Using Uncertain Time Intervals in Linked Data

Sekino, T.
International Research Center for Japanese Studies, Japan, E-mail: sekino@nichibun.ac.jp

Abstract
Uncertain temporal data often included in a dataset are critical, especially in history and archeology. In this study, we proposed a method to represent uncertain time intervals in uncertain temporal data and examine the relationships between them using the Linked Data technology. We created uncertain time intervals by specifying the ranges for the beginning and the ending points, and its features were represented by possible and reliable time intervals. Moreover, we divided the states of relationships between the two uncertain time intervals as "reliable," "possible," and "impossible" relations for each Allen's relation. Using a sample model, we demonstrated that procedures for creating uncertain time intervals and examining the relationships between them can be realized as RDFs in the Linked Data world.

1. Introduction

Uncertain data are an extremely important problem when dealing with spatiotemporal data. Numerous cases contain uncertain data, especially in history and archeology. However, usual information systems such as databases and data analysis tools are designed to accept only specific format data and refuse uncertain data. For temporal data, most information systems assume that data used are determined at date or a more detailed level. Therefore, users have to forcibly fit uncertain temporal data to the system. For example, a user often tentatively sets the data date as January 1 to fit uncertain data at the year level. However, this method of forcibly fitting hides the fact that the original data are uncertain and may cause misunderstanding for other users. Obtaining appropriate analysis and retrieval results using these data is challenging and requires mechanisms and procedures to deal appropriately with uncertain temporal data.

Many studies represent uncertain time and examine relationships between them (Billiet and De Tré, 2016). Studies attempting to represent uncertainness using possibility theory are mainstream. They used the fuzzy set theory (e.g., Nagypál and Motik, 2003 and Billiet et al., 2011), rough set theory (e.g., Asmussen et al., 2009 and Qiang et al., 2009), and both theories (e.g., Qiang et al., 2010) to represent uncertainness of time. Furthermore, most of these studies apply Allen's relations; the relations between two time intervals are classified into 13 types of relative relations (Allen, 1983).

Regardless of the numerous studies on uncertain time, studies on forcibly fitting the uncertain temporal data to the format for information systems are ongoing. These studies, using the possibility theory, describe more detailed behavior of uncertain time. However, information, tools, and fundamental data to support applying these theories to actual data are scant. Consequently, although vast uncertain temporal data are accumulated in various scientific fields, applying those theories to them and using these uncertain temporal data are not yet proceeded. Thus, a simple method to deal with uncertain temporal data is required.

Considering this background, we propose a procedure to deal with uncertain temporal data. Thus, we used the Linked Data, which are linkable to each other and are often associated with open sciences and semantic Web technology. In recent years, a lot of RDF (Resource Description Framework) data are available on the Internet as the Linked Data (Bizer et al., 2009). The concept and mechanism of Linked Data are suitable to represent uncertain temporal data, because they are easy to associate the uncertain temporal data with data that gives hints to explain and solve the uncertainness. Therefore, in this study, we represented uncertain temporal data by a simple model that can be applied to Linked Data; we examined the procedure for creating RDF data from actual data according to the model.

2. Representation of Uncertain Time Intervals

We assumed that uncertain temporal data comprise uncertain time intervals. A time interval is an interval between two time instants representing its boundaries. The beginning point is included in the interval, whereas the ending point excluded. When the beginning and ending point of a time interval are the same, it is a time instant. For simplicity, we

Online ISSN 2673-0014/

assumed that all time intervals are continuous between both boundaries.

Either or both boundaries in an uncertain time interval are uncertain. We represented the boundaries as ranges in which the time instants of a true boundary were located. For example, a time interval represented in words "a building that existed from the 15th century until the 1560s" represents a typical uncertain time interval. True dates when the building was completed and demolished must be mentioned, though both dates are represented as ranges in which the dates may exist. In this uncertain time interval, as long as the beginning point is in the 15th century (between January 01, 1401, and December 31, 1500) and the ending point is in the 1560s (between January 01, 1560, and December 31, 1569), all time intervals can be considered true time intervals. This indicates that an uncertain time interval has more than one state that may be determined. We determined the uncertain time interval at a state in which the beginning and ending point is $\widehat{\omega}_b$ and $\widehat{\omega}_e$, respectively; a time interval representing this state is defined as a "determinate time interval" denoted as $\widehat{\omega}$. Using this definition, we redefine an "uncertain time interval" as a set of all possible determinate time intervals (Figure 1) denoted as ω (Equation 1):

$$\widehat{\omega} \coloneqq \{x | \widehat{\omega}_b \leq x < \widehat{\omega}_e\}$$
$$\omega \coloneqq \{s | s = \widehat{\omega}\}$$

Equation 1

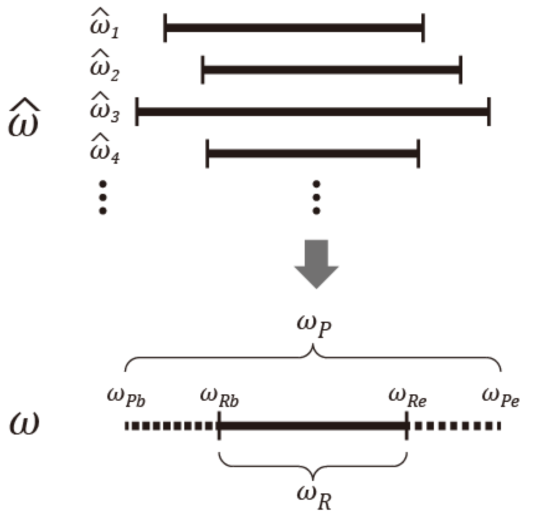

Figure 1: Basic concept of uncertain time interval and relation with determinate time intervals. Feature of uncertain time interval is represented by possible and reliable time intervals.

