4C1-IOS-4b-2

# Inclusion of Temporal Semantics over Keyword-based Linked Data Retrieval

Md-Mizanur Rahoman*1 Ryutaro Ichise*2

*1 Department of Informatics
The Graduate University for Advanced Studies
Tokyo, Japan
mizan@nii.ac.jp

*2 National Institute of Informatics
Tokyo, Japan
ichise@nii.ac.jp

Temporal features, such as date and time or time of an event, always expose some concise semantics over any kind of information retrieval, and so over linked data information retrieval. On one hand, we see, contemporary research tries to adapt linked data information retrieval with easy and familiar keyword-based retrieval to hide the complexity of data's underlying technologies. On the other hand, we find, linked data information retrieval perspective, most of them overlook the power of temporal feature inclusion. Considering the both, this study, investigates the importance of temporal feature inclusion over linked data information retrieval. We propose a keyword-based linked data information retrieval framework which can incorporate temporal features and can give more concise results. Our investigation justify the significance of temporal feature inclusion over linked data retrieval.

## 1. Introduction

Temporal feature related information, such as information related to date and time or time of an event, is helpful in finding appropriate result or in discovering new relationship. Though temporal feature related information extraction over usual document-based data has quite long history of research, same kind of research is rarely seen over linked data. The reason either could be that research community has overlooked this issue or linked data has different structure than the usual document-based data. In this study we will attribute those issues and proposes an easy to use linked data information retrieval framework which can retrieve temporal feature related information.

Linked data follow loose data publishing strategy which means data publishers, without knowing other publishers' data schema, can publish data using their own data schema, and construct a global data. For this loose data publishing strategy, linked data inherently hold data heterogeneity. This heterogeneity gets magnified when data publishers publish temporal feature related data because temporal feature itself inherit diverse presentation strategy. So temporal feature related information retrieval over linked data, particularly by general purpose users, is a challenge. Moreover, literature review [Vandenbussche 11] also indicates that very few initiatives were taken to adapt temporal feature related linked data information retrieval. Our motivation is to incorporate temporal feature related information retrieval over linked data with an easy to use QA system. We use keyword-based QA system [Rahoman 12], because keyword-based system is thought familiar system to the general purpose users. Moreover, the system has developed some predefined templates which is thought an efficient technique. So contribution of this study is how we can adapt temporal feature to this QA system.

Contact: Md-Mizanur Rahoman, The Graduate University for Advanced Studies, Tokyo, Japan. mizan@nii.ac.jp

## 2. Related Work

DARPA's Trans-lingual Information Detection, Extraction, and Summarization (TIDES) in 2000 is one of an earlier attempt to retrieve temporal feature related information over document-based data. In 2002, Automatic Content Extraction (ACE), and in 2004, ACE Time Normalization (TERN) conducted event based information extraction task which brought several temporal feature related information extraction researches. So temporal feature related information extraction over usual document-based content is an active research field. On contrary, temporal feature related information extraction over linked data is relatively new. In 2011, Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE) workshop is considered one of the first linked data temporal feature related information extraction venture. In that event, we find study [Vandenbussche 11] was directly related to temporal feature related information extraction while other studies were mostly attributed to event related research.

## 3. Adaptation of temporal semantics

This section will deals with temporal feature attachment to QA system. Saquete et. al., showed that temporal feature related query questions always hold some indicator words which are called as *signal words* [Saquete 09]. For example, query question *Which artist were born on the 29th of December 1960?* holds temporal feature indicating word *on* that follows temporal feature *the 29th of December 1960*. So word *on* is considered as *signal word*. In this study, we assume that input keywords will contain *signal word*. Input keywords which hold *signal words*, we call them as *signal keywords*.

Figure 1 shows overall process flow of our proposed system. It is mainly segregated into two phases: query text processing and semantic query. In first phase, process *ordering key generator* divide input keywords according to *signal words* and generates two set of keywords: Q-FKS and

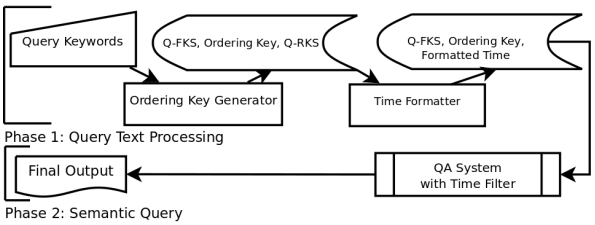


Figure 1: Overall workflow of our proposed system

Q-RKS, then put appropriate "ordering key" (described in details below) to those keyword sets. Then process *time formatter* convert explicit time to some common format. In second phase, our proposed system uses previous phase processed input and impose time filter to produces intended result.

