NTCIR-12 MOBILECLICK: Sense-based Ranking and Summarization of English Queries

Monalisa Dey
JIS College of Engineering
Kalyani, Nadia
India
+91-8981121692
monalisa.dey.21@gmail.com

Anupam Mondal*
Jadavpur University
Kolkata
India
+65-83984727
anupam@sentic.net

Dipankar Das
Jadavpur University
Kolkata
India
+91-9432226464
ddas@cse.jdvu.ac.in

ABSTRACT
NTCIR-12 MobileClick task has been designed to rank and summarize English queries. The primary aim of this task was to develop a system which is capable of minimizing interaction between the human users and mobile phones while extracting relevant data with respect to given queries. Organizers provided the data represented as information units (iUnits). Each of the iUnits describes a pertinent query associated with other information like type or category, relevance, sense and knowledge-based relations [1] [2] [4]. The task is divided into two sub-tasks namely ranking and summarization. The ranking sub-task focuses on identifying the important iUnits related to a query. In the summarization sub-task, the output has to be designed as a two-layered model where the first layer will identify the important iUnits and the second layer will compile those important iUnits and generate a summarized output for the query. In this present task, we have employed several sentiment lexicons like SentiWordNet[1], SenticNet[2] etc. with tabulation based approaches to identify the important query-based iUnits for ranking and summarization. Our sense-based system has achieved a score of 0.8859 mean Q-measure for ranking and score of 11.7033 mean M-measure for summarization tasks, respectively.

Team Name
JUNLP

Subtasks
iUnit Ranking (English), iUnit Summarization (English)

Keywords
MobileClick, Query, iUnit, Ranking, Summarization, Sentiment, Text-rank, Wup-similarity

1. INTRODUCTION
Web-based searching and co related information extraction with summarization is treated as a challenging task due to the unawareness of the knowledge-based classification of the search query. Once the search query is given, a user receives information-based links (URL's) related to the searched query.

To extract the required information from these links, context analysis is essential and, it helps to identify the proper ranking of the provided links. Especially, when users search a query on mobile phones, due to the small screen size, it becomes very exasperating. A substantial amount of interaction with the device is essential to filter out the required information from the large amount of data provided. For this reason, developing a better search engine for mobile phones is crucial. The main aim of this task is to develop a system which will help mobile phone users to attain the related information with minimum effort.

The task has been divided into two subtasks namely ranking and summarization. The query related detail is represented in the form of information units (iUnits) which help to define the appropriate and atomic pieces of information related to the query. The assumption was that among all provided iUnits, the most germane iUnits should identify the knowledge-based gloss of the query. Our developed system ranks a set of iUnits provided for the individual queries against the degree of similarity with respect to the queries. The ranking score of the iUnits has been calculated by a sense-based approach.

In the summarization subtask, these extracted ranked iUnits play an important role. The sense-based approach helps to prepare an extractive query related to two layered summary from the highly ranked iUnits. As the output has to be a structured one, the first layer contains the most important iUnits and links to the second layer, whereas, the second layer connects such links with iUnits for providing query related summarization as an output. The extractive approach has been adopted due to the nature of the task where an iUnit is considered to be the extraction unit. Figure 1 shows the general framework of our developed system. The developed system has limited the number of characters on the screen to 420 due to the size constraint of the mobile screen.

The rest of the paper is organized as follows. Section 2 gives a description of the proposed system along with the distribution of the test dataset. Section 3 provides the extensive experimental results with the evaluation measures used to test the efficiency of the system and Section 4 finally offers concluding remarks.

