HITSZ-ICRC at NTCIR-12 Temporal Information Access Task

Yongshuai Hou
houyongshuai@hitsz.edu.cn

Cong Tan
viptancong@gmail.com

Xiaolong Wang
wangxl@insun.hit.edu.cn

Jun Xu
hit.xujun@gmail.com

Qingcai Chen
qingcai.chen@gmail.com

Intelligence Computing Research Center
Department of Computer Science and Technology
Harbin Institute of Technology Shenzhen Graduate School, 518055 Shenzhen, China

ABSTRACT
This paper presents the methods HITSZ-ICRC group used to Temporalia-2 task at NTCIR-12, including subtask Temporal Intent Disambiguation (TID) and subtask Temporal Diversified Retrieval (TDR). In the TID subtask, we merged results of rule based method and word temporal intent classes vector based method to estimate temporal intent classes distribution on English queries and Chinese queries. The rule based method was improved from the method we used in Temporalia-1. The word temporal intent classes vector based method estimated temporal intent classes distribution by normalizing the sum of temporal intent classes vectors of all words in the query. In the TDR subtask, for the temporal information retrieval, we used TIR system in Temporal-1 to get ranked documents list for each temporal subtopic; for the temporally diversified ranking, we used all documents in result lists of the four temporal subtopics as candidate documents set for a query topic, and ranked each document in the candidate set based on: the document relevant score to each subtopic, the temporal intent classes of temporal expressions in each document and the temporal information of previous ranked documents for the topic. We only tried our methods for TDR subtask on English topics.

Keywords
temporal intent disambiguation, temporal information retrieval, temporally diversified ranking.

Team Name
HITSZ-ICRC

Subtasks
Temporal Intent Disambiguation (English and Chinese)

Temporal Diversified Retrieval (English)

1. INTRODUCTION
With the data published on the Web becoming faster and the data size becoming larger daily, more and more web information becomes time-dependent, and the temporal intent cannot be ignored for web search [7, 9]. Temporal aspects of web search have gained a lot of research focus in recent years. But there is fewer dataset and venue for researchers to compare their methods for temporal web search.

Temporal Information Access (*Temporalia*) task hosted at the NTCiR Workshop on Evaluation of Information Access Technologies (NTCiR) created a venue and provided annotated datasets for researcher to compare their methods for temporal information access task[4]. The Temporalia-2 task[6] at NTCIR-12 is a follow-up and upgraded task after Temporalia-1 [5] at NTCIR-11.

Temporalia-2 task included *Temporal Intent Disambiguation* (TID) subtask and *Temporally Diversified Retrieval* (TDR) subtask in English and Chinese. The TID subtask required participants estimate a temporal intent distribution over four temporal intent classes (*Past*, *Recency*, *Future* and *Atemporal*) for a given query following the time understanding of the query. The TDR subtask required participants to do same job as TIR subtask in Temporalia-1[5], which is retrieving a set of documents relevant to each of four temporal intent classes subtopics for a given topic, and at same time, TDR subtask required participants to return a set of documents that is temporally diversified for the topic.

The HITSZ-ICRC group at Harbin Institute of Technology Shenzhen Graduate School participated in the both subtasks of the Temporalia-2. For the TID subtask, we improved the rule based method and multi-results merging method used in Temporalia-1 [2] to make the methods matching TID requirements, and we designed a word temporal intent vector based method for TID subtask. For the TDR subtask, we used the methods in Temporalia-1 to run the TIR part of the subtask, and designed subtopics relevance vector based method and time expressions classes vector based method to run temporally diversified ranking part of the subtask.

The remainder of this paper is organized as follows. Section 2 describes the methods used for the TID subtask, explains formal run submissions and analysis the evaluation results for the TID subtask. Section 3 describes the methods designed for the TDR subtask, introduces data used, details formal run submissions and analysis the evaluation results for the TDR subtask. Section 4 concludes the paper.

2. TEMPORAL INTENT DISAMBIGUATION
The TID subtask required participants to estimate a temporal intent distribution of four temporal intent classes including *Past*, *Recency*, *Future* and *Atemporal* for a given query. This is an upgraded challenge from *Temporal Query Intent Classification* (TQIC) subtask in Temporalia-1 [5], which only required participants giving a single temporal intent class for a given query.

