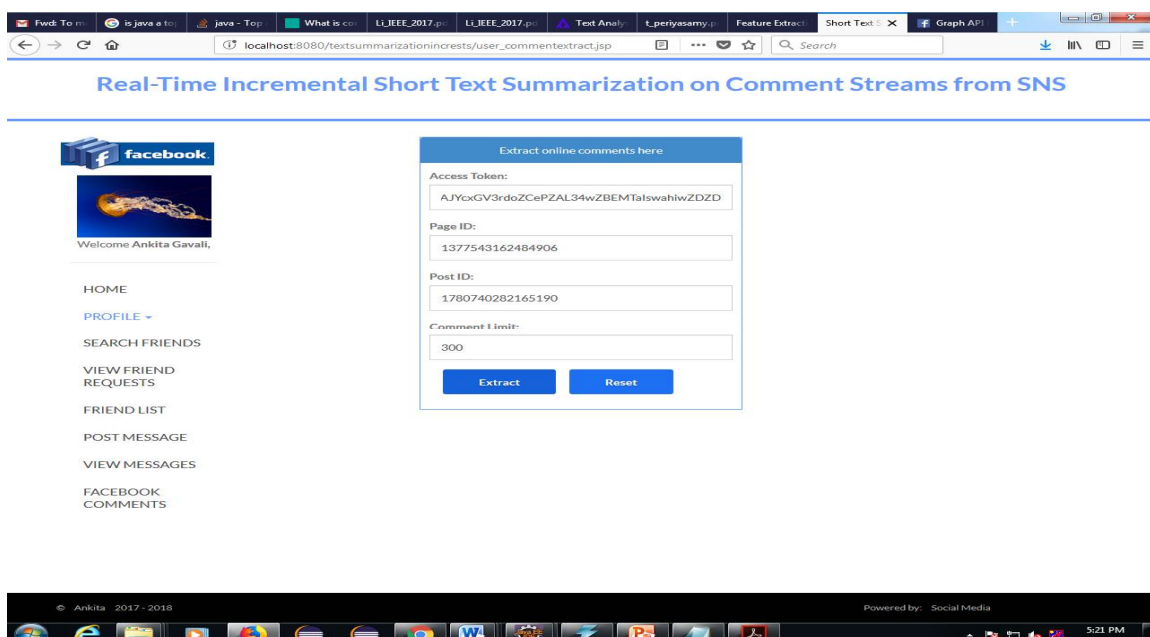


Fig.2 System model of the proposed framework

V. RESULTS

1) Facebook comments extracts using Facebook API



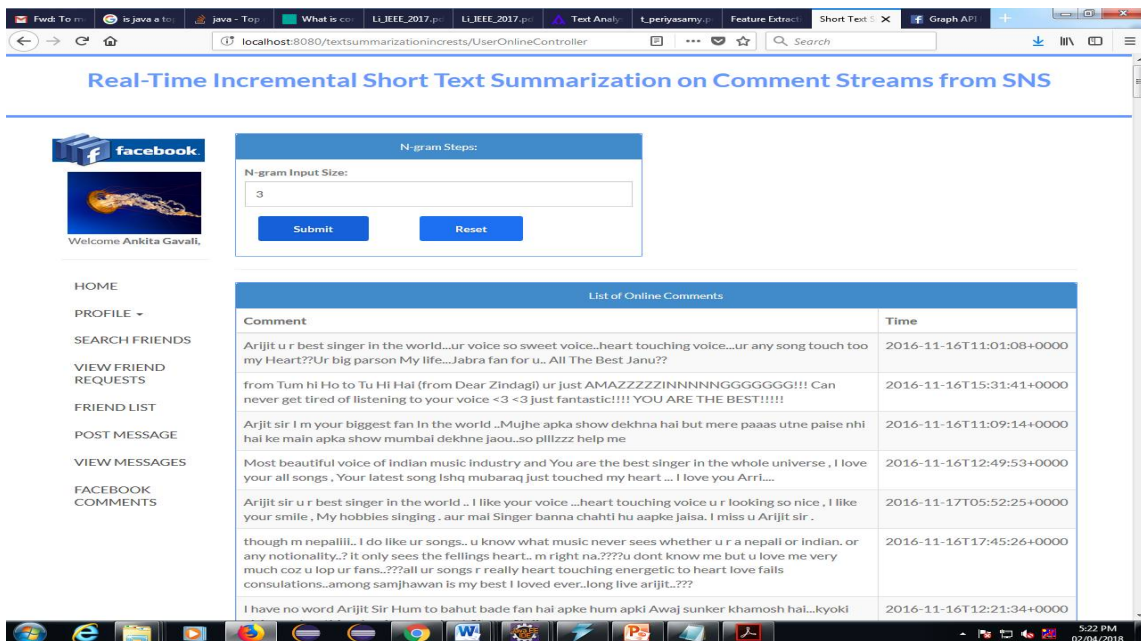
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

