

FINDING RELATED FORUM POSTS THROUGH TWITTER DATA USING MATCHING ALGORITHM

Dr.K.N.S.Lakshmi, B.Madhu Sowmya Department of CSE SVPEC, Visakhapatnam, A.P, India

Abstract— Forum posts has the specific problem of finding related posts to a post at hand. By considering across the related documents the contents of posts are generally consider are whole. Here similarity process are done between two posts with respective segments and should be of same intention. All posts are generally fragmented in the form of group to attain the goal bunches. Now similarities are generally cross view the forums with sections and that will be of same intention.

Keywords—Segmentation

I. INTRODUCTION

Forums are generally a online discussion site, where people hold there conversations by posts. It is like a message board and different from chat rooms. A traditional approach for finding related document that perform content comparisons across content of posts, the contents are compared by different posts. The relatedness of two posts can then be based on a comparison across segments that serve the same goal. Every posts are generally considered as segments. Segments are generally said as parts (or) sections. In This the relatedness between two posts should be based on similarities respective to segments. The segmentation methods play important role by developing work with monitoring the no of text features ,it identify by parts of post. While this process performing significant jumps are occurred because of that segmentation are occur. Now segmentation of all posts are generally clustered in the form of intention cluster so that the similarities are calculated across segmentation with same intention. plays important role.

1) DATA MINING:

. Data mining also called as knowledge Discovery refer to extracting interesting patterns and useful knowledge from huge amount of data. Data mining in the databases is a new interdisciplinary field, merging ideas from statistics, machine learning, databases and parallel computing. Knowledge Discovery in Database(KDD) is a process of finding useful information and patterns in data where as Data mining is the use of algorithms to extract the information and patterns derived by the KDD process.

CLUSTERING:

Clustering in data mining is a discovery process that groups a set of data. The principle of clustering is "Maximizing the Intra

cluster similarity and minimizing the Inter cluster similarity." That is, a cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in the other clusters. A cluster of data can be treated collectively as one group in many applications.

SEGMENTATION

Segmentation is a key data mining technique. The key to segmentation is to decide how to split the database up. Segment is a group of consumers that react in a similar way to a particular marketing approach. So the key to segmentation is to decide how to split the database up.

II. EXISTING SYSTEM

There is no goal base division in Existing System. Question Answering Communities (QAC) plays important role they coordinate the work for posts. Division techniques are isolated into 2 general gatherings. The first is topical division where contiguous sets of content pieces are thought about for general likeness in light of terms or points. The second gathering of division strategies comprises of Transcribed oral-talk procedures utilized as a part of the investigation of translated oral correspondence utilizing semantic criteria.

1.3.1. Advantages of Existing System

 \Box No complexity in computations.

1.3.2.Disadvantages of Existing System

- □ Less Efficiency. .Complexity of TF/IDF algorithm is high.
- □ Difficult to finding related forum posts
- □ Segmentation increases cost
- \Box Searching is difficult
- □ Complexity of TF/IDF algorithm is high



III. PROPOSED SYSTEM:

3.1.SEGMENTATION OF POSTS

The challenging task is finding right segmentation process, from segmented documents, that are occurred in large body of work, if it is in the form of document d, there are 2^{jdj1} possible segmentations are occur.

Every segmentation process should be in the form of

- coherent
- largely disconnected from its adjacent segments.

Segmentation is the intention-based, these two properties translate to a segmentation where every segment conveys a single clear intention conveys by the adjacent segments.

3.2. SEGMENT GROUPING

Segments with similar intentions are created same group and segments with different intentions in different groups. It is modeled with vector of features, array of information are taken here. Now each cluster are generally communicates respectively the same goal.I to denote a cluster, and C to denote the set of the generated clusters.

Vector of weights that are based on the feature values are created by us. Vector with the letter F. Now consider two types of weights that capture the strength of the use of each CM categorical value, of each feature.

Now each CM value with in the segment are measured then the comparisons are done to categorical values that belong to same communication appearing in segments.Using the notion of the distribution table DSb_{CMr} of a communication mean CMr introduced in Section we define the vector F_s of weights, one weight for each feature.

3.3.SEGMENTATION REFINEMENT:

They should have same document with same segments that are end up with same cluster with same intention., if they have the same intention but are not same cluster then consequence document. The segments that belong to the same document in a cluster are concatenated into one.