[1] http://sentiwordnet.isti.cnr.it/

[2] http://sentic.net/

* Corresponding author: Anupam Mondal [Emailid: anupam@sentic.net]

2. PROPOSED SYSTEM

2.1 Data

NTCIR-12 organizers provided the training and test data for ranking the ranking subtask. For the summarization subtask, only the test data was provided. Test data set for the ranking subtask consists of two files, one for queries and one for iUnits with respect to each of the queries. The query file contains 100 queries, whereas the iUnit files contains 4342 iUnits for all the 100 queries. Test data for the summarization subtask consists of three files,

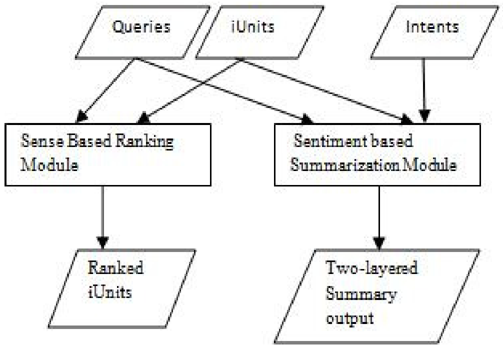

Figure 1. Framework of our system

viz. queries, iUnits and intents. The notion of intents has been introduced in the NTCIR-12 MobileClick2 task, which is constructed by clustering of iUnits with corresponding cluster labels, where each of the labels is termed as intent.

2.1.1 IUnit Ranking

In order to identify the rank of information units, we have developed a sense-based approach. The word or context level sense signifies the knowledge and sentiment information related to the context. The sentiment or knowledge information helps us to understand the proper context-based affinity between the iUnits and their related query. In order to determine the ranking of iUnits, we have calculated the sense-based difference between query and context of the iUnits. We have provided the rank of these iUnits against the sense-based differences. The run file of the ranking subtask is represented as query id, iUnit id and rank (sense-based). Figure 2 shows a snapshot of our run file with the provided test data where QID indicates the query id and UID indicates the corresponding selected iUnit id and RANK represents the score given to each iUnit. Figure 3 illustrates the ranking framework for identifying the rank of the iUnits for a particular query.

SentiWordNet and SenticNet sentiment lexicons help us to extract the sense-based score of the iUnits with a tabulation approach. These lexicons provide a concept-based positive, negative and neutral sense with their corresponding polarity score.

2.1.2 iUnit Summarization

Summarization output of the iUnits related to each query is represented as a two layered architecture, where the first layer consists of the most important iUnits and links (intents).

QID	UID	RANK
MC2-E-0003	MC2-E-0003-0021	1
MC2-E-0003	MC2-E-0003-0022	25
MC2-E-0003	MC2-E-0003-0024	34

Figure 2. Snapshot of the Run File for Ranking

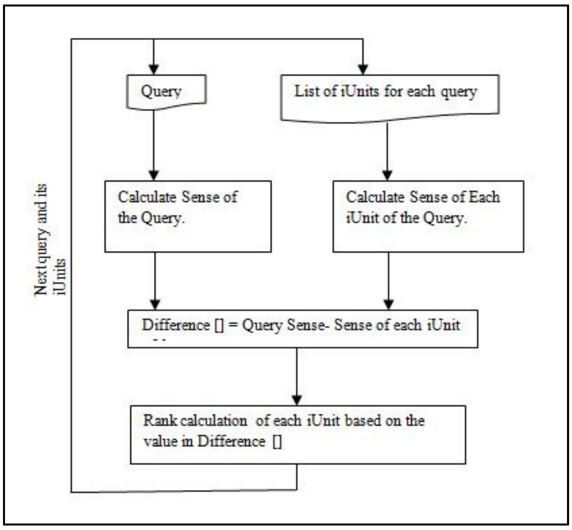

Figure 3. Framework of the Ranking Algorithm

To summarize iUnits against the queries, we have followed the Text-Rank and Wup-Similarity based models [3][5].

