

オブジェクト指向データベースにおける 時間軸データモデル

福田紀彦[†] 天笠俊之[†] 金森吉成[†] 増永良文^{††}

[†] 群馬大学工学部情報工学科 ^{††} 図書館情報大学

本論文では、事象の時間が interval とそのスカラー量である duration とから構成されるというモデルを提案した。このモデルでは、interval は、interval の start point を表す座標軸と end point を表す座標軸の 2 次元空間で表現される。この 2 次元空間に duration を表す座標軸を加え、3 次元空間として時間を表現する。さらに、interval を real interval と null interval とに分けて捉え、それらの複合した composite interval の概念も導入した。本論文でのモデルは、Little 等、Chandra 等の interval のみに基づくモデルの拡張となっている。

このモデルと Allen の与えた 2 項 temporal relation との関連についても議論した。本モデルを医療分野に応用する場合の問題についても検討した。

Temporal Axis Data Models for Object-Oriented Databases

Norihiko Fukuda[†] Toshiyuki Amagasa[†]
Yoshinari Kanamori[†] Yoshifumi Masunaga^{††}

[†] Department of Computer Science, Gunma University

^{††} University of Library and Information Science

In this paper, we have proposed the temporal data model based on intervals and it's scalars, durations, of which the time of fact is composed. The model is represented on three dimensional space. One axis of the space is start points and the other one is end points. In addition to, we adopted durations as third axis. Accordingly, the time is given in three dimensional space.

Furthermore, we have discriminated between real intervals and null ones, and introduced the conception of composite intervals mixed with them.

The model in the paper expands those ones proposed by Little et al. and Chandra et al. based on only intervals. We have discussed the relations between the model and binary temporal relations given by Allen, and moreover described the problems when the model is applied in the medical fields.

1 はじめに

マルチメディアデータベースにおいては、各メディアの時間管理が重要な課題となっている。この問題の必要性は、(1) 時間に関する質問処理をするための時間管理と、(2) 検索されたメディアの結果を時間的同期を取りながら提示するために必要な時間管理とに分けて考えることができる。前者については、関係データベースを対象として Snodgrass 等 [1, 2] によって、TQuel 質問言語の研究が代表的な例として挙げることができる。また、オブジェクト指向データベースにおける時間導入モデルが Wuu 等 [3]、Käfer 等 [4] によって報告されている。しかし、これらのすべての時間モデルは (2) の問題に対しての考慮が全くないので (1) と (2) を別個に考えざるを得ない。

ところが、最近、Allen[5] の提案した interval に基づく時間モデルによって上述の (1)、(2) を統一的に扱える可能性の道が開けてきた。Little 等 [6] によるマルチメディアデータの時間管理はこの方向を示していると言える。Chandra 等 [7] も interval に基づく時間モデルの有効性を (1) の視点から発表している。

本研究においてもマルチメディアデータベースをオブジェクト指向データベースで実現するという前提のもとで時間管理の問題を議論する。すなわち、Little 等 [6]、Chandra 等 [7] のモデルを拡張した時間モデルを提案する。

この論文では、実世界の時間構造をより忠実にモデル化するために、interval のスカラー量である duration を加え、また、interval を real interval と null interval とに分けて捉え、それらの複合した composite interval の概念も導入した。さらにこのモデルを時間構造の複雑な医療分野に適用する場合についても議論する。