Advantages of Proposed System

- Complexity is low.
- Searching became easy.
- Finding related forum posts became more efficiency.
- Searching became now easy

IV. ALGORITHM:

3.4.1.Matchin Algorithm:

Document matching is one of the best technique plays important role by collecting of documents that are generally related to reference document dq .Now the dq reference document are measure the relatedness between other documents d0 are lie in the form of IR technique.

3.4.1.1.Matching with respect to a specific Intention:

Every document with some specific intention are projected on each cluster. The specific intention are made by measuring the related documents to reference document d0. Text comparison are computed the relatedness between the documents like IR technique i.e. TF/IDF model.TF/IDF method and its probabilistic variance BM25 consists of a term weighting scheme that weighs a term in a document considering the number searching That variance computes the weight of a termt in a document d⁰. If s_q and s^0 are the segments of the documents d_q and d^0 , respectively, in the intention cluster. where fsq (t) denotes the frequency of the term t in the segment sq, jIj the cardinality of the intention cluster, and jI^tj the number of segments in the intention cluster.

3.4.1.2. Matching with respect to All the Intentions:

This algorithm consist of top-nlists generated across the different intentions, ., the set M are used to generate the k most related documents to the reference document $d_q R$ is created as new list contains in every document in lists in M. Each document are associated with the sum of the scores with which this document appears in the various lists in M. The k elements in R with the highest score are returned as answer to the request of the matching documents to the reference document d_q . High value for n compared to the value of k, on the other hand, will favor documents that appear in many lists even with not very high scores. We have empirically found that a good choice is an n equal to 2 k

3.4.2. Clustering Algorithm

Clustering is the task of dividing the data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar trait s and assign them into cluster. It is the most important unsupervised learning problem

K-Mean Method

If a vector of documents (Dl, D2 . . \cdot Dn) is given ,K-means clustering Algorithm will partition the n documents into K clusters (K :::; n) such that cosine distance between them is minimum.

International Journal of Engineering Applied Sciences and Technology, 2022 Vol. 7, Issue 11, ISSN No. 2455-2143, Pages 131-137 Published Online March 2023 in IJEAST (http://www.ijeast.com)





- 1. Randomly select initial centroid that divides the documents into k clusters.
- 2. Compute Cosine distance of each document from the centroid of each of the clusters. Assign that document to the cluster with the closest centroid.
- 3. Repeat step 2 until there is no change in newly formed clusters.



3.4.3. The Proposed Method As Follows:



3.4.3. 1 TF/IDF Algorithm

It is generally said as term frequency–inverse document frequency. It is often used as a <u>weighting factor</u> in searches of information retrieval, <u>text mining</u>, and <u>user modeling</u>. TF-IDF, a vector space based representation is the common technique used for text processing.

TF–IDF weighting scheme are often used for in scoring and ranking a document's <u>relevance</u> given a user <u>query</u>. TF–IDF can be successfully used for <u>stop-words</u> filtering in various subject fields, including

<u>Text Summarization</u>

Classification

The TF–IDF IS calculated as: TF-> Term Frequency

11 > Term requency

IDF->Inverse Document Frequency

 $\operatorname{tfidf}(t, d, D) = \operatorname{tf}(t, d) \cdot \operatorname{idf}(t, D)$

where t=term d=referencing document D=set of Documents





IV. RESULTS AND DISCUSSION:

i. OPEN THE COMMAND PROMPT AND RUN



ii. IT IS THE FIRST PAGE

		- trached and - range
Accuración de la construcción de la	Famat last Vite Page	
apertitative Reviews - 31 Reviews of Experimention	· season production and the season and	
benity law and divorce softlifters gloacatcher Meth	Tac. +48.05211.8747.0791	6
mordiargenery Contai Inaliti Feran, conference report, Publiced. IC20.01ml	Petr@silen(w)ervie.re	
international Health Forum. Anno	Collimentos.	
Bolf Conview - MCDCA - World Farger for Healthing Intel	Program practicity,	
Paged Min.	reade, prince the interview of a	
Paget Mm	KaNother planning	
*age15.ktm	PRATE FUEL M.A.	
Page 11 Ann.	printer a Net Aligne Cherry and a state	
Page 72.8mm	a Petra 14	
fage13.htm	Englished Seea	
Tage14.htm	Pagin as an exhibition	
Tage 15.4mm	Pergenti as a sume	
Page16.htm	Cabillete - Findarte - Alexa	
hage17.mm	Purther online offers within the Donar Portfolio "Yough & Medical Technologies" to Messe Donalized Contri-	
Page18.htm	Putter Tixte Fant	
Tage 75,800	MEDC46kame - Ozrumbreits brand	
fage1 bim	PREDICA - Descention?/Germany	
Tege 25.Mtm	CIOMP ABIED - Dilatel/doff / Germany	
hage21.htm	PREHICKHEE PUTERPARTICIPAE - Classificativ Germany	
Fage/22.htm	Production (Product - Manimum / Product	
*age33.809	BATTO'AL FAIR ADD. Designed	
Page J4.htm.	SMCOCH, FAM BRATE - 150 Paulo / Brate	
Tage 25.htm	BREDCK: FAR CHRA - Suther / China	
Summer State Section	Ball Ya'ar Juli Balla Aleman and Isaa Talle Inda	13