The TextRank Model

To extract the important iUnits for the first layer summary, we have used a graph based method which is a modified version of the text-rank algorithm [3].]. Text-rank model primarily uses a text to graph conversion approach. So the ranked iUnit (words, sentences etc.) is represented as vertices of the graph and the interconnections between these iUnits (vertices) are established by the meaningful relations between the query and iUnits. A meaningful relation is identified by a concept voting based approach. The connection (edge) of the vertices shows the voting relation between these vertices. Vertices having greater scores are assumed to be more important than the others. The model is defined as follows: let G = (V, E) be a directed graph with a set of vertices V and a set of edges E, where E is the subset of V x V. For a given vertex V_i, let in $In(V_i)$ be the set of vertices that point to it and $Out(V_i)$ be the set of vertices that V_i points to. Equation (1) shows technique to calculate the score under the model.

$$S(V_i) = (1-d) + d * \sum_{j \in \mathrm{In}(V_i)} \frac{1}{|Out(V_j)|} S(V_j) \qquad (1)$$

Where, d is a damping factor that can be set in the range 0 to 1, which has the role of integrating the probability of jumping from a given vertex to another random vertex in the graph. The factor d is usually set to 0.85.

To convert the iUnits into a graph, a similarity matrix has been constructed. This matrix provides a score of similarity measure between the iUnits by identifying the content overlaps between them. Equation (2) shows the similarity score extraction process, where the sentences are taken as S_i and S_j with N_i words, where the sentence is represented as $S_i = w_1^i, w_2i \ldots w_N^i$.

$$Similarity(S_i, S_j) = \frac{\left|\{w_k | w_k \in S_i \,\&\, w_k \in S_j\}\right|}{\log(|S_i|) + \log(|S_j|)} \quad (2)$$

After obtaining the similarity score, we compared it with our defined threshold score of 10 in order to identify the important iUnits for the first layer of the summarization subtask. Table 1 shows the output of our developed system in the form of the selected iUnits and their contents for the extracted query "*hulk hogan*" having an ID of **"MC2-E-0001"**. In the Table, UID indicates the selected iUnit id and CONTENT indicates the text corresponding to the iUnit id.

For the second layer summary, the intents that are provided for each query are linked with their associated iUnits. The wup-similarity model has been used in this process.

Wup_similarity Measure

Wu and Palmer (Wup) similarity model also follows a graphical-measurement like the text-rank model. Text-rank model follows a voting approach for creating a similarity matrix, whereas, the wup-similarity model counts the number of edges. The principle of wup-similarity computation is based on the distance between the concept nodes (e.g. C1 and C2) and root concept (R) node of the graph. The distances between the concepts node C1 and C2 from the root node R are indicated as N1 and N2, where the distance N separates the closest common ancestor (CS) of C1 and C2 from the root node R. Equation (3), shows the similarity measurement between two concept nodes (C1 and C2).

$$WP_{sim} = \frac{2 * N}{(N_1 + N_2)} \quad (3)$$

We have applied the wup-similarity approach on the queries and their related intents both for the test data set provided in order to obtain the context similarity score. We have set the threshold of wup-similarity score to 1(one) for identifying the most related iUnits for each of the intents. Table 2 indicates the iUnits selection approach based on the similarity threshold score of 1 where QID, IID, UID and SCORE indicate the query id, intent id, iUnit id and wup-similarity score. For the query id MC2-E-0001, and intent id MC2-E-0001-INTENT0001, the selected iUnit id's based on the score, are represented in the Table.

3. EVALUATION

We have submitted two different run files for the ranking and summarization under NTCIR-12 Mobile-Click task (English).