2.1 Methods for TID Subtask

We considered TID subtask as a task to calculate a probability distribution vector with four temporal intent dimensions (*Past*, *Recency*, *Future*, *Atemporal*) for a given query. For example, the probability distribution of the query "*value of silver dollars 1976*" over temporal intent class *Past*, *Recency*, *Future* and *Atemporal* is 0.727, 0.273, 0 and 0, we presented the distribution as a 4 dimensions vector [0.727, 0.273, 0, 0], which was named temporal intent vector here. Given a query Q, its temporal intent classes vector was denoted to $QI=[p, r, f, a]$, where $p+r+f+a=1$, and the TID task becomes to estimate the value of vector QI.

We used rule based method and word temporal intent vector based method to calculate temporal intent vector QI for a query, and merged results of the two methods to get final result for TID subtask.

2.1.1 Rule based method

The rule based method was improved from the method we used for TQIC subtask in Temporalia-1[2], which based on time-sensitive word dictionary, date distance between date in query and query issue time, verb tense.

By observing the queries in dry run set, we found that many features to classify temporal intent of some queries are obvious. For example, if the word "*prediction*" appears in a query, *Future* intent probability will be higher for the query, like query "*NFL Playoffs Predictions*"; if a query includes date expression, the temporal intent distribution can be estimated by the date distance between the date expression and query issue time, like query "*value of silver dollars 1976*"; for queries in English, the verb tense also is an important feature to estimate temporal intent, like "*when was electricity invented*". Based on those features, we created three groups rules to estimate temporal intent probability distribution for a given query .

Time-sensitive word dictionary: creating a dictionary for time-sensitive words and set the temporal intent vector for each word based on dry run queries; at the query intent classification step, judging whether the query contains time-sensitive words in the dictionary, if the query contains time-sensitive word, used the word temporal intent vector as query temporal intent vector.

Date distance: comparing the date expression in query and query issue date to get the date distance, set the query temporal intent vector based on the date distance.

Combining date distance and verb tense: combined the verb tense and date distance to create rules to set query temporal intent vector.

The improvements to the rule based method we used in TQIC task are:

(1) Each rule gives a single temporal intent class in TQIC; each rule was changed to give out a temporal intent vector in TID task. For example, we changed temporal intent of time-sensitive word "*schedule*" from temporal intent class *Future* to temporal intent vector [0, 0.3, 0.5, 0.2].

(2) The temporal intent class of a query can be decided by only one rule in TQIC; the temporal intent vector of a query was calculated by normalized the sum of temporal intent vectors of all rules which can cover the query in TID subtask.

(3) Temporal class of query that cannot be covered by rules designed was set to *Atemporal* as default class in TQIC, and a default temporal intent vector was assigned to the query cannot be covered by all rules designed in TID.

2.1.2 Word temporal intent vector based method

Each user query Q can be seen as a sequence of words, and each query can be presented as $Q=\{w_1, w_2, \ldots, w_n\}$.

We hypothesized that each word has a temporal intent vector WI, and all temporal intent vectors of words in a query have contribution to the temporal intent vector QI of the query. Here we calculate QI of a query using following formula:

$$QI = \frac{\sum_{i=1}^{n} WI_i}{\left|\sum_{i=1}^{n} WI_i\right|} \quad (1)$$

Where WI_i is the word temporal intent vector of the ith word in the query.

For a given word, its temporal intent vector WI is calculated by counting all queries in training dataset that contain the word, and WI is calculated by following formula:

$$WI = \frac{\sum_{i=1}^{m} QI_i}{\left|\sum_{i=1}^{m} QI_i\right|} \quad (2)$$

Where QI_i is temporal intent vector of the ith query that contains the word in training dataset.

Each <*word*, *WI*> pair was save to word dictionary to be used for QI calculating in formal run step. And temporal intent vectors of all words out of word dictionary were set to zero vector, and if all the words in the query were out of the dictionary, the QI of the query will be set to a default vector.

2.1.3 Multi-results merging method

The rule based method have high accuracy to classify user query, but all rules should be designed manually, and the rules designed cannot cover all queries in dry run dataset and formal run dataset; the word temporal intent vector based method can cover more user queries in dataset, but the classify accuracy is low. So we designed a method to merge results from different methods to improve the final temporal intent disambiguation result by weighted sum the temporal intent vectors for the query.