We considered a time interval consisting all determinate time intervals $\widehat{\omega}$ included in an uncertain time interval ω, and defined as a "possible time interval", denoted as ω_P. The possible time interval indicates the whole temporal range where the true time interval associated with the uncertain time interval ω may be located. Assuming the time intervals in this study, the beginning point of the possible time interval (ω_{Pb}) is the minimum value of the beginning points of all determinate time intervals ($\widehat{\omega}_b$), and the ending point (ω_{Pe}) is the maximum value of the ending points of all determinate time intervals ($\widehat{\omega}_e$) (Equation 2):

$$\omega_P \coloneqq \bigcup_{\lambda \in \Lambda} \widehat{\omega}_\lambda$$
$$\omega_{Pb} = \min_{\lambda \in \Lambda} \widehat{\omega}_{b_\lambda}$$
$$\omega_{Pe} = \max_{\lambda \in \Lambda} \widehat{\omega}_{e_\lambda}$$

Equation 2

By contrast, we considered the time interval common to all determinate time intervals $\widehat{\omega}$ included in an uncertain time interval ω. We defined this time interval as "reliable time interval" denoted as ω_R. The reliable time interval indicates the temporal range usually included in any true temporal intervals associated with the uncertain time interval ω. Similarly, for the possible time interval, the beginning point of the reliable time interval (ω_{Rb}) is the maximum value of the beginning points of all determinate time interval ($\widehat{\omega}_b$), and the ending point (ω_{Re}) is the minimum value of the ending points of all determinate time interval ($\widehat{\omega}_e$) (Equation 3):

$$\omega_R \coloneqq \bigcap_{\lambda \in \Lambda} \widehat{\omega}_\lambda$$
$$\omega_{Rb} = \max_{\lambda \in \Lambda} \widehat{\omega}_{b_\lambda}$$
$$\omega_{Re} = \min_{\lambda \in \Lambda} \widehat{\omega}_{e_\lambda}$$

Equation 3

We represented the feature of an uncertain time interval ω using these possible and reliable time intervals (Figure 1). The beginning points ($\widehat{\omega}_b$) and the ending points ($\widehat{\omega}_e$) of determinate time intervals are located between ω_{Pb} and ω_{Rb} and between ω_{Re} and ω_{Pe}, respectively (Equation 4):

$$\forall \widehat{\omega}_b \, \forall \widehat{\omega}_e \, (\omega_{Pb} \leq \widehat{\omega}_b \leq \omega_{Rb} \wedge \omega_{Re} \leq \widehat{\omega}_e \leq \omega_{Pe} \wedge \widehat{\omega}_b \leq \widehat{\omega}_e)$$

Equation 4

These ranges correspond to the ranges for the beginning and ending points of usual uncertain time intervals. In the preceding example, the range between ω_{Pb} and ω_{Rb} corresponds to "the 15th century", and that between ω_{Re} and ω_{Pe} corresponds to "the 1560s." Therefore, the range between January 01, 1401, and December 31, 1569, which is the possible time interval, is the period when the building may have existed. By contrast, the range between January 01, 1501, and December 31, 1559, which is the reliable time interval, is the period when it is certain that the building existed. The existence of the building is clearly denied before January 01, 1401, and after December 31, 1569.

A reliable time interval is a subset of a possible time interval. Therefore, when both time intervals are equal (i.e., $\omega_P - \omega_R = \emptyset$), the determinate time interval derived from them is only one. It shows the true time interval. By contrast, when both time intervals are not equal (i.e., $\omega_P - \omega_R \neq \emptyset$), it shows that one or both borders are uncertain.

In the possible time interval, the beginning point (ω_{Pb}) is always equal or smaller than the ending point (ω_{Pe}) (Equation 5). However, this relationship may not always hold for the reliable time intervals. Because the reliable time interval is a subset of the possible time interval, the beginning and ending points of the reliable time interval are never located out of the possible time interval. However, the beginning point (ω_{Rb}) may be sometimes larger than the ending point (ω_{Re}). For example, in the case of "a building that existed from the first half of the 15th century to 1440s," according to Equation 3, ω_{Rb} is December 31, 1449, and ω_{Re} is January 01, 1440. In this case, because of the overlapping of the temporal ranges for the beginning and the ending points, no range represents the certainty when the building existed. Hence, when ω_{Rb} is larger than ω_{Re}, it is defined that the reliable time interval does not exist (Equation 6). Similarly, in the case of words "a building that existed in the 15th century", ranges for both $\widehat{\omega}_b$ and $\widehat{\omega}_e$ are the 15th century. As ω_{Rb} is equal to ω_{Pe}, and ω_{Re} is equal to ω_{Pb}, the reliable time interval does not exist in this case, according to Equation 5 and Equation 6:

$$\forall \omega_{Pb} \forall \omega_{Pe} \left(\omega_{Pb} \leq \omega_{Pe}\right)$$