2 時間軸データモデル

2.1 区間の定義と表現

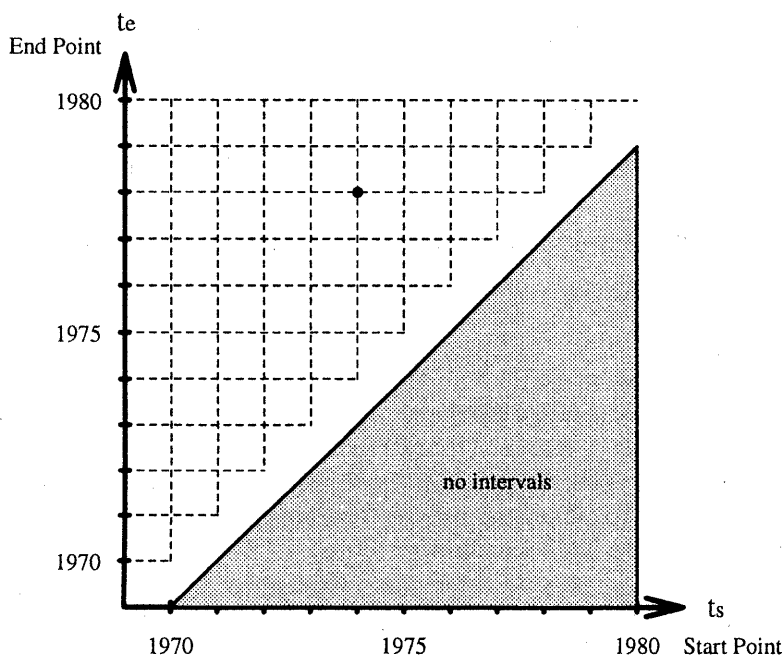


図 1: 時間軸データモデルによる interval の表現

時間軸データモデルでは、実世界で起こる事象 (event) の時間を区間 (interval) として考える。区間は、その事象が始まった時点 (始点または、start point) と終わった時点 (終点または、end point) の情報を持っている。この論文では、ある事象の区間を $I = \langle t_s, t_e \rangle$ で表し、 t_s が始点、 t_e が終点をそれぞれ示し、 $t_s \leq t_e$ の関係にあることにする。このように事象の時間を区間に基づいて捉えることは、Allen[5] が最初に提案した。本論文での区間の定義方法は、Little 等 [6] や Chandra 等 [7] の区間の定義と同じである。

図 1 は、事象の区間を表現する座標系を示した図である。横軸は始点を表す座標軸であり、縦軸は終点を表す座標軸である。この図の場合、例として各軸の単位に西暦を採用し、1970 年から 1980 年までを表している。また、この座標系の右斜め下の部分は、 $t_s > t_e$ の関係から、区間が存在しないことを示している。つまり、 $I_1 = \langle 1978, 1974 \rangle$ というような区間は存在しない。この座標系を用いて、ある事象の区間を表現すると、この座標系の点として表現される。図 1 の座標系の点は、ある事象の区間 $I_2 = \langle 1974, 1978 \rangle$ を表現していることを示す。