Given a query Q with n temporal intent vectors from different methods, the ith temporal intent vector is denoted as QI_i, and the final temporal intent vector QI is calculated by following formula:

$$QI = \frac{\sum_{i=1}^{n} w_i QI_i}{\left|\sum_{i=1}^{n} w_i QI_i\right|} \quad (3)$$

Where w_i is the weight assigned to the *i*th temporal intent vector.

2.2 Results Evaluation for TID Subtask

In TID subtask, Stanford CoreNLP toolkit [8] was used to extract POS and normalized date for each query in English, and Jieba[1] toolkit was used to segment and extract POS for each query in Chinese.

To train and turn methods, the dry run dataset was used as training dataset. There are only 93 queries in English and 52 queries in Chinese in dry run datasets of TID subtask. To increase size of the training dataset, the dry run and formal run datasets in TIQC subtask of Temporalia-1[5] were use to train and turn methods for the English TID subtask. Query in TIQC subtask was given with single temporal intent category, which cannot be used to TID subtask directly. We changed the temporal category of each query in TIQC to a temporal intent vector to make the query can be used to TID subtask. For example, the temporal category *Past* of query "*current price of gold*" in TIQC subtask was changed to temporal intent vector [0, 1, 0, 0].

In formal run step, there are 300 queries in English and 300 queries in Chinese to disambiguate temporal intent. We submitted 3 results for each language queries for TID subtask, which includes: TID-E-1, TID-E-2 and TID-E-3 for queries in English, TID-C-1, TID-C-2 and TID-C-3 for queries in Chinese.

Results of the TID subtask were evaluated with two metrics: Averaged Absolute Lose (*AvgAbsLoss*) and Averaged Cosine Similarity (*AvgCosin*) [6]. The evaluation results of our submissions are shown in Table 1 and Table 2.

Table 1. Results evaluation for TID subtask in English

RunID	AvgAbsLoss	AvgCosin
TID-E-1	0.1465	0.8496
TID-E-2	0.1647	0.8499
TID-E-3	0.2049	0.7443

All the results submitted for TID subtask in English are runs of methods described in section 2.2. TID-E-1 is the run of the rule based method; TID-E-3 is the run of the word temporal intent vector based method trained with dry run dataset of TID subtask; TID-E-2 is the run of the multi-results merging method, by merging TID-E-1, TID-E-3 and another run of word temporal intent vector based method trained with dry run dataset and formal run dataset of TIQC subtask of Temporalia-1.

Table 1 shows that the performance of multi-results merging method is better than the other two methods, which means merging results from different methods can improve performance of single method for TID subtask in English. With checking the detail evaluating results, we found that the number of queries whose *AvgCosin* lower than 0.3 is 19, 6 and 33, and the number of queries whose *AvgCosin* equal to 1.0 is 34, 2 and 16 in run TID-E-1, TID-E-2 and TID-E-3. We can see that the lower performance queries were improved but the higher performance queries were reduced in result of the merging method.

In the run TID-E-1, queries with obvious features can get higher performance, as "*Canadian Dollar Exchange Rate*" (010), "*history of halloween*" (080) and "*When Did WW2 Start*" (219); but if queries with no obvious feature or rules failed to cover get lower performance, as query "*April Jobs Report*" (101), "*the power of now*" (263) and "*full moon may*" (298).

The performance of run TID-E-3 is lowest. After checking the evaluation result, we found that contribution of different word temporal intent vector to query temporal intent vector is different, and some words had no contribution, so it should add a key word extracting for each query to training and testing steps in future experiment.

Table 2. Results evaluation for TID subtask in Chinese

RunID	AvgAbsLoss	AvgCosin
TID-C-1	0.1606	0.8601
TID-C-2	0.1308	0.8854
TID-C-3	0.2080	0.7522

The runs for queries in Chinese in TID subtask used same methods as the runs for queries in English. TID-C-2 is the run of the rule based method; TID-C-3 is the run of the word temporal intent vector based method trained with dry run dataset of TID subtask; TID-C-1 is the run of the multi-results merging method, by merging TID-C-2 and TID-C-3. For we have no other training dataset in Chinese, the results merging method only run on TID-C-2 and TID-C-3.

From Table 2 we can see that, the performance of the results merging method is higher than word temporal intent vector based method but lower than rule based method. The word temporal vector based method had no contribution to results merging method for TID subtask in Chinese.