iii.NOW PRE PROCESSING IS TO BE PERFORMED

Versionistis 2 quest, las Versiones - 31 Revisione et Esperitano com-, Linguistione Esti Manchi las and disconte a statistico et discontentes com Manchiscophinary Califaci Houffe Essaine , ocarberence e supert Publiked Manchiscophinary Califaci Houffe Essaine , ocarberence e supert Publiked Manchiscophinary (Califaci Houffe Frances Inc.) Modelana attat	Proved Tast _ Vite Page Fit - et a Stat I STYTETS Infigurencessition (e) Stytestes Project ydawing	
Pagy02.htm Paget 8.00m Paget 8.00m Paget 9.00m Paget 9.00m Paget 9.00m Paget 9.00m Paget 9.00m Paget 9.00m Paget 9.00m Paget 9.00m	Ann (ar Canton and Anna anna a	P 19 Reser Classemer Control
1990-2.000 1990-2.000 Papell 21.000 Papell 21.000 Papel 23.000 Papel 24.000 Papel 24.0000 Papel 24.00000 Papel 24.0000 Papel 24.00000 Papel 24.00000 Papel 24.00000 Papel 24	File gans File of gaps Million Anno State (Second Second	

International Journal of Engineering Applied Sciences and Technology, 2022 Vol. 7, Issue 11, ISSN No. 2455-2143, Pages 131-137 Published Online March 2023 in IJEAST (http://www.ijeast.com)



iv.RESULT AFTER PREPROCESSING.

CET MAY AND AND AND A	Consideration of the state of the second states of the	a second s		
				Comparis Document Worghts Concat
The contraction of the	VVviii V	Lincal Freis	Entropy of Firms	Haradae Freis
Panarro Josson 1	2001 C.	23	4442	3582.5
Printiger B. School	C	91	0.172	0.0.0
Press of the Avenue	x.6	3	24	12.5
STUDIES & S. BURNESS		4	DOLD.	0.2.8
Presson W. P. Britson		1.	48.0	13.5 ()
Concession in Protocols 1	h	1	197(1).	348.0
and the second	ACCERR		18	0.h
and the second	AEBA	14 /	1.5+	1 (F14) 14
Constant of the second	AFTER	1.1	11	39.9
- mages 110, for 110 a	ADAM	1	194	30.0
Publish & V. Actives	26.5 m	10	3.48	99.5
Transport # III, To Basis	NURO.	1.	20 a	136.0
Page 10.60m	Children		12.000 (MS Pu
Parager 2. Adams		11	THE R.	11世纪1
Palagas (210, Rotato)	245	6	and the second s	
Panager (7.4. Belance			a de la compansión de la c	12. The second s
Transactor of the Archanter	81	6	1.10	100.00
Patrate 273 Actes	BAR CH.	1	100	14.4
Taning in the Station	and the local sector of th		11	170.11
inter 2 in Artest	AND CONTRACTOR OF		1 and an and a second s	64.5
A LOSS THE REAL OF	DA P OPPA		1 Control of Control o	14.5
and the second	AND CLORAR AN AND ME IN		444	50.0
many r. titres	AND Y VARIABLE A	3	6.8	10.0
-miller 210. Meters	HEOTH .	F.	10	17 M
-14-19-10 (7:30 Al-19-10	4668	*	01	8.0.0
haun 3.ham.	BOOKER	3	214	33.0
*#80.29.84m	INFLACING (T)	2	82	48.0
Index Private Article	49.0T	12	97	149.5
rean 39 Anno		15	46H	1837.0
man and the budgets	CARL STOLEN		107	66.0
man 3.4. htm	EARCADINII		48.	123.0
THE OWNER WATER OF THE OWNER	E REPERSATE II		147	2 M. B
Salar 3-6 holes	HAMPER .		10 A	13.8 C
Contract in P. Station	Difference (the fame		125	12.0
inclusion (bills designed)	District Distance		127	10 M
And the lot of the lot	and the second se	5 · · · · · · · · · · · · · · · · · · ·	71	- 12 0 -
and the sector		17-		2.2
	00.000	14 ·····	the second se	- 4 C u
Cardina and the state of the st	Contraction and a second se			11.0