2.2 期間の定義と表現

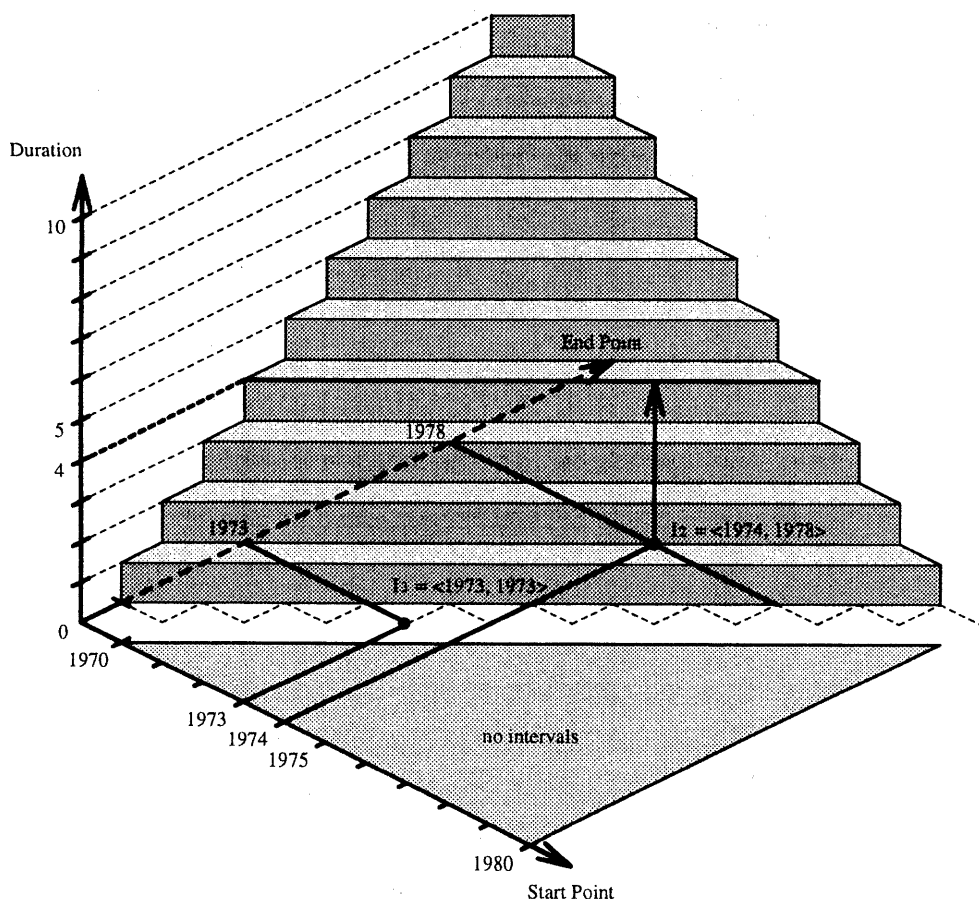


図 2: 時間軸データモデルによる duration の表現

ある事象の区間は、スカラー値として期間 (duration) を持っている。すなわち、期間 $D = |I| =$

$|t_e - t_s|$ である。時間軸データモデルでは、期間を図1の座標系を拡張させて表現することができる。図2は、図1に期間の軸を付け加えた座標系を示す。

図2の期間の軸 (Duration) は、1年間から10年間までを表している。図2は、図1の区間 I_2 の期間が $1978 - 1974 = 4$ 年間であることを示している。

また、図2は、同じ期間を持つ区間の点 (すなわち、 $\langle 1970, 1974 \rangle$, $\langle 1971, 1975 \rangle$, $\langle 1972, 1976 \rangle$, ...) が期間 $D = 4$ 上に1列に並ぶことを視覚的に示している。それに、期間が零の区間 (すなわち、 $\langle 1970, 1970 \rangle$, $\langle 1971, 1971 \rangle$, $\langle 1972, 1972 \rangle$, ...)、についても同様にこの座標系で扱うことができることを示している (図2では、 $I_3 = \langle 1973, 1973 \rangle$)。

図1、図2のように区間を表現して、時間を統一的に記述するモデルを本論文では時間軸データモデルと呼ぶことにする。

2.3 時間軸データモデルにおける質問処理

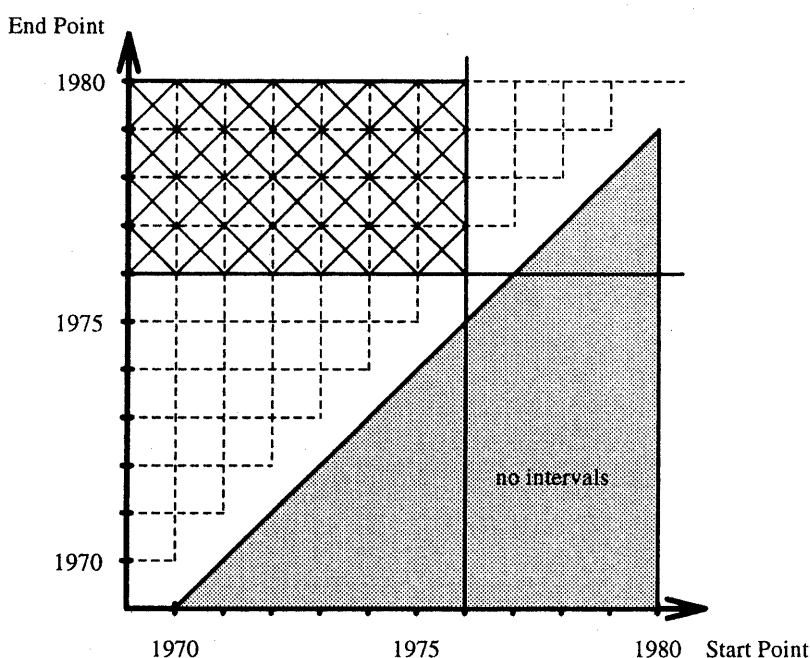


図 3: 時間軸データモデルにおける質問処理

ここでは、このモデルの有効性を説明するために、例として代表的な質問処理を4つ挙げる。

1. 区間の始点による質問の検索の場合。例：1974年に始まった事象は？
2. 区間の終点による質問の検索の場合。例：1978年に終わった事象は？
3. ある時点に存在している区間による質問の検索の場合。例：1976年に存在する事象は？
4. 期間による質問の検索の場合。例：4年間続いた事象は？

この時、各々の条件に対して、時間軸データモデルに基づく質問処理の方法を次に示す。

1. の条件に関しては、図1の横軸 (区間の始点を表す座標軸) の1974年の点から横軸と垂直をなす線に存在するすべての点が条件に該当している事象である。また、2. の条件は、縦軸に関して1. と同様なことを行なうことによって該当する事象を検索することができる。

3. の条件に関しては、条件を次のように解釈する。

- 1976 年以前に始まった事象は?
- 1976 年以降に終わった事象は?