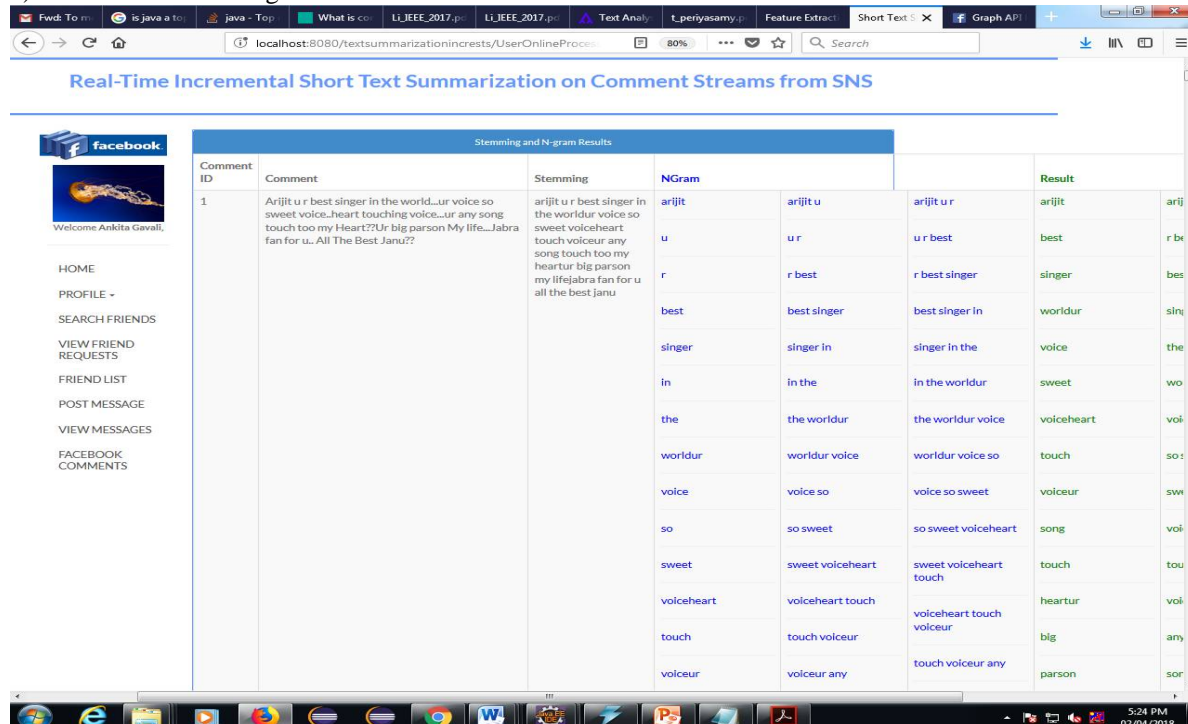
Vol. 7, Issue 5, May 2018

2) View Comments Page



Comment	Time
Arijit u r best singer in the world...ur voice so sweet voice...heart touching voice...ur any song touch too my Heart??Ur big parson My life...Jabra fan for u... All The Best Janu??	2016-11-16T11:01:08+0000
from Tum hi Ho to Tu Hi Hai (from Dear Zindagi) ur just AMAZZZZZINNNNNNGGGGGG!!! Can never get tired of listening to your voice <3 <3 just fantastic!!!! YOU ARE THE BEST!!!!	2016-11-16T15:31:41+0000
Arjit sir I m your biggest fan In the world ..Mujhe apka show dekhna hai but mere paaas utne paise nhi hai ke main apka show mumbai dekhne jauu..so plllzzz help me	2016-11-16T11:09:14+0000
Most beautiful voice of Indian music industry and You are the best singer in the whole universe , I love your all songs , Your latest song Ishq nubaraq just touched my heart ... I love you ArriL...	2016-11-16T12:49:53+0000
Arijit sir u r best singer in the world .. I like your voice ...heart touching voice u r looking so nice , I like your smile , My hobbies singing , aur mai Singer banna chahti hu aapke jalsa. I miss u Arijit sir .	2016-11-17T05:52:25+0000