v.COMPUTE DOCUMENT WEIGHTS

or Clarenced Creaseservisioned	Test Date Segmentations of	reg Carrierts Adaptertree	and the second se	the second s
				Bookd MIV's Markets
counts.	Weinerskill	Local Fred	City and Press	Photome Freis Cusine Weight
Avters .	2.2	2442	11557.0	83.88310004547441
Antony Con-	10	117	10.0	16540410230615113
A BARREL AVE	1	12.4	19.8.5	0.7307570030110070
a second s		101	53.5	2 6092611720100220
1781 P	1	40	23.6	0.00013054672301665
and the second se		1724	344.0	2 5302579046404540
LACIE MAN		142	16.61	1 4295 1 34802 1905 72
ADD		12.34	1200.0	8.000765455935644
adam Lag Trilling	14	94	144 m	In This Area and a second second second
ArtiAlty	12	28	15.0	11 120000 24700 3005 2
Reference Lots &	19	1110	28.44 16	B # 35 3 3 5 F# 37 94003
Add (10)	1	2.4	14.0 %	0.73875769181109278
Cont Cana	14	4803	124.5	A DOLY LOUGH THE PROPERTY
adarb.	144	10.6.3	1207.0	128 400 1270 4477 1270
area Cabula	1.0	Tab	116.0	1 0 1 36 15 36 39 30 70 70 4
Address (Address)	15	1945	140.0	14 HARDE ATRONOM (41)
Address (Address)	15	110	100	A STREAM PATRIAGES
Address (A.V.		11.12	100.0	The transmission of a second second
the second second	12	1017	112.2	Construction of the state of th
and a state of the		17.1	100.0	D COMPACE VOID A TRANSPORT
A ALL AND	1	100		D 1993 31 STATES AND ADDRESS
and a close of		-1127	177.5	The second second second second
states of an in the lot of the lot		- 17T-1	10.0	The second s
Roberts Total Land	1.	100	100	In the second second second second
Automa and Automatica	1	113	100	1 LAND POLYTAN LINE PALL
Infern Control		14.97		The second
test and an and a second second			40.0	
Address of the local division of the local d			11.12	1.477019203023029923
and the second second second	15	102	10.0	1 SAME A PROPERTY AND A DESCRIPTION
	12	Land.		1 - Contraction of the Contracti
Address of the second	15	1.44	3647.0	. 10//##09##054379#
Address of the Party of the Par		- FBC		
Arton Contractor	1			
Adam COUNTRIGAT	419. IV.		124.0	0.07123149804211242
Adams Control of the		181	381.0	0.10072407004732408
THE REPORT OF	10	120	18.0	77.53567827072003836
And I have been a			10.0	10 10075241405013374
Construction of the second	18	17.		x 4 x 1 x 9 3 0 0 0 0 0 1 2
ANN CLEAR	100			3 0 12 5 5 0 5 2 3 4 5 5 7
teres Constanting	1	- Const	1.0	3 8122220054300124
Adda CLICK	17	42.5	- 183.0	15.5446284848427202