Table 1. iUnit Text Corresponding to iUnit ID's

UID	CONTENT
MC2-E-0001-0006	most recognized wrestling star worldwide
MC2-E-0001-0009	became the face of pro wrestling
MC2-E-0001-0010	made his debut in the American Wrestling Association
MC2-E-0001-0015	regularly attended wrestling cards at the Tampa Sportatorium in high school
MC2-E-0001-0022	won his first wrestling championship

Table 2. Wup_similarity scores between intents and iUnits

QID	IID	UID	SCORE
MC2-E-0001	MC2-E-0001-INTENT0001	MC2-E-0001-0004	1.4010989011
MC2-E-0001	MC2-E-0001-INTENT0001	MC2-E-0001-0010	1.29298642534
MC2-E-0001	MC2-E-0001-INTENT0001	MC2-E-0001-0011	1.03626373626
MC2-E-0001	MC2-E-0001-INTENT0001	MC2-E-0001-0015	1.94471802707
MC2-E-0001	MC2-E-0001-INTENT0001	MC2-E-0001-0018	1.16483516484

The iUnit ranking subtask is evaluated with the score of normalized Discounted Cumulative Gain (nDCG) for different cutoff thresholds k.

Equation (4) shows the calculation process for the evaluation metric Discounted Cumulative Gain (DCG).

$$nDCG@K = \sum_{r=1}^{K} \frac{GG(u_r)}{\log_2(r+1)} \quad (4)$$

Table 3. Performance of our system in the iUnit ranking task

QID	RUN ID	nDCG@3	nDCG@5	nDCG@10	nDCG@20	Q
Mean	87	0.7012	0.7268	0.7807	0.8506	0.8859

Equation (5) indicates the normalized version of DCG (nDCG) for measuring the score of iUnits.

$$nDCG@K = \frac{DCG@K}{iDCG@K} \tag{5}$$

where, iDCG indicates the DCG ideal ranked list of iUnits, which help to rank these iUnits based on the global importance of the query.

Another metric, Q-measure is introduced to verify the iUnit ranking output of the query. Q-measure is evaluated by a recall-based graded-relevance approach whereas nDCG is measured by a rank-based graded-relevance approach. Table 3 shows the scores of nDCG and Q-measure for our submitted runs of the ranking subtask.

In case of type based evaluation of our results, we have classified the provided queries into four different classes namely people, places, random and sentence type. Table 4 shows the example of these classes with query id provided in the test data.

Figure 4 and Figure 5 shows the graphical representation of the average nDCG and Q-measures received for the different types of query with varying threshold values

Table 4. Classification of Queries

QID	TYPE
MC2-E-0001 - 0020	Type 1 (People)
MC2-E-0021 - 0040	Type 2 (Places)
MC2-E-0041 - 0080	Type 3 (Random)
MC2-E-0081 - 0100	Type 4 (Sentence type)

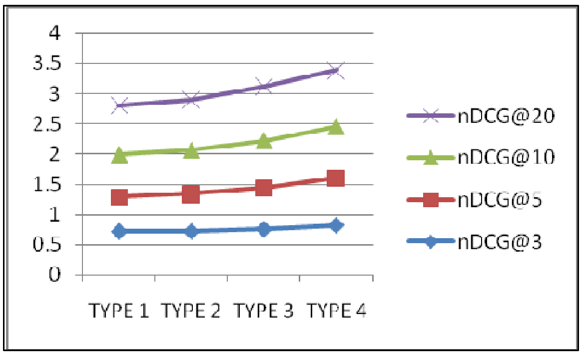

Figure 4. Mean nDCG at Different Thresholds for Four Query Types

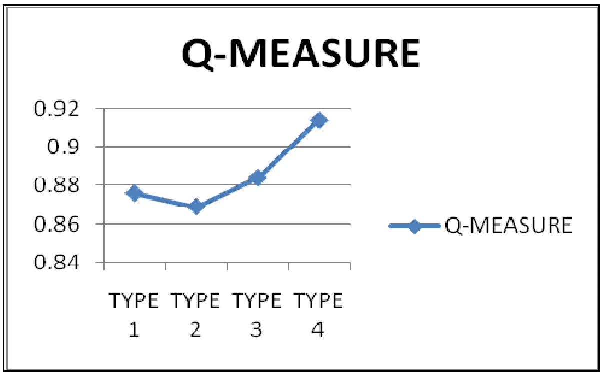

Figure 5. Mean Q-measure at Different Thresholds for Four Query Types

In the summarization subtask, the performance of the system submitted, is calculated using the M-measure metric. Equation (6) illustrates the M-measure score evaluation process.

$$M = \sum_{t \in T} P(t)U(t) \tag{6}$$

where, where T is a set of all possible trail-texts, P(t) is a probability of going through a trail t, and U(t) is the unit-measure score of the trail.