though m nepaliii.. I do like ur songs.. u know what music never sees whether u r a nepali or Indian, or any notionality..? It only sees the fellings heart.. m right na????u dont know me but u love me very much coz u lop ur fans..???all ur songs r really heart touching energetic to heart love falls consultations...among samjhawan is my best I loved ever..long live arijit..???	2016-11-16T17:45:26+0000
I have no word Arijit Sir Hum to bahut bade fan hai apke hum apki Awaj sunker khamosh hai...kyoki	2016-11-16T12:21:34+0000

3) N-Gram Results Page



Comment ID	Comment	Stemming	N-Gram	Result	
1	Arijit u r best singer in the world...ur voice so sweet voice...heart touching voice...ur any song touch too my Heart??Ur big parson My life...Jabra fan for u... All The Best Janu??	arijit u r best singer in the worldur voice so sweet voiceheart touch voiceur any song touch too my heartur big parson my lifejabra fan for u all the best Janu	arijit u r best singer in the worldur voice so sweet voiceheart touch voiceur any	arijit u r best singer best singer in singer in the in the worldur the worldur voice worldur voice so voice so sweet so sweet sweet voiceheart voiceheart touch touch touch voiceur touch voiceur any	arijit best singer worldur the sweet voiceheart touch heartur big parson

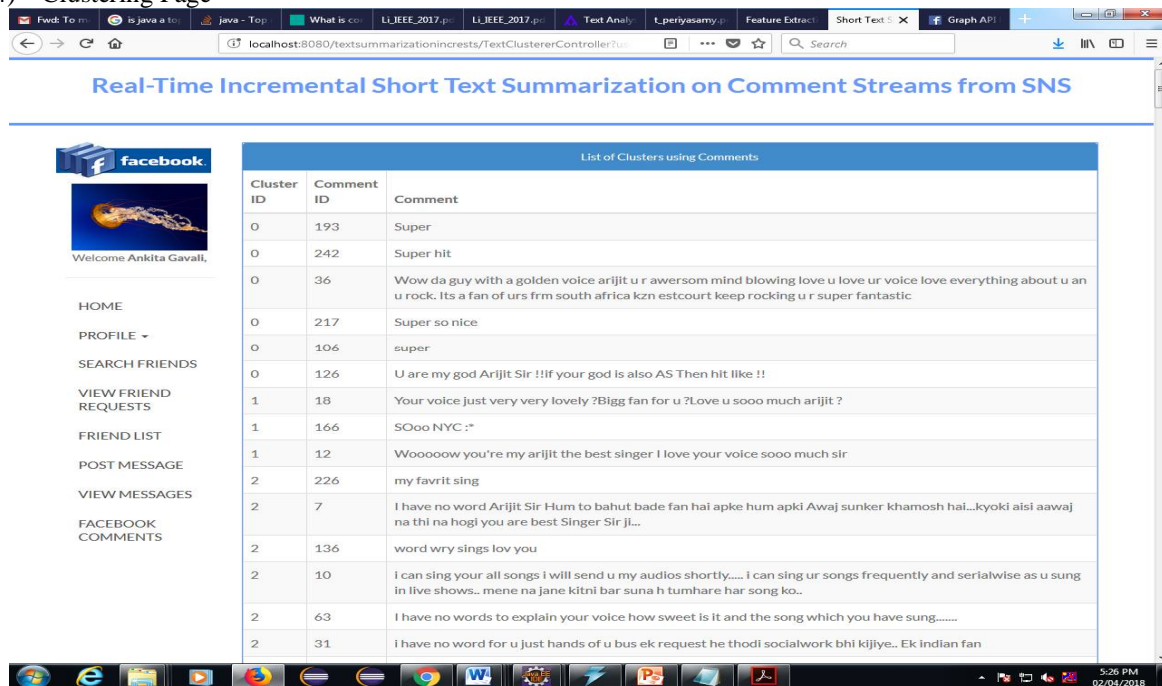
International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