The rules for queries in Chinese covered fewer queries than rules for queries in English. This is because that Chinese queries has no tense feature and rules for Chinese quries is fewer than English queries. The performance on Chinese queries is higher than English queries, this is caused by using same default temporal intent vector setting for queries in English and Chinese. The default temporal intent vector is more suitable for Chinese queries.

3. TEMPORAL DIVERSIFIED RETRIEVAL

In TDR subtask, each topic contains one topic description and four subtopic questions in four temporal intent classes (*Past*, *Recency*, *Future*, and *Atemporal*). The TDR includes two parts task here: the first part task required participants to retrieve and rank a set of documents for each temporal subtopic and required documents relevant to the subtopic in content and temporal intent, the work is same to the TIR subtask in Temporalia-1 [5]; the second part task asked participants to return a list of documents that is temporally diversified for the topic. For the TIR part task, we used methods for TIR subtask in Temporalia-1 [2, 5] to retrieve and rank documents for each subtopic.

The temporally diversified retrieving and ranking part task includes two jobs: documents temporally diversified retrieving and documents temporally diversified ranking. Documents temporally diversified retrieving requires to retrieve a set of documents which are relevant to the topic in different temporal subtopics. Here we used results of the TIR part task as temporally diversified retrieving results, and all documents in ranked lists of

[1] https://github.com/fxsjy/jieba

the four temporal subtopics for a topic were used as candidate documents for documents temporally diversified ranking. Here we focus on how to temporally diversified rank the candidate documents for the given topic.

3.1 Method for Temporally Diversified Ranking

For describing easily, we first explain some notations used later. Giving a topic *T*, there are four subtopics for *T*, $\{ST_1, ST_2, ST_3, ST_4\}$. The documents list for subtopic ST_i in TIR part task is denoted as $D_i = \{(d_1, r_1), (d_2, r_2), \ldots, (d_m, r_m)\}$, i=1, 2, 3, 4, and d_i denotes a document relevant to subtopic ST_i, r_i denotes the relevant score between ST_i and d_i. The candidate documents set *CandiD* for temporally diversified ranking for topic *T* is the union of set D_1, D_2, D_3 and D_4. For a document *d* in *CandiD*, it may be relevant to any subtopic ST_i, so the relevant score between *d* to each subtopic can be denoted as $R=(r_1, r_2, r_3, r_4)$, and r_i is the relevant score between subtopic ST_i and *d*. The candidate documents set *CandiD* can be denoted as: $CandiD=\{(d_1, R_1), (d_2, R_2), \ldots, (d_n, R_n)\}$.

For a document *d* in *CandiD*, its temporally diversified score *divR* is decided by two factors: relevant score *temR* between *T* and *d* in content and temporal intent; temporal difference *diF* between *d* and documents already ranked in temporally diversified list for *T*.

The temporally diversified documents list for topic *T* was denoted as $RL=\{(d_1, R_1), (d_2, R_2), \ldots, (d_{i-1}, R_{i-1})\}$, where i-1 is the rank position for document d_{i-1}, and the temporally diversified ranking score *divR* for document *d* which will be placed to the *i*th position in *RL* is calculated by following formula:

$$\text{divR}=\text{diF}\times temR \tag{4}$$

To find the most suitable document for the *i*th rank position in *RL*, we calculate *divR* for all documents in *CandiD* except the documents already in *RL*. The document with highest *divR* is placed to the *i*th rank position in *RL*.

Temporal relevant score *temR* for document *d* is decided by its relevance between *d* and each subtopic, r_1, r_2, r_3 and r_4. Here *temR* is calculated with a simple way by following formula:

$$temR = r_1 + r_2 + r_3 + r_4 \tag{5}$$

The different factor *diF* is decided by temporal difference and position distance between *d* and each document in *RL*. We designed two ways to calculate *diF*: subtopics relevance vector based way and document time expressions classes vector based way.

3.1.1 Subtopics relevance vector based

The current document d_i will be placed to the *i*th ranking position should be temporally diversified to documents in *RL* by considering all four temporal classes. The temporal diversify is determined by the temporal difference between the current document and each document already ranked in *RL*, and the document temporal difference is a key factor for *diF*.

