# Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 12 (3): 284-288, 2012
ISSN 1812-5654 / DOI: 10.3923/jas.2012.284.288


# Managing News Archive Using Temporal Data Modeling

Noraida Haji Ali and Sumazly Sulaiman
Department of Computer Science, Faculty of Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

**Abstract:** This study focuses on the usage of temporal concept in news archiving. Specifically, we present temporal data using timestamp technique and information retrieval using indexing technique for news history data management. News can be classified into four temporal characteristics such as title of news, time status, categories and year era. Time status and year era characteristics involve temporal logic of the retrieval process which can be categorized into five modes, i.e., before, after, interval, meet by and meets. Based on this concept, the combination of reference material of news archiving becomes easier, more efficient and accurate.

**Key words:** Temporal databases, temporal model, news archiving, time-stamping, relational model

## INTRODUCTION

News generally involves any occurrences that can be seen by a reporter or another person and what they have seen will purposely translate to inform to other people. The function of news to the society is to maintain the peace and to generate the social changing. A news archiving is an important thing to the news that happen by classifying the information or stories and explaining the relationship between them that happen to be segregated by time and space. The ordinary news will spread widely through the variety of mediums such as through publishers (as a printer-media) and through broadcast (as an electronic media). Therefore time is very important to ensure that the news broadcasted or not. Time can be divided into three aspects as follows:

- **Recency:** Is the time when news is announced as does public interest? If the news of any issues is spread before hand, it will confuse to all readers. As such, the readers? belief towards that news will decrease. For example, news about match results of any sports if the match sports have not even held yet
- **Immediacy:** Is the time where the news will be broadcasted immediately? This aspect will avoid any news that could not be trusted again. For example, news about the flood condition or any main roads that should be closed, so that people will know the real condition immediately
- **Currency:** Is the time to ensure news suitability based on its environment or any discussion issues? This aspect will avoid any riot from the public if the news is against to any current incident. For example, it is irrelevant to write any kindness of someone at the same time when the issue of his arrest of any overlapped cases, yet had not been strongly proved by the court

In data archiving technology system, it shows the expeditions of space in data storage, time and facilities to store and to recall the required data.

Time is an attribute of all real-world phenomena. Consequently, efforts to incorporate the temporal domain into Database Management Systems (DBMSs) have been ongoing for more than a decade (Tsotras and Kumar, 1996). Temporal database stores time-varying information. It supports different aspects of time and user-defined time is not a part of the temporal database concept. Temporal databases record 'when certain actions are taking place' or 'when data is true' in the mini world. So, the temporal database can represent the evolution of an entity in the database with respect to time. In order to achieve these features, different kinds of time-stamps will be added to data. One type of time-stamps, transaction time is the time that represents when the data is added or stored in the database. Another type of time-stamp, valid time is the time that represents when data is true or valid in the mini world. When both valid-time and transaction-time time-stamps are used in a temporal database, then it is known as bitemporal database. There are different ways of representing the time-stamps-as time instants and time interval (Edelweiss *et al*., 2000; Jensen and Snodgrass, 1999; Elmasri and Navathe, 2000).

Gao *et al*. (2005) presented a comprehensive and systematic study of join operations in temporal

**Corresponding Author:** Noraida Haji Ali, Department of Computer Science, Faculty of Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

databases, including both semantics and implementation. Tanin *et al*. (2009) proposed an approach to exploit the spatial and temporal nature of data and queries such as motion characteristics. A feasible way of implementing a temporal database is by mapping temporal features onto a conventional commercial database management system. Research by Unnikrishnan and Paramod (2009) did analysis of different possible ways of implementing coalescing operator of temporal databases by mapping the temporal data model into the underlying relational model followed by Oracle was done. The temporal time-stamps, implicit in the temporal database, are represented as does explicit attributes. Most researches focused on implementation of temporal features in the database. Not more to apply the temporal concepts in real problem such news archiving. This study aimed to discuss how the temporal concepts can apply to manage the news archive which is date is a very important features in the news.