VI. BUILT VSM MATRIX

Cites the	tional litteragemetrics	and Taxo Claix Degree	and designed in a	strag literaphy of	Adapted with the same of											6.064.0	-
											1	. tear		a sector a be	*****	104	(the
and a c	Ny Manite Co	careers have a series of	June		in the second se	_			_		_			-	_		
Oriele .	01	09	0.0.				-1-1-1-1				1-1-1-		1-1-1-			1-1-1-	
	O DESCRIPTION OF	In the Loss of the party	10. At 4									distant a					1.1
	TO BELLENGARD	10-0-4-10-7-0-2403-444	20.004														-
	DO RESIDENCES	A BOARD AND AND AND	20.078									des des des		and the second second		in the last	-
	00.004044.000	B \$78407.4440.	10.0														-
	N STRUCTURE .	10 3074-4 FEMILIAN 1	28.878														
	20.843361108.	In addition and a local	20.000.0												h-h-h-	1-1-1-	-
	22.581914021	10 March 2007283	30.8953														-
	N. 0185330124-	10.000.04.007710-	22.201-1													1	
-	at semana see	0.001112100.343	10.018				1.1.1.1.1				- in the second		1.1.1	1.1.1		1.1.1.	1
	A 48 10 10 2 3 4	In the owner of a state	0.000-1										1-1-1-		-		
	ALC: NOT THE REPORT OF	10.89718320590	10000	and the second second				1.000				de la	1-1-1-	1-1-1-		1-1-1-	-
	23.0707457000	10 0000 T122381	21.0.04	and the property of the			1000	1.000			de la la la	de la	1-1-1-	1	-	1-1-1-	-
	A BREAKERSTAL	0.0.1.41010.4340	20.000												h-h-h-		-
<u> </u>	10.0000474914	0.004000A0004	24.000														-
_	DO REFERENCE	IN MINT COURSES	20 899													1-1-1-	
	TO EXPERIMENT OF LARSE		10.020						in production for			de la desta	Jack Strates	designed as		h-h-h-	-
	A 2007 18707	In star type, A top ball	20.020		_							de la la					
	D. WOTTANDARS	I BE TRUE AND A	10.005									des des des				4-4-4-	-
	20.0017241100	In mapping the second															-
	10.302424900	0.01-00100-0018	37.822														-
	28.241379558	IN STATEMAN AND TH	20.8484										1-1-1-			1-1-1-	
	D. PARONOPUS	10.002031410000	20.045														-
<u> </u>	10.0000407.00.	10.000 million 2 (00 - 20	20.000									1					
	20.303764053	0.80559515.040	20.8781		_					_							
5	33 TA 94 DA 96 DA 96 DA	3.83297¥30966	30 Beerld									de la la					
<u> </u>	(0. Prinzen168)	B-#7582414503	30.083						_								
ł	20.4432310306	10 M.J. 10 4 JULY 10 7	20.884		_												-
·	20.31109988888	15 BOATTY / 5 25 55	20. P.O					_									
·	20.010132710.	19-904007796423.	10.044							_							-
	CONTRACTORY.	IN MAR CARLETTER.	20.000														1
	10.5 FUEDOULTE.	0.81368540525	PATR	and the local data in the								la l	-			1-1-1-1-1	1
-	N. 39499273483	THE REPORT OF A PARTY OF A PARTY OF A	20.7.84										1				1
	0.896832296	19 W00002374078	-R.818								-	in the second		1		1	
	20.200304043	0.800003175760	20.847	and the second second second									1			la la la	1.1
h	0.3839984657	0.80110054820	0.0/9	and and and and and				1 1 1 1 1 1 1 1									
	20.458972493	10 mm / hw > Y / h # #	10.021								-					1	1
	20.2210304224	0.000000000.004	30.5FB	and the second sec				-		-			-				1
e	DURING MARINESS	0.07401031547	10.054														
10	201.010.004.010.004	10.877.4/57Amber	30.875 L														

International Journal of Engineering Applied Sciences and Technology, 2022 Vol. 7, Issue 11, ISSN No. 2455-2143, Pages 131-137 Published Online March 2023 in IJEAST (http://www.ijeast.com)



V. OUTPUT

Papelläter	Page4Lites				
Papelitan	Pagelitta	 UND Obasel Hill 			
Papellitite	PagetElites	+ UNP Heath Pare			
PaperSann	Pepettites	• uno ripper pore			
Pagei738x	Paget2.tes	 INVESTIGATION 			
Fupit.200	PagetLines	Login			
- Paprittand	Page14.885				
papelitation	Paget5.htm				
pege \$3.0kmi	Page18.iom				
PageTates	Pagettilites				
Pagekillin	Papettilles				
Servers and Operating Systems _ HPE Block, Discounters and Levens Commands.Mov	Page/Eliters	Office of Global Health Educatio			
Top 100 Healthcare Blogs, Unitedies & Deardetiers To Fallow to 2015 taxis	Page21.009				
What are some of the best online law by small in toda, - Gaets Anni	Page 20.mm				
What knows extent to discore legal matters . Quirachted	Page21.ims	A MARKET AND A MARKET AND A MARKET AND A MARKET			
	Page24.889	Toggle navigation Office of Glo			
	Page25.mm	Education			
	Page 26, http:	Education			