4) Clustering Page



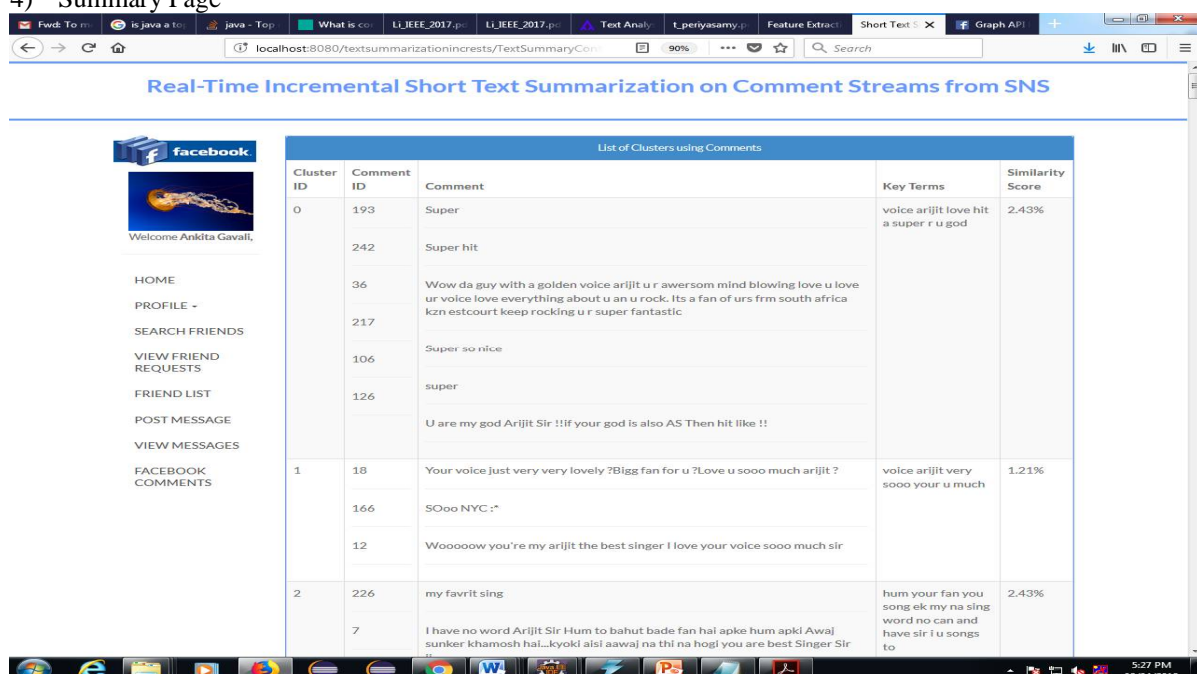
Real-Time Incremental Short Text Summarization on Comment Streams from SNS

facebook
Welcome Ankita Gavali.