Equation 5

$$\forall \omega_{Rb}\ \forall \omega_{Re} \left(\omega_{Pb} \leq \omega_{Rb} \leq \omega_{Pe} \wedge \omega_{Pb} \leq \omega_{Re} \leq \omega_{Pe}\right)$$
$$\omega_{Rb} \leq \omega_{Re} \Rightarrow \omega_R \neq \emptyset$$
$$\omega_{Rb} > \omega_{Re} \Rightarrow \omega_R = \emptyset$$

Equation 6

This concept about the uncertain time intervals is close idea to rough time intervals based on the rough set theory (Bassiri et al., 2009). The possible and reliable time intervals correspond with upper and lower approximations in the rough time interval, respectively. However, we defined the uncertain time intervals as sets of determinate time intervals that may be the true time interval (Equation 1); therefore, they are different from the rough time intervals focused only on boundaries. Furthermore, in this study, we clarified the condition for existence of the reliable time intervals while corresponding to the actual data (Equation 6). Although a reliable time interval does not exist due to the condition $\omega_{Rb} > \omega_{Re}$, both values are held, because these values are essential to indicate ranges for the beginning and ending points in the uncertain time interval. By contrast, when a lower approximation in a rough time interval does not exist, the beginning and ending points do not exist.

3. Relationships between Uncertain Time Intervals

Allen (1983) classified the relative positions between two time intervals into 13 types (Table 1). This classification is widely used as "the Allen's relations" to examine relationships between time intervals. Here, the relations between two determinate time intervals *s* and *t* can be represented as a function $relT\ (s,t)$ (Equation 7):

$$s \coloneqq \{x | s_b \leq x < s_e\}$$
$$t \coloneqq \{y | t_b \leq y < t_e\}$$

$$\begin{aligned}
\text{before } (s,t) &\coloneqq (s_e < t_b) \\
\text{after } (s,t) &\coloneqq (s_b > t_e) \\
\text{during } (s,t) &\coloneqq (s_b > t_b \wedge s_e < t_e) \\
\text{contains } (s,t) &\coloneqq (s_b < t_b \wedge s_e > t_e) \\
\text{overlaps } (s,t) &\coloneqq (s_b < t_b \wedge s_e > t_b \wedge s_e < t_e) \\
\text{overlapped-by } (s,t) &\coloneqq (s_b > t_b \wedge s_b < t_e \wedge s_e > t_e) \\
\text{meets } (s,t) &\coloneqq (s_b < t_b \wedge s_e = t_b \wedge s_e < t_e) \\
\text{met-by } (s,t) &\coloneqq (s_b > t_b \wedge s_b = t_e \wedge s_e > t_e) \\
\text{starts } (s,t) &\coloneqq (s_b = t_b \wedge s_e < t_e) \\
\text{started-by } (s,t) &\coloneqq (s_b = t_b \wedge s_e > t_e) \\
\text{finishes } (s,t) &\coloneqq (s_b > t_b \wedge s_e = t_e) \\
\text{finished-by } (s,t) &\coloneqq (s_b < t_b \wedge s_e = t_e) \\
\text{equals } (s,t) &\coloneqq (s_b = t_b \wedge s_e = t_e)
\end{aligned}$$

Equation 7

Using Equation 7, we examined the relations between two uncertain time intervals. Figure 2 shows the Allen's "before" relation between the uncertain time intervals *a* and *b*.

According to Equation 7, when a determinate time interval $\hat{a}$ of *a* is located before a determinate time interval $\hat{b}$ of *b*, and the ending point of $\hat{a}$ is located before the beginning point of $\hat{b}$ (i.e., $\hat{a}_e < \hat{b}_b$), the relation between the determinate time intervals is the "before" relation.

Table 1: Thirteen relations between two time intervals according to the Allen's relations (Allen, 1983)

Relation	Relative position	Relation	Relative position
s before *t*	s t	*s* after *t*	t s
s during *t*	t s	*s* contains *t*	s t
s overlaps *t*	s t	*s* overlapped-by *t*	t s
s meets *t*	s t	*s* met-by *t*	t s
s starts *t*	s t	*s* started-by *t*	s t
s finishes *t*	t s	*s* finished-by *t*	s t
s equals t	s t		

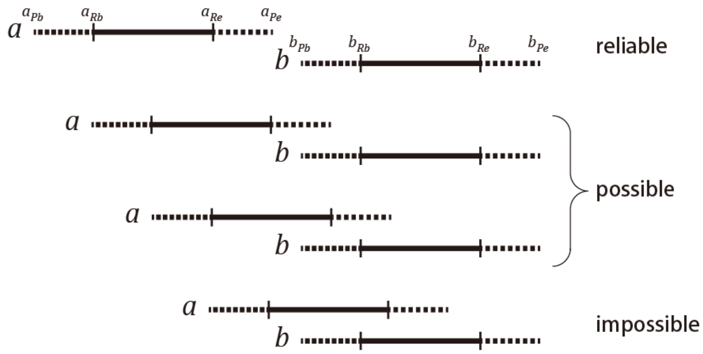

Figure 2: A change in the states of the "before" relation due to the difference in relative positions of two uncertain time intervals