## THE TEMPORAL CONCEPT IN NEWS ARCHIVING

Date is an important attribute in news archiving where attribute data temporal being used in the news archiving model is based on time of the transaction. All data of news facts are stored in the database on time point. The time of transaction of the database is the time when the news is stored with the serialized-consecutive transaction. The values of the transactional time never exist after the current time. Every accumulative news will only happen on the current time point. Though, every news that to be reported to the public normally can be classified based-on the occurrences or related events. Therefore, for any time point there are varieties of classifications such as news in politics, entertainments, law and regulations and etc. To get the consistent database, we need to involve the temporal model data which the concept is temporal logical. The temporal logical is defined as a logic mode which the mode operation later on can explain time. For example, always, sometimes, next time, last time, since and so on. These modes are relevance to the researchers where date is an important attribute to generate news.

Based on the news necessity, the writer defines the date element called time point as:

$$\text{Time point} = (t, t+1, ..., t+j)$$

Where:

t = Starting date
t+1 = Date incremented by one time point from starting date
t+j = The final date that is current date

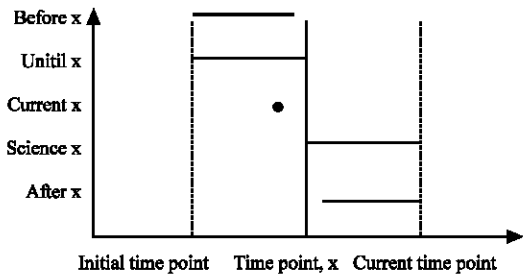


Fig. 1: Temporal operator relationship with time point

For temporal model to archive the news, model proposed by Allen (1981, 1983) will be adopted for time point and during time and used as a primitive to manage the temporal data. Let T represents a set of time elements or time points and d represents a function from T to $R_0$ (real number set). An element of time, t, is said during if d (t)>0. Thus it can be represented as:

$$T = I \cup P$$

Where:
I = A set of during time
P = A set of time point

Different temporal operators manage the question for during time. Allen has defined five temporal operators to explain their relationship with the time point Fig. 1.

## NEWS TEMPORAL DATA MODEL

Generally, the temporal relationship can be represented as:

$$\text{Temporal relationship } X (S, V, TS, TE)$$

Where:

X = Database
S = Surrogate
V = Data attribute
TS = Starting time
TE = Ending time

Based on the temporal relationship, X.TS<X.TE, data in the main database will be sorted in ascending order by the value of TS in order the attainment to become more impressive than any other point of time. These sequence order of times are called time stamping. Indexing is used in order to access data from the database easily.

**Temporal classification:** In the process of generating temporal table, some classification criterion must be constructed because it is closely related with the selected criterion by users. The classification can be shown as in Table 1.

**Temporal information representation:** Temporal information representation describes how the logical temporal combine with the time elements in the time status category. Year era category can be represented using time stamp method, therefore, each temporal information has the right and the left times to describe the duration of certain time. Time stamping can be represented as:

$$Ti = \{t_i | \omega_i (L) \leq t_i \leq \omega_i (R)$$

Where:

$t_i$ = The ith time point in the represented time space
$\omega_i (R)$ = The ith right time point
$\omega_i (L)$ = The ith left time point, $i \epsilon N$ = (integer number)

**Temporal representation for time status criteria:** Time status involves 4 data news time elements, i.e., date, week, month and year. A temporal representative, time stamp, will represent each time element in generating relational temporal. Each time elements will be represented by a temporal representative called time stamp in generating relational temporal. The relation in Table 2 describes how each status of the time element is presented.

**Temporal representation for yearly era category:** In yearly era category, it is defined as ten years complete time period where it started at the early year till the end of the year for each era. Era is considered as a year-attribute. The time stamp for each era is the era temporal representative. This can be shown in a Table 3.

For example, to determine the 90's era news criteria the date of the news involved, can be defined as:

$$\omega_n (L) = 1/1/1990$$

$$\omega_n (R) = 31/12/1999$$

if

$$t_1 = 6/6/1995$$

Then, it can be classified in the 90's era

where:

$$\omega_1(L) \leq 6/6/1995 \leq \omega_1(R)$$

Table 1: Temporal classification

| News topic | Based on the news issues |
|---|---|
| Time status | Date-news arise on certain date basis |
| | Week-news arise on weekly basis |
| | Month-news arise on monthly basis |
| | Year-news arise on yearly basis |
| News category | News arise based on type of category |
| News year-era | News arise based on every 10 years |
| Keywords | News arise based on keyword of certain topic |