### 3.1 Query Text Processing

This section talks about query text processing. Temporal question answering perspective, Saquete et. al., showed a technique which called "ordering key" [Saquete 09]. "Ordering key" helps to attach temporal features to the input keywords. Process *ordering key generator* divides input keywords into two keyword sets, then establishes "ordering key". Two keyword sets are:

1. *Q-Focus keyword set* (Q-FKS): keywords which specify the information that user is searching for. *Signal words* prior keywords fall into Q-FKS.

2. *Q-Restriction keyword set* (Q-RKS): keywords which specify restricting temporal features that are used to restrict required information. *Signal words* follower keywords fall into Q-RKS.

If input keywords for previous example question were crafted as *artist, birthday(i.e., Q-FKS)* and *on the 29th of December 1960* (i.e., Q-RKS), then according to *signal word*, "ordering key" would be "=" (equal) between Q-FKS and Q-RKS keywords. Saquete et.al., showed several "ordering key" for various *signal keywords* [Saquete 09]. Then process *time formatter* uses Stanford parser [Chang 12] on Q-RKS keywords to convert all kind of temporal values to a common format.

### 3.2 Semantic Query

QA system described in [Rahoman 12] automatically constructs SPARQL query for input keywords. Semantic query phase uses QA system [Rahoman 12] and extracts Q-FKS input keywords related SPARQL query. For date and time related part of SPARQL query, *QA system with time filter* redefines query by adding filter clause considering "ordering key" and previous phase formatted temporal value. Final SPARQL query generates intended result.

## 4. Experiment

We use Question Answering over Linked Data 1 (QALD-1) open challenge question set in our experiment. We select only those questions which relate temporal feature in their answering. Since our system depends on keywords, we intuitively retrieve keywords for those questions which simultaneously balance to the dataset and the query questions.

Table 1: QALD-1 temporal feature related question answering performance by proposed system

| Participant question set | # of questions | Performance of proposed system | | |
|---|---|---|---|---|
| | | Recall (avg) | Precession (avg) | F1 Measure (avg) |
| DBPedia QALD-1 | 4 | 1.000 | 1.000 | 1.000 |
| MusicBrainz QALD-1 | 18 | 0.765 | 0.765 | 0.765 |

To show the performance of our system, we execute questions from each participant question set and check average recall, average precision and average F1 measure for each set. Table 1 shows QALD-1 temporal feature related question answering performance by our proposed system.

We find that our proposed system achieves gold standard performance for DBPedia dataset questions. On contrary, our system perfumes low for MusicBrainz dataset questions. Our detail investigation finds that this performance drop is not because of lacking in temporal feature attachment rather it is because, our proposed system is not able to generate Q-FKS related information.

## 5. Conclusion

Temporal feature always holds precise semantics. Like other data, linked data also conform this narrative. This study shows that how temporal feature attached linked data information can be retrieved. In future we want to extend our system for more flexible keyword-based system because current system depends on exact keyword matching.

## References

[Chang 12] Angel X. Chang, Christopher Manning. SUTime: A library for recognizing and normalizing time expressions. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, 3735–3740, 2012.

[Rahoman 12] Md-Mizanur Rahoman, Ryutaro Ichise. An automated template selection framework for keyword query over linked data. In *Proceedings of the 2nd Joint International Semantic Technology Conference*, 175–190, 2012.

[Saquete 09] Estela Saquete, José Luis Vicedo González, Patricio Martínez-Barco, Rafael Muñoz, Hector Llorens. Enhancing QA systems with complex temporal question processing capabilities. *Journal of Artificial Intelligence*, 35:775–811, 2009.

[Vandenbussche 11] Pierre-Yves Vandenbussche and Charles Teissédre. Events retrieval using enhanced semantic web knowledge. In *Proceedings of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web*, 112–116, 2011.