HOME
PROFILE -
SEARCH FRIENDS
VIEW FRIEND REQUESTS
FRIEND LIST
POST MESSAGE
VIEW MESSAGES
FACEBOOK COMMENTS

Cluster ID	Comment ID	Comment
0	193	Super
0	242	Super hit
0	36	Wow da guy with a golden voice arjlit u r awersom mind blowing love u love ur voice love everything about u an u rock. Its a fan of urs frm south africa kzn estcourt keep rocking u r super fantastic
0	217	Super so nice
0	106	super
0	126	U are my god Arjlit Sir !!if your god is also AS Then hit ilke !!
1	18	Your voice just very very lovely ?Bigg fan for u ?Love u sooo much arjlit ?
1	166	SOoo NYC :*
1	12	Wooooow you're my arjlit the best singer I love your voice sooo much sir
2	226	my favrit sing
2	7	I have no word Arjlit Sir Hum to bahut bade fan hai apke hum apki Awaj sunker khamosh hai...kyoki aisi aawaj na thi na hogi you are best Singer Sir jL...
2	136	word wry sings lov you
2	10	I can sing your all songs i will send u my audios shortly..... i can sing ur songs frequently and serialwise as u sung in live shows.. mene na jane kitni bar suna h tumhare har song ko..
2	63	I have no words to explain your voice how sweet is it and the song which you have sung.....
2	31	I have no word for u just hands of u bus ek request he thodi socialwork bhi kijiye.. Ek indian fan

4) Summary Page



Real-Time Incremental Short Text Summarization on Comment Streams from SNS

facebook
Welcome Ankita Gavali.

HOME
PROFILE -
SEARCH FRIENDS
VIEW FRIEND REQUESTS
FRIEND LIST
POST MESSAGE
VIEW MESSAGES
FACEBOOK COMMENTS

Cluster ID	Comment ID	Comment	Key Terms	Similarity Score
0	193	Super	voice arjlit love hit a super r u god	2.43%
	242	Super hit		
	36	Wow da guy with a golden voice arjlit u r awersom mind blowing love u love ur voice love everything about u an u rock. Its a fan of urs frm south africa kzn estcourt keep rocking u r super fantastic		
	217	Super so nice		
	106	super		
1	126	U are my god Arjlit Sir !!if your god is also AS Then hit ilke !!	voice arjlit very sooo your u much	1.21%
	18	Your voice just very very lovely ?Bigg fan for u ?Love u sooo much arjlit ?		
	166	SOoo NYC :*		
2	12	Wooooow you're my arjlit the best singer I love your voice sooo much sir	hum your fan you song ek my na sing word no can and have sir i u songs to	2.43%
	226	my favrit sing		
	7	I have no word Arjlit Sir Hum to bahut bade fan hai apke hum apki Awaj sunker khamosh hai...kyoki aisi aawaj na thi na hogi you are best Singer Sir		

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 5, May 2018

VI. CONCLUSION

To enable the capability of comment stream summarization on SNS, we model a novel incremental clustering problem and propose the algorithm IncreSTS, which can incrementally update clustering results with latest incoming comments in real time. With the output of IncreSTS, we design a visualization interface consisting of basic information, key-term clouds, and representative comments. This at-a-glance presentation enables users to easily and rapidly get an overview understanding of a comment stream. From extensive experimental results and a real case demonstration, we verify that IncreSTS possesses the advantages of high efficiency, high scalability, and better handling outliers, which justifies the practicability of IncreSTS on the target problem.

REFERENCES

- [1] H. Becker, M. Naaman, and L. Gravano, "Learning similarity metrics for event identification in social media," in Proc. 3rd ACM Int. Conf. Web Search Data Mining, 2010, pp. 291–300.
- [2] H. Becker, M. Naaman, and L. Gravano, "Selecting quality twitter content for events," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 442–445.
- [3] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: Interactive topic-based browsing of social status streams," in Proc. 23rd Annu. ACM Symp. User Interface Softw. Technol., 2010, pp. 303–312.
- [4] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 450–453.
- [5] D. Chakrabarti and K. Punera, "Event summarization using tweets," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 66–73.