VI. CONCLUSION

This paper presents an approach to calculate the semantic similarity between two words, sentences or paragraphs. A novel approach proposed by us for matching a reference post to the k most related posts in a collection. In our method, segmentation are done across the posts that convey similar with some intentions. We presented several experiments regarding the right segmentation criteria, the effectiveness of the segmentation algorithms and the formation of intention clusters that prove that a rather intuitive concept, that of the author intentions to communicate a certain message, can be effectively captured by an automated process. Then extractive summarization is used to extract feature terms and the summarized sentences were ranked based on feature frequency of the posts, measuring the relatedness score after having distinguished the different

VII. FUTURE SCOPE

- As we are just transferring only the text messages, we can also transfer the files, images etc.
- As we are using MUTUAL AUTHANTICATION PROTOCOL for the communication we can also use OPTIMALITY OF KEYING PROTOCOL.
- In this project we have been used symmetric keys for the communication we can also use ID of the sensors for the communication.

REFERENCES

- [1]. M. Chen, X. Jin, and D. Shen, "Short text classification improved by learning multi-granularity topics," in IJCAI, 2011, pp. 1776–1781.
- [2]. J. Jeon, W. B. Croft, and J. H. Lee, "Finding semantically similar questions based on their answers," in Proceedings of the 28th ACM SIGIR Conference, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 617–618
- [3]. S. Robertson, S.Walker, and M. Hancock-Beaulieu, "Okapi at TREC-7: Automatic ad hoc, filtering, VLC and interactive track," TREC '98, pp. 199–210, 1998

[4]. Jiawei Han, Micheline Kamber, Data Mining and Concepts.

The state

- [5]. Maqgret H. Dunham, Data Mining and Introductory to Advanced Topics
- [6]. .ClusterAnalysis:http://www.tutorialspoint.com/data mining/dm cluste r analysis.htm
- [7]. DataMining: http://www.webdocs.cs.ualberta.ca/~zaiane/courses/c omput690
- [8]. PorterStemmerAlgorithm: http://snowball.tartarus.org/porter/stemmer.htm
- [9]. TF-IDF: <u>http://www.tfidf.com</u>
- [10]. K-Means-Algorithm:<u>http://sites.google.com/site/dataclusteringa</u> lgorithms/k-means-clustering- algorithm
- [11]. SentimentAnalysis Lexalytics: https://www.lexalytics.com/technology/sentiment
- [12]. G. Salton, A. Singhal, C. Buckley, and M. Mitra, "Automatic text decomposition using text segments and text themes," in ACM Hypertext, 1996, pp. 53–65
- [13]. S. Louvigne, N. Rubens, F. Anma, and T. Okamoto, "Utilizing social media for goal setting based on observational learning," in ICALT, 2012, pp. 736– 737.
- [14]. K.Wang, Z. Ming, and T. Chua, "A syntactic tree matching approach to find similar questions in community QAservices," in ACM SIGIR, 2009, pp. 187 – 194.
- [15]. A. Shtok, G. Dror, and Y. Maarek, "Learning from the past: Answering new questions with past answers," in WWW, 2012, pp. 759–768.
- [16]. K. Jones, C. Van Rijsbergen, B. L. Research, and D. Department, Report on the Need for and Provision of an Ideal Information Retrieval Test Collection, ser. British Library Research and Development reports, 1975.



- [17]. J. Kekalainen, "Binary and graded relevance in ir," Inf. Processing & Management, vol. 41, no. 5, pp. 1019 – 1033,2005.
- [18]. Z.-Y. Ming, T.-S. Chua, and G. Cong, "Exploring domain specific term weight in archived question search," in Proceedings of the 19th ACM CIKM, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 1605–1608.
- [19]. H. Wen, W. Zhongyuan, W. Haixun, Z. Kai, and Z. Xiaofang, "Short text understanding through lexical-semantic analysis," in IEEE ICDE, 2015
- [20]. J. Jeon, W. B. Croft, and J. H. Lee, "Finding semantically similar questions based on their answers," in Proceedings of the 28th ACM SIGIR Conference, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 617–618.