Here, if a_{Pe} is smaller than b_{Pb}, irrespective of the state the uncertain time interval *a* and *b*, it always becomes $\hat{a}_e < \hat{b}_b$; so, although the two time intervals are uncertain, it is certain there is the "before" relation between them. Expanding this concept to the other relations, when the determinate time intervals $\hat{a}$ and $\hat{b}$ satisfy Equation 8, they are always the given relation, and are in state of reliable.

$$\forall \hat{a}\ \forall \hat{b}\ relT\ (\hat{a}, \hat{b})$$

Equation 8

In this study, we defined this relation as "reliable relation." In the "before" relation, the condition of the reliable relation is:

$$\forall \hat{a}\ \forall \hat{b}\ (\hat{a}_e < \hat{b}_b)$$
$$a_{Pe} < b_{Pb} \Rightarrow \forall \hat{a}\ \forall \hat{b}\ \text{before}\ (\hat{a}, \hat{b})$$

In the same manner, for other relations between uncertain time intervals, we obtained the conditions satisfying the reliable relations (Table 2). Furthermore, we examined the cases where the uncertain time interval *a* moves closer to *b*. In the second state in Figure 2, the possible time intervals of *a* and *b* overlap. In this case, some determinate time intervals $\hat{a}$ may be before some determinate time intervals $\hat{b}$, and some of them may overlap. Therefore, the "before" relation is possible, but is does not always satisfy that. The third state in Figure 2 shows the possible and reliable time intervals of *a* and *b* overlapping each other. In this case, both the states, in which the relation between *a* and *b* is the "before" relation or not, are possible. In the fourth state of Figure 2, the reliable time intervals of *a* and *b* are overlapping each other. In this case, it is certain that all determinate time interval $\hat{a}$ and $\hat{b}$ overlap each other, and therefore, *a* and *b* never represent the "before" relation. In this study, we defined the second and third states in Figure 2 as the "possible relation", and the fourth state as the "impossible relation."

Table 2: Conditions that two uncertain time intervals are states in the reliable or the possible relation.

Relation	Reliable	Possible
before	$a_{Pe} < b_{Pb}$	$a_{Re} < b_{Rb}$
after	$a_{Pb} > b_{Pe}$	$a_{Rb} > b_{Re}$
during	$a_{Pb} > b_{Rb} \wedge a_{Pe} < b_{Re}$	$a_{Rb} > b_{Pb} \wedge a_{Re} < b_{Pe} \wedge b_{Pb} < b_{Pe}$
contains	$a_{Rb} < b_{Pb} \wedge a_{Re} > b_{Pe}$	$a_{Pb} < b_{Rb} \wedge a_{Pe} > b_{Re} \wedge a_{Pb} < a_{Pe}$
overlaps	$a_{Rb} < b_{Pb} \wedge a_{Re} > b_{Rb} \wedge a_{Pe} < b_{Re}$	$a_{Pb} < b_{Rb} \wedge a_{Pe} > b_{Pb} \wedge a_{Re} < b_{Pe}$ $\wedge a_{Pb} < a_{Pe} \wedge b_{Pb} < b_{Pe}$
overlapped-by	$a_{Pb} > b_{Rb} \wedge a_{Rb} < b_{Re} \wedge a_{Re} > b_{Pe}$	$a_{Rb} > b_{Pb} \wedge a_{Pb} < b_{Pe} \wedge a_{Pe} > b_{Re}$ $\wedge a_{Pb} < a_{Pe} \wedge b_{Pb} < b_{Pe}$
meets	$a_{Re} = a_{Pe} = b_{Pb} = b_{Rb}$ $\wedge a_{Rb} < b_{Pb} \wedge a_{Pe} < b_{Re}$	$a_{Re} \leq b_{Rb} \wedge a_{Pe} \geq b_{Pb} \wedge a_{Pb} < b_{Rb}$ $\wedge a_{Re} < b_{Pe} \wedge a_{Pb} < a_{Pe} \wedge b_{Pb} < b_{Pe}$
met-by	$a_{Pb} = a_{Rb} = b_{Re} = b_{Pe}$ $\wedge a_{Pb} > b_{Rb} \wedge a_{Re} > b_{Pe}$	$a_{Pb} \leq b_{Pe} \wedge a_{Rb} \geq b_{Re} \wedge a_{Rb} > b_{Pb}$ $\wedge a_{Pe} > b_{Re} \wedge a_{Pb} < a_{Pe} \wedge b_{Pb} < b_{Pe}$
starts	$a_{Pb} = a_{Rb} = b_{Pb} = b_{Rb} \wedge a_{Pe} < b_{Re}$	$a_{Pb} \leq b_{Rb} \wedge a_{Rb} \geq b_{Pb} \wedge a_{Re} < b_{Pe}$ $\wedge b_{Pb} < b_{Pe}$
started-by	$a_{Pb} = a_{Rb} = b_{Pb} = b_{Rb} \wedge a_{Re} > b_{Pe}$	$a_{Pb} \leq b_{Rb} \wedge a_{Rb} \geq b_{Pb} \wedge a_{Pe} > b_{Re}$ $\wedge a_{Pb} < a_{Pe}$
finishes	$a_{Pb} > b_{Rb} \wedge a_{Re} = a_{Pe} = b_{Re} = b_{Pe}$	$a_{Rb} > b_{Pb} \wedge a_{Re} \leq b_{Pe} \wedge a_{Pe} \geq b_{Re}$ $\wedge b_{Pb} < b_{Pe}$
finished-by	$a_{Rb} < b_{Pb} \wedge a_{Re} = a_{Pe} = b_{Re} = b_{Pe}$	$a_{Pb} < b_{Rb} \wedge a_{Re} \leq b_{Pe} \wedge a_{Pe} \geq b_{Re}$ $\wedge a_{Pb} < a_{Pe}$
equals	$a_{Pb} = a_{Rb} = b_{Pb} = b_{Rb}$ $\wedge a_{Re} = a_{Pe} = b_{Re} = b_{Pe}$	$a_{Pb} \leq b_{Rb} \wedge a_{Rb} \geq b_{Pb}$ $\wedge a_{Re} \leq b_{Pe} \wedge a_{Pe} \geq b_{Re}$