Our submitted run for the summarization subtask, received **11.7033** as the M-Measure score. Figure 6 shows the mean M-measure achieved for the four query types. The figure indicates that the M-measures for type 2 (places) and type 4 (Sentence type) queries are poor. Type 2 has produced 11 faulty queries with very low M-measure score out of 20 queries. Especially, the queries with id numbers 22, 25, 27, 36, 37, 40 and 50 have produced a score less than 3. We have observed that the iUnits of these queries mainly contain numeric values like telephone numbers, timings, pin codes etc. In Type 4, 8 queries, out of 20 queries, we have obtained a score less than 6. Table 5 shows some extracted queries along with their iUnits context.

4. CONCLUSION

This paper presents an approach to the NTCIR-12 MobileClick task that is, English query related iUnits ranking and summarization for mobile phones. We have developed a sense-based ranking system with text-similarity and a wup-similarity based two layer summarization system.

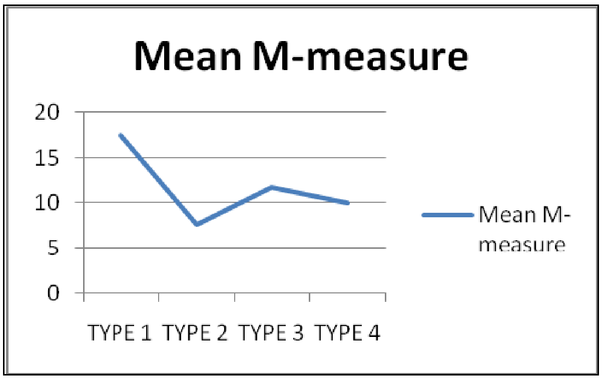

Figure 6. Mean M-measure for Four Query Types

Table 5. iUnits of Queries 37 and 85

QID	iUnit ID	iUnits
MC2-E-037	MC2-E-0037-0009	9 South Street, Town Centre, Worthing
	MC2-E-0037-0010	Tel: 023 9237 0606
	MC2-E-0037-0011	MC2-E-0037-0011
	MC2-E-0037-0012	MC2-E-0037-0012
MC2-E-085	MC2-E-0085-0002	Cc: Carbon copy
	MC2-E-0085-0003	To: indicates primary recipients
	MC2-E-0085-0007	Bcc: Blind carbon copy

Our system produced a noticeable output score of 0.8859 for the ranking and 11.7033 for summarization tasks using test data sets out of 397 numbers of runs submitted in total. The paper explains the task design, system descriptions and evaluation methodology along with analysis of the results. For future work, we plan to improve the accuracy of our system by incorporating more fine grained features.

5. REFERENCES

[1] A. Bah Rabiou and B. Carterette. Udel @ NTCIR-11 MobileClick Track. In Proceedings. of NTCIR-11 Conference, 2014.

[2] M. P. Kato, M. Ekstrand-Abueg, V. Pavlu, T. Sakai, T. Yamamoto, and M. Iwata. Overview of the NTCIR-10 1CLICK-2 Task *In Proceedings of NTCIR-10*, pages 243–249, 2013.

[3] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 1998.

[4] T. Sakai, M. P. Kato, and Y.-I. Song. Overview of NTCIR-9 1CLICK. *In Proceedings of NTCIR-9*, pages 180–201, 2011.

[5] Wu, Z., and Palmer, M. Verb semantics and lexical selection. In 32nd Annual Meeting of the Association for Computational Linguistics, 133–138, 1994.