The document subtopics relevance vector *R* in *candiD* shows the relevance between the document and each subtopic in the four temporal classes, so the temporal difference between two documents can be presented by the difference between subtopics relevance vectors of the two documents. So the different factor *diF* can be calculated by following formula:

$$diF = \sum_{j=1}^{m} \frac{1}{j} \left| R_{i-j} - R_i \right| \tag{6}$$

There are four temporal classes for each topic, so it's only need to compare to the last 3 documents in RL, we set m=3 here.

3.1.2 Time expressions classes vector based

Temporal difference of two documents also can be presented by the difference of the time expressions in the two documents. But the number and value of time expressions in different documents are different, it cannot be compare directly.

Time expressions in each document were annotated out and normalized in the corpus [4]. So each time expression can be easily classified to temporal class *Past*, *Recency* or *Future* based on its relation with search date. Numbers of time expressions in each temporal classes for different documents can be compare directly.

The numbers of time expressions in different temporal classes in a document was used to present the document temporal information here. Temporal information of a document can be presented as time expressions classes vector *texV*=[*p*, *r*, *f*], where *p*, *r*, and *f* is the number of time expression in class *Past*, *Recency* and *Future* in the document.

After getting time expressions classes vector for each document, comparing temporal difference of two documents becomes to comparing the time expressions classes vectors of the two documents. To calculate *diF* based on time expressions classes vector, we replaced vector *R* with vector *texV* in formula (6) and get formula to calculate *diF* based on *texV* as following:

$$diF = \sum_{j=1}^{m} \frac{1}{j} \left| \frac{\text{tex}V_{i-j}}{\left|\text{tex}V_{i-j}\right|} - \frac{\text{tex}V_i}{\left|\text{tex}V_i\right|} \right| \tag{7}$$

Where $texV_i$ is the time expressions classes vector of the document to be placed to the *i*th rank position of *RL*, $texV_{i-j}$ is the time expressions classes vector of the *(i-j)*th document already ranked in *RL*, and m is set to m=3 here.

3.2 Results Evaluation for TDR Subtask

In TDR subtask, for the TIR part task we indexed the "LivingKnowledge news and blogs annotated sub-collection" corpus [4] and searched each subtopic using two systems in Temporalia-1 [5]:

(1) Solr system [5]: using Apache Solr[2] (version 4.6.0) with BM25 [10] weighting scheme to index the corpus and search each subtopic. The query string input to the system for each subtopic is *topic* + *description* + *subtopic*. We changed the system setting to output relevant score for each document in result list.

(2) HITSZ_BW system [2]: using Lucene[3] in Java with BM25 model to build index for the corpus and search candidate documents for each subtopic, and using

[2] http://lucene.apache.org/solr/

[3] http://lucene.apache.org/

relevant score weight sum method to rank candidate documents for each subtopic. System inputting for each subtopic includes *topic* + *subtopic* and the subtopic class information was used.

For the temporally diversified ranking part task, we run the subtopics relevance vector based method and the time expressions classes vector based method on results of the two systems for TIR part task. For the time expressions classes vector method, we classified time expressions used the method we designed for TIR subtask in Temporalia-1 [2].

For the evaluation, ranked list for a specific temporal subtopic was evaluated by the metric *nDCG* [3], and the temporally diversified ranked list for a topic was evaluated by the metric *a-nDCG* [1] and *D#-nDCG* [11], and *D#-nDCG* was used as main metric for temporally deversified ranking.

In TDR subtask, we only run our system on topics in English. There are 50 search topics in formal run, each topic with four specific temporal subtopic. We submitted 3 runs for the TDR subtask: TDR-E-2, TDR-E-3 and TDR-E-4.

TDR-E-2: Ranked list for each specific temporal subtopic is the result of system HITSZ_BW; temporally diversified ranked list is the result of the subtopic relevance vector based method.

TDR-E-3: Ranked list for each specific temporal subtopic is the result of Solr system; temporally diversified ranked list is the result of the time expressions classes vector based method.

TDR-E-4: Ranked list for each specific temporal subtopic is the result of system HITSZ_BW; temporally diversified ranked list is the result of the time expressions classes vector based method.

The evaluation results of the 3 runs are shown in Table 3 and Table 4. Table 3 shows the performance of the two systems for the TIR part task for each specific temporal subtopic; Table 4 shows the performance of different temporally diversified ranking methods with different candidate documents. All runs were evaluated with a cutoff value of 20.