Applying this concept to other relations, when the determinate time intervals $\hat{a}$ and $\hat{b}$ satisfy Equation 9, they may be the given relation, namely, they are in state of possible relation.

$$\exists \hat{a}\, \exists \hat{b}\; relT(\hat{a}, \hat{b})$$

Equation 9

In case of the "before" relation, the condition of the possible relation is:

$$\exists \hat{a}\, \exists \hat{b}\; (\hat{a}_e < \hat{b}_b) \equiv \neg(\forall \hat{a}\, \forall \hat{b}\; (\hat{a}_e \geq \hat{b}_b))$$
$$a_{Re} < b_{Rb} \Rightarrow \exists \hat{a}\, \exists \hat{b}\; \text{before}(\hat{a}, \hat{b})$$

Similarly, we obtained the conditions satisfying the possible relations for other relations between uncertain time intervals (Table 2). The impossible relation can be expanded to other relations as Equation 10:

$$\forall \hat{a}\, \forall \hat{b}\; \neg relT\; (\hat{a}, \hat{b})$$

Equation 10

The conditions satisfying impossible relation for each relation between two uncertain time intervals can be easily derived from the conditions of the possible relation shown in Table 2 because they are negation of the condition of the possible relation.

Even if the reliable time intervals do not exist (i.e. $a_{Rb} > a_{Re}$ or $b_{Rb} > b_{Re}$), boundaries of the reliable time intervals shown in Table 2 are valid. Figure 3 shows an example of uncertain intervals *a* and *b* which do not have reliable time intervals. A condition in which "before" relation between the uncertain time intervals *a* and *b* is impossible, is $a_{Re} \geq b_{Rb}$, and the condition can be derived form a condition for possible "before" relation ($a_{Re} < b_{Rb}$) in Table 2. These uncertain time intervals *a* and *b* in Figure 3 satisfy this condition and thus it is impossible that they have the "before" relation. The whole range of the ending point in *a* is located after the whole range of beginning point in *b* in Figure 3, and this indicates that relation of *a* "before" *b* is actually impossible.

As mentioned above, this study's concept about uncertain time interval is close idea to rough time intervals. However, the concept about uncertain time interval has great advantage in examining relationship between time intervals. An uncertain time interval always holds values of both boundaries of the reliable time interval, even if the reliable time interval does not exist like the situation in Figure 3. Therefore, we can examine relationships between uncertain time intervals, always using conditions in Table 2. By contrast, when a rough time interval does not have a lower approximation corresponding with the reliable time interval, the time interval does not hold values of boundaries of the lower approximation, and we cannot apply the same procedures like Table 2. It may be possible that values of boundaries of the upper approximation are used instead of the lower approximation.

In this situation, Allen's "before" relation looks like possible, when the two upper approximations overlap each other. However, as the case of Figure 3, there are situations that the "before" relation is actually impossible even though possible time intervals, corresponding with the upper approximations, are overlap. This means the result examined using the upper approximation instead of the lower approximation is fault. Hence, it is difficult to examine relations between two rough time intervals when they do not have lower approximations.

When considering whether relations between two uncertain time intervals *a* and *b* are reliable, possible, or impossible for all 13 relations, if *a* or *b* is uncertain (i.e., $a_P - a_R \neq \emptyset \vee b_P - b_R \neq \emptyset$), they are one of the following states: 1) reliable for one relation and impossible for the remaining 12 relations 2) possible for more than one relation and impossible for the remaining relations. While, if *a* and *b* is not uncertain (i.e., $a_P - a_R = \emptyset \wedge b_P - b_R = \emptyset$), they are the state in which reliable for one relation and impossible for the remaining 12 relations, and are the same as the original Allen's relation.