Table 2: Temporal representative for time status

| Category time status | Type of status | Time stamp representative |
|---|---|---|
| Date | Current | $t_i = Now(\grave{u}_i(R))$ |
| | Before | $\omega_i (L) < t_i$ |
| | After | $\omega_i (R) > t_i$ |
| | During | $\omega_i (L) \leq t_i (L)$, $t_i (R) \leq \omega_i (R)$ |
| | Since | $\omega_i (R) \leq t_i$ |
| | Until | $\omega_i (L) \leq t_i$ |
| Week | 1-3 weeks | $\omega_i (R) - (w_j * 7) \leq t_i \leq Now (\omega_i (R))$, where $w_j$ is j's week, for j = 1,2,3 |
| Month | 1-11 months | $\omega_i (R) - (m_j * Nod) \leq t_i \leq Now (\omega_i (R))$, where $m_j$ is j's month, for j = 1,2, ...,11<br>Nod = {28,29,30,31} based on month of the current date |
| | Along | $\omega_i (L) \leq t_i \leq \omega_i (R)$ (mj), where $m_j$ is j's month, for j = 1,2, ....,11<br>$\omega_i (L)$ is the early month of $m_j$ |
| Year | 1-9 years | $\omega_i (R) - (y_j * Nofyear) \leq t_i \leq Now (\omega_i (R))$, where $y_j$ is j's year, for j = 1,2, ...,9<br>Nofyear = {365,366} based on year of the current |
| | Along | $\omega_i (L) \leq t_i \leq \omega_i (R)$ (yj), where $y_j$ is j's year, for j = 1,2, ....,9<br>$\omega_i (L)$ is the early year of $y_j$ |

Table 3: Temporal representative for year era

| Code | Time stamp representation |
|---|---|
| $T_1$ | $\omega_1(L) \leq t_1 \leq \omega_1 (R)$ |
| $T_2$ | $\omega_2(L) \leq t_2 \leq \omega_2 (R)$ |
| . | . |
| . | . |
| . | . |
| $T_n$ | $\omega_n(L) \leq t_n \leq \omega_n(R)$ |

$\omega_n(L)$ is the early year of the nth year era and $\omega_n(R)$ is the last year of the nth year era.

## RELATIONAL DATA MODEL

In the relational data model, news that functioned as history data news, generally consists of six tuples as shown below:

$$B (D, T, NF, JB, KK, KJ)$$

Where:

B = A news database
D = The date for a certain topic
T = The topic
NF = The file name for a certain topic

JB = The type of news
KK = The code category of the topic
KJ = The code for category type

Thus, each $X_i$ in the database can be presented as:

$$B (D_j, T_{jk}, NF_{jk}, JB_{jk}, KK_{jk}, KJ_{jk})$$

Where:

D = {<d,j>| j$\epsilon$N$\cap$d$_j\epsilon$B}, N is an integer
$D_j$ = The date for the jth news
T = {<<t,j>d,k>| j$\epsilon$N$\cap$<t,d>$\epsilon T_{jk}$}
$T_{jk}$ = The kth topic of the j date
NF = {<<d,j>nf,k>| j$\epsilon$N$\cap$(nf,d>$\epsilon NF_{jk}$}
$Nf_{jk}$ = The file name for a kth topic of the j date
JB = {<<d,j>jb,k>| j$\epsilon$N$\cap$(jb,d>$\epsilon JB_{jk}$}
$Jb_{jk}$ = The type of news for a $k^{th}$ file of the j date
KK = {<<d,j>kk,k>| j$\epsilon$N$\cap$(kk,d>$\epsilon KK_{jk}$}
$Kk_{jk}$ = The code category for a $k^{th}$ topic of the j date
KJ = {<<d,j>kj,k>| j$\epsilon$N$\cap$(kj,d>$\epsilon KJ_{jk}$}
$Kj_{jk}$ = The code for a $k^{th}$ topic of the j date

The relationship between records can be defined in the algebraic forms. This will describe the continuity and its relationship between them. For example, on the date 1/1/1998 there are four news titles. Each title has its own contents and each content will be categorized as described below:

$D_{j=1} \neg T_{k=1,2,3,4} \neg NF_{k=1,2,3,4} \neg JB_{k=1,2,3,4} \neg KK_{k=1,2,3,4} \neg KJ_{k=1,2,3,4}$ (1/1/1998)