Table 3. nDCG@20 of each specific temporal class subtopic and all subtopics

RunId	atemp.	future	past	rec.	all
TDR-E-2	0.5372	0.479	0.5896	0.5046	0.5276
TDR-E-3	0.6359	0.6086	0.6299	0.5893	0.6159

Table 4. Results evaluation of formal runs for temporally diversified ranking

RunID	nDCG@0020	D#-nDCG@0020
TDR-E-2	0.7236	0.8647
TDR-E-3	0.7273	0.8619
TDR-E-4	0.5808	0.7885

From Table 3 we can see that, the performance of Solr system is better than HITSZ_BW system on specific temporal subtopic retrieval task, which is same as in Temporalia-1 [5].

The Table 4 shows that, for the temporal diversified ranking task, with same candidate documents, performance of subtopic relevance vector based method is better than the time expressions classes vector based method; with same temporally diversified ranking method, performance of using candidate documents of Solr system is better, which means better performance temporally diversified ranking is based on system of better performance on TIR part task.

4. COCLUSIONS

This paper presents the methods HITSZ-ICRC group used for TID and TDR subtask in Temprolia-2 task at the NTCIR-12. For TID subtask, we employed rule based method, word temporal intent vector based method and multi-results merging method to disambiguate temporal intent of user query. For TDR subtask, we designed subtopic relevance vector based method and time expressions classes vector based method to temporally diversified ranking documents for a topic, and we tried our methods on results from different TIR systems. Results evaluation shows that the methods we used were effective for TID and TDR subtask.

5. ACKNOWLEDGMENTS

This paper is supported in part by grants: NSFCs (National Natural Science Foundation of China) (61402128, 61473101, 61173075 and 61272383) and Strategic Emerging Industry Development Special Funds of Shenzhen (JCYJ20140508161040764 , JCYJ20140417172417105 and JCYJ20140627163809422).

6. REFERENCES

[1] Clarke, C.L.A., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S. and MacKinnon, I. 2008. Novelty and Diversity in Information Retrieval Evaluation. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2008), 659–666.

[2] Hou, Y., Tan, C., Xu, J., Pan, Y., Chen, Q. and Wang, X. 2014. HITSZ-ICRC at NTCIR-11 Temporalia Task. *Proceedings of the 11th NTCIR Conference, Tokyo, Japan* (2014).

[3] Järvelin, K. and Kekäläinen, J. 2002. Cumulated Gain-based Evaluation of IR Techniques. *ACM Trans. Inf. Syst.* 20, 4 (Oct. 2002), 422–446.

[4] Joho, H., Jatowt, A. and Blanco, R. 2014. NTCIR Temporalia: A Test Collection for Temporal Information Access Research. *Proceedings of the 23rd International Conference on World Wide Web* (New York, NY, USA, 2014), 845–850.

[5] Joho, H., Jatowt, A., Blanco, R., Naka, H. and Yamamoto, S. 2014. Overview of ntcir-11 temporal information access (temporalia) task. *Proceedings of the NTCIR-11 Conference* (2014).

[6] Joho, H., Jatowt, A., Blanco, R., Yu, H. and Yamamoto, S. 2016. Overview of NTCIR-12 Temporal Information Access (Temporalia-2) Task. *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies* (2016).

[7] Joho, H., Jatowt, A. and Roi, B. 2013. A Survey of Temporal Web Search Experience. *Proceedings of the 22Nd International Conference on World Wide Web* (New York, NY, USA, 2013), 1101–1108.

[8] Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J. and McClosky, D. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. *Proceedings of 52nd Annual*

Meeting of the Association for Computational Linguistics: System Demonstrations (2014), 55–60.

[9] Moulahi, B., Tamine, L. and Yahia, S.B. 2015. When time meets information retrieval: Past proposals, current plans and future trends. *Journal of Information Science*. (2015), 1–24.

[10] Robertson, S., Zaragoza, H. and Taylor, M. 2004. Simple BM25 Extension to Multiple Weighted Fields. *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management* (New York, NY, USA, 2004), 42–49.

[11] Sakai, T. and Song, R. 2011. Evaluating Diversified Search Results Using Per-intent Graded Relevance. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2011), 1043–1052.