4. Creating Uncertain Time Intervals

Because we represented the boundaries of uncertain time intervals as the temporal range according to Equation 4, uncertain time intervals were created by specifying the ranges of the boundaries. We used the periods based on calendars, such as date, month, year, and century to specify these ranges. In these periods, as the boundaries were accurately determined, similar to the time interval *a* of Figure 4 a created uncertain time interval was used as the period as the range of the boundary as it is. However, things associated with periods, such as eras, persons, and events, are used to specify the boundaries of time intervals (e.g., "from the Tang Dynasty" or "during the president's term"). Although these periods are often uncertain time intervals, we used them to create uncertain time intervals, because ranges for the boundaries can be specified using the possible time intervals of the periods (time interval *b* in Figure 4). This indicates that uncertain time intervals can be recursively used to create other uncertain time intervals.

Although it is assumed that both boundaries of an uncertain time interval have true value, there are time intervals without fixed boundaries in practice. For example, historical periods (e.g. middle ages) often gradually shifted to the next periods and did not suddenly shift to the next at a certain time instant. Therefore, such time intervals do not have fixed boundaries. In this situation, ranges of the ending of the previous period and the beginning of the next period overlap. Similarly, periods used in natural science (e.g., the last glacial period) do not have fixed boundaries.

It is also possible that we apply uncertain time intervals of this study to such time intervals without fixed boundaries. That is because description and examination of relationships for uncertain time intervals can be done without determining true value of boundaries. When ranges of the ending and beginning points of two continuous historical periods overlap, relations between events in the overlapped range are possible relation.

A word of "circa" is also often used to indicate a range of boundaries of a time interval (e.g., "circa 1300 BCE").

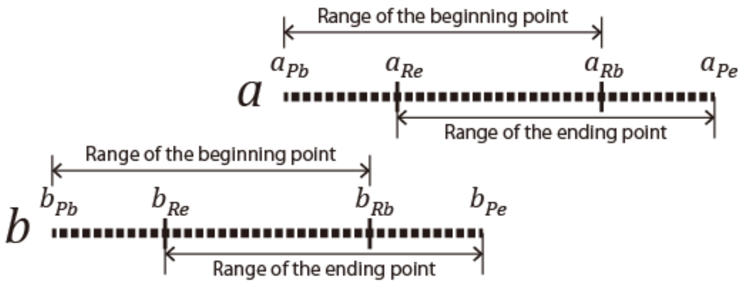

Figure 3: A relative positions of two uncertain time intervals without reliable time intervals

Figure 4: A schematic of creating uncertain time interval. Time interval *a* is not uncertain, but time interval *b* is uncertain (uncertain time interval). Uncertain time interval *c* is created from *a* and *b* specifying the range for the beginning and ending points.

It is necessary to know the range indicated by each "circa" in order to generate an uncertain time interval from such description about boundaries. However, in practice, since it is usually difficult to obtain the range, we often use the range derived from significant figures instead. For example, we can assume "circa 1300 BCE" as a range from 1400 BCE to 1200 BCE. In such expression about a boundary, it is appropriate to assume widely the range of the boundary. The wider the range of a boundary, the wider a possible time interval and the narrower a reliable time interval are generated. Therefore, the following fatal error can be avoided: 1) a range which is not actually a part of a time interval, is included in the reliable time interval, 2) a range which is actually a part of a time interval, is not included in the possible time interval.

5. Using the Linked Data and RDFs

We created the uncertain time intervals recursively. We used data in the same database and resources on the Internet to create time intervals. Thus, uncertain time intervals had high affinity with the Linked Data, a mechanism to link resources each other on the Internet.

Although RDF is usually used for the Linked Data, for representing determinate time intervals as RDF data, the method describing calendrical periods such as dates, months, and years is problematic. According to the RDF regulation, the calendrical periods should be expressed as literal values with ISO 8601 format (W3C, 2014a). As these literal values are not linked to other resources or values, their temporal ranges cannot be obtained using the Linked Data mechanism. Moreover, the ISO 8601 format does not assume representations such as "1560s," and lacks flexibility. To solve this problem, Sekino (2017) released RDF resources of the calendrical periods. These resources cover dates and periods of months, years, decades, and centuries. As these are resources and not literal values, the beginning and ending points represented by Julian Day are linked. The Julian Day is the total day count from noon on January 01, 4713 BCE (Dershowitz and Reingold, 2007) and expressed as a real number. Therefore, it is suitable for calculating the relative position on the temporal axis between the temporal data. Moreover, the calendrical period resources of Sekino (2017) supports periods based on calendars other than the Gregorian calendar. Thus, resources for events described using local calendar, such as luni-solar calendar, can be created using their original period description as they are. Because the beginning and ending points of the periods based on the local calendar are also represented by Julian Days, the relationships between the periods can be examined even if the periods are based on different calendars.

Figure 5 shows an example of uncertain time intervals represented by the RDF. We represented the existence period (from the 15th century to the 1560s) of the building aforementioned as an RDF resource using calendrical period resources. Furthermore, we represented the lifetime of a person born in the building and died in 1572 using the existing period of the building. In this case, we specified the range for the beginning point of the lifetime using an uncertain time interval recursively, which represents the existence period of the building. Therefore, we represented the lifetime as an uncertain time interval using links between resources (Figure 6). We represented the positions of the possible and reliable time intervals of the lifetime on the temporal axis as from 2232773.5 to 2295596.5 and 2294500.5 to 295230.5 in Julian Days, respectively.