If j = 1 and k = 3, then:

$$\forall X_i \exists B = \{D_1, T_{13}, NF_{13}, JB_{13}, KK_{13}, KJ_{13}\} \ni X_i \epsilon B$$

**Relational model of generating temporal table:** To realize the accessing information based on the temporal concept, each selected temporal classification criteria has its own temporal table. The formation of this table is hidden from the user. The method used is based on time representative at the file level. Based on the main database news, there are only three tuples i.e., D, I and KK involved in choosing the criteria to retrieve data from users. Let R be the choosing criteria, then:

$$R \epsilon \{D, T, KK\}$$

Thus, the news temporal tables produced can be represented by, $B_R$ and:

$$B_R^i = \{R_d^i, R_t^i, Nf_{kk}^i\} \epsilon B$$

Where:

$$R_D^i = \sum_{j=1}^{n} D_j$$

$$R_T^i = \sum_{j=1}^{n} \sum_{k=1}^{j} T_k$$

$$NF_{kk}^i = \sum_{j=1}^{n} \sum_{k=1}^{j} KK_k$$

The temporal table $B_R^i$ will store the data according the criteria of user's selection. Subsequently, the users will be given an alternatives to choose either they want to search the keywords in the content of news title or not. The $keyword_k$ symbol is used to represent the kth keyword. This keyword is based on the user's input.

**The necessary of separating data temporal:** In establishing this model, we need to separate the history data news and temporal data. Besides, it aims is also to realize the temporal aspect for temporal engine.

Other factors are as follows:

- The growth of the records in both news and temporal data are differed from each other. The generated temporal schedule is directly proportional to the selection criteria by users. In other words, the growth of temporal schedule is faster than the growth of records for news
- The separation of both schedules will provide the following functions effectively
- Data retrieval in the temporal schedule using indexing method will be faster
- To improve a variety of facilities in temporal SQL
- To provide a highly reliable software. The separation of data will reduce the data retrieval time from the file

## CONCLUSIONS

As a conclusion, temporal model data can be involve in news archiving since every news has its own time point. Date is an important attribute in news archiving where attribute data temporal being used in the news archiving model is based on time of the transaction. This study introduces a new temporal representative for time status and year era. In time status, four different type of status introduced to handle the news depend of its date; date, week, month and year. Each type of status has different logic mode of temporal such as after, before,

during etc. These logics mode will make sure the data accessing always right and meet the users need. Beside that, this concept can be reduced the time of accessing and to maintain the consistent of database. To realize the accessing information based on the temporal concept, each selected temporal classification criteria has its own temporal table. This table has its own tuple which represent data of the table. In future works, others logic temporal can be apply to news archiving or others system development.

## REFERENCES

Allen, J.F., 1981. An interval-based representation of temporal knowledge. Proceedings of the 7th International Joint Conference on Artificial Intelligence, (IJCAI'81), Morgan Kaufmann Publishers Inc., San Francisco, CA, USA., pp: 221-226.

Allen, J.F., 1983. Maintaining knowledge about temporal intervals. ACM Commun., 26: 832-843.

Edelweiss, H.N., P.N. Hbler and P. Nogueria, 2000. Implementing a temporal database on top of a conventional database: Mapping of the data model and data definition management. Proceedings of the 15th Brazilian Symposium on Databases, October, 2000, Brazil, pp: 1-15.

Elmasri, R. and S.B. Navathe, 2000. Fundamentals of Database Systems. 3rd Edn., Addison-Wesley, New York.

Gao, D., C.S. Jensen, R.T. Snodgrass and M.D. Soo, 2005. Join operations in temporal databases. VLDB J., 14: 2-29.

Jensen, C.S. and R.T. Snodgrass, 1999. Temporal data management. Knowledge Data Eng., 11: 36-44.

Tanin, E., R. Zhang and L. Kulik, 2009. Spatio-temporal database research at the university of Melbourne. SIGMOD Rec., 38: 35-39.

Tsotras, V.J. and A. Kumar, 1996. Temporal database bibliography update. ACM SIGMOD Rec., 25: 41-51.

Unnikrishnan, K. and K.V. Pramod, 2009. On implementing temporal coalescing in temporal databases implemented on top of relational database systems. Proceedings of the International Conference on Advances in Computing Communication and Control, January 23-24, 2009, ACM, pp: 153-156.