Using the obtained Julian Days, relationships between the lifetime of the person and existence period of the building can be examined according to Table 2. These are possible for "started-by", "overlapped-by", and "met-by" relations and are impossible for the remaining relations. However, it seems odd to apply "started-by" for the relation between existence periods of a building and a person, because it is not usually possible that birth of a person in a building and completion of the building are simultaneous. Similarly, "met-by" is also difficult to apply for the relation. By contrast, it is possible to apply "started-by" for relation between existence periods of a building and a sculpture or a mural attached to the building. These facts mean that which type of relation can be applied to a relation, is depending on the context of the relation. In practice, it is difficult that we select appropriate Allen’s relations to large amount of data, depending on the context, and therefore, we will simplify this operation into the same procedure. Since fatal errors can be avoided by taking range of a boundary of a time interval widely, as mentioned above, it is better that relations of "started-by" and "met-by" are included in the simplified operation.

We can perform this comparison directly using the RDF data. For retrieving the RDF data, we used SPARQL, a query language widely used (W3C, 2008). As the SPARQL has a mechanism that defines new functions used in the query sentence, we provided functions to examine the relationships between uncertain time intervals, according to Table 2. Using these functions, we retrieved and analyzed the temporal data containing uncertain time intervals.

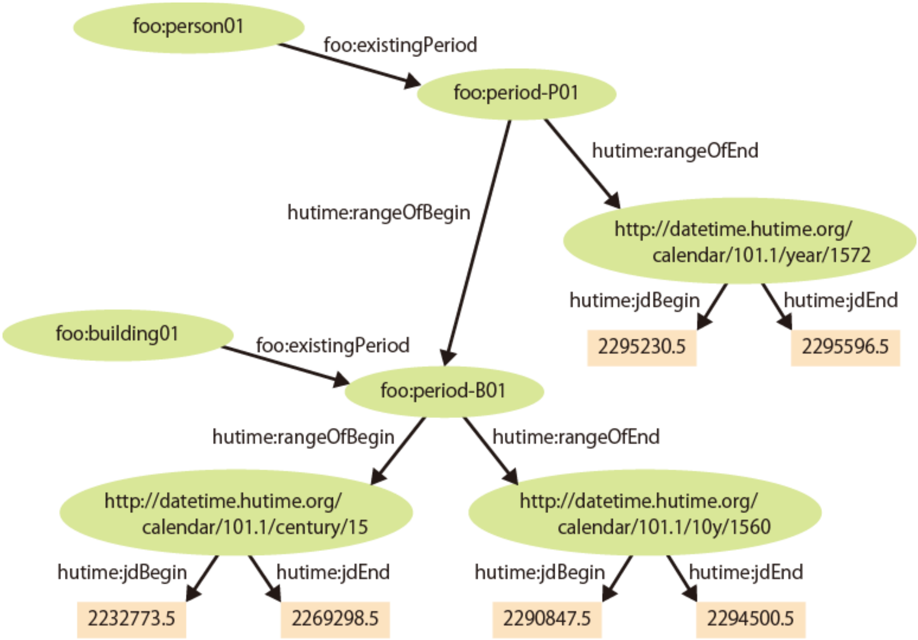

Figure 5: An RDF graph illustrating the existence period of a building and lifetime of a person. A range for the beginning point of the person's lifetime is specified by the existence period of the building. This means the person was born in the building. Ranges for the other beginning and ending points are specified by calendrical period resources.

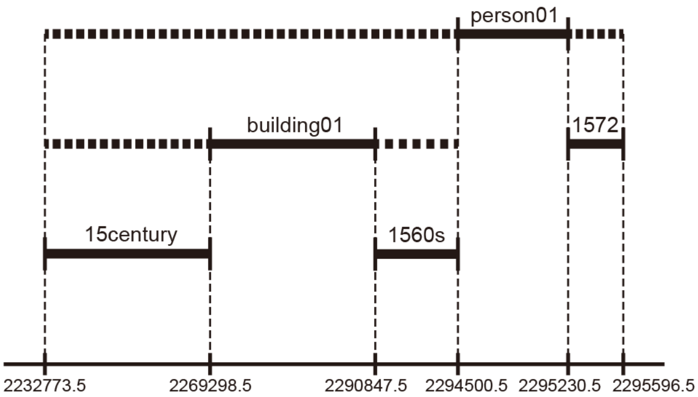

Figure 6: Time intervals drawn according to the RDF data in Figure 5. Although the relative positions of the intervals are accurate, durations of the intervals are not necessarily accurate, because some intervals are drawn with short due to a limitation of space for printing. Numbers on the temporal axis placed on the bottom are Julian Days that can be obtained using the links between RDF resources.

For example, a user can retrieve events that may be within the given range, in addition to the events that are definitely in the range. The following is an example of an assumed SPARQL query and returns events that may have occurred during the 17th century.

```
PREFIX hutime: <http://resource.hutime.org/ontology/>
PREFIX hfunc: <http://resource.hutime.org/function/>
PREFIX hcal: <http://datetime.hutime.org/calendar/>
PREFIX rdfs:
   < http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX time: <http://www.w3.org/2006/time#>
SELECT ?event
WHERE { ?event rdfs:type hutime:timeInterval .
   FILTER ( hfunc:possibleRelation($event,
     time:intervalDuring, hcal:1001.1/cantury/17) ) .
 }
```

In order to examine relations according to Table 2 it is required to obtain positions of a temporal axis (i.e. values of Julian Day) for boundaries of possible and reliable time intervals. In the above sample query, it is assumed that each event is associated with these values using stipulated vocabularies, and these operations are done in the function "hfunc:possibleRelation". To realize such operation, it is necessary to construct related vocabularies, and implement functions according to the vocabularies, in the future works. Recently, OWL-time which is a standard vocabulary to describe information about time, was published as a W3C recommendation (W3C, 2017). Although it is difficult to directly use OWL-Time for uncertain time interval of this study, the constructed vocabularies should be associated with OWL-Time using RDF Schema (W3C, 2014b).

6. Conclusion

In this study, we represented the feature of an uncertain time interval by two types of time intervals (Figure 1): the possible time interval indicating the whole temporal range where the true temporal interval may be located, and the reliable time interval indicating the temporal range always included in any true temporal interval. We created the uncertain time intervals by specifying ranges for the beginning and ending points, using the periods based on calendars and periods associated with thing such as events, persons, and buildings. We showed that this procedure can be realized as the RDF data in the Linked Data. Moreover, we revealed conditions that two uncertain time intervals become to be states in the reliable or possible relation for each Allen's relation (Table 2). We implemented these conditions as a function for SPARQL; therefore, users can retrieve RDFs of uncertain temporal data containing uncertain time intervals under arbitrary conditions described in the queries of SPARQL, thereby indicating that the appropriate handling of uncertain temporal data can be realized on the Linked Data world. We believe that using uncertain temporal data will progress in various disciplines.

Acknowledgment

This study is mainly supported by JSPS KAKENHI "Semantic chronology – Construction of infrastructure for visualization and utilization of knowledge along temporal axis" (15H01723) and partially supported by JSPS KAKENHI 16H01897, 16H01830, 16H02918, 17H00773, and 23300097.

References

Allen, J., 1983, Maintaining Knowledge about Temporal Intervals. *Communications of the ACM*, Vol. 26(11), 832-843.

Asmussen, K., Qiang, Y., DeMaeyer, P. and Van DeWeghe, N., 2009, Triangular Models for Studying and Memorising Temporal Knowledge. *Proceedings of the International Conference on Education, Research and Innovation*, 1849-1859.

Bassiri, A., Malek, M. R., Alesheikh, A. A. and Amirian, P., 2009, Temporal Relationships between Rough Time Intervals. *In ICCSA 2009, Part I. LNCS*, Vol. 5592, edited by Gervasi, O. et al. (Springer).

Billiet, C. and De Tré, G., 2016, The Role of Computational Intelligence in Temporal Information Retrieval: A Survey of Imperfect Time in Information Systems, *In Flexible Query Answering Systems 2015*, edited by Andreasen, T. et al. (Springer).

Billiet, C., Pons, J. E., Matthé, T., De Tré, G. and Pons Capote, O., 2011, Bipolar Fuzzy Querying of Temporal Databases. *In FQAS 2011. LNCS*, edited by Christiansen, H. et al. (Springer).

Bizer, C., Heath, T. and Berners-Lee, T., 2009, Linked Data - The Story So Far, *International Journal on Semantic Web and Information Systems*, Vol. 5(3), 1-22.

Dershowitz, N. and Reingold, E. M., 2007, *Calendrical Calculations*, (Cambridge University Press).

Nagypál, G. and Motik, B., 2003, A fuzzy model for representing uncertain, subjective, and vague temporal knowledge in ontologies. *In CoopIS 2003, DOA 2003, and ODBASE 2003. LNCS*, edited by Meersman, R. and Schmidt, D. C. (Springer).

Qiang, Y., Asmussen, K., Delafontaine, M., Stichelbaut, B., De Tré, G., De Maeyer, P. and Van De Weghe, N., 2009, Visualising Rough Time Intervals in a Two-Dimensional Space. *Proceedings of IFSA World Congress/EUSFLAT Conference*, 1480-1485.

Qiang, Y., Delafontaine, M., Asmussen, K., Stichelbaut, B., De Maeyer, P. and Van De Weghe, N., 2010, Modelling Imperfect Time Intervals in a Two-Dimensional Space, *Control and Cybernetics*, Vol. 39(4), 983-1010.

Sekino, T., 2017, Basic Linked Data Resource for Temporal Information. *Proceedings of the 2017 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC)*, 76-82.

W3C, 2008, SPARQL Query Language for RDF, https://www.w3.org/TR/rdf-sparql-query/.

W3C, 2014a, RDF 1.1 Concepts and Abstract Syntax, http://www.w3.org/TR/rdf11-concepts/.

W3C, 2014b, RDF Schema 1.1, https://www.w3.org/TR/rdf-schema/.

W3C, 2017, Time Ontology in OWL, https://www.w3.org/TR/owl-time/.