以上の 2 つの条件を同時に満たす座標系の点を検索することによって、3. の条件を満たす事象をすべて検索することができる。図 3 の左斜め上の部分 (ハッチング部分) は、3. の条件を満たす事象の区間の点が存在している部分であることを示している。

4. の条件に関しては、前節で述べた通り、4 年間続いたすべての事象は、図 2 の期間の軸が 4 年間である区間の点の列から検索することができる。この質問処理は、 $I = \langle t_s, t_e \rangle$ において、 $|t_e - t_s| = 4$ を満たす t_s, t_e の組合せをすべて探すことに相当するが、このモデルでは、期間軸で期間 $D = 4$ の値を見るだけで該当する区間を直ちに求めることができる。これはこのモデルの有効性を示している。

従って、時間に関する質問処理は、上記のように検索を実行する基本的な演算を定義し、これらの演算に基づく時間の質問処理言語を与える問題となる。これについては、別の機会で述べたい。

3 関連する研究との関係

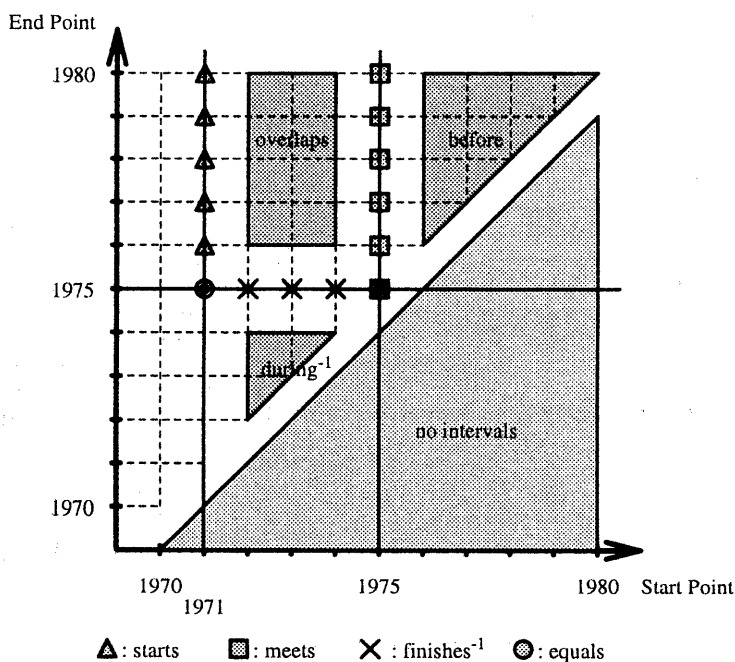


図 4: 時間軸データモデルにおける 2 項 temporal relation の表現

Allen[5] では、事象の時間を区間で表現し、区間どうしの 2 項関係を 13 個の temporal relation で与えている。

図 4 は、例として、区間 $I = \langle 1971, 1975 \rangle$ に対する 7 個の 2 項 temporal relation (equals 以外の 12 個は、互いに inverse 関係にある 6 個からなる。結局、7 個の temporal relation で表現できることを与えた) を表現した図である。ここでの 2 項 temporal relation の *before*, *meets*, *overlaps*, *during-1*, *starts*, *finishes-1*, *equals* は、Little 等 [6] の記法に従っている。

時間軸データモデルでは、区間が本来持っている性質(始点、終点、期間)をそのまま座標系によって表現しているので、7個の2項 temporal relation を図4のように表現できる。これによって、2つの区間の座標系での位置関係によって、temporal relation を識別することが簡単にできる。また、この位置関係によって、temporal relation による制約も与えることができる。

一方、Little 等 [6] のモデルでは、区間に関する6個のパラメータで2項 temporal relation を定義している。時間軸データモデルでは、2つの区間の始点と終点の4個のパラメータで2項 temporal relation を与えていることになる。この点からも本モデルが有効であることがわかる。

また、Little 等 [6] は、マルチメディア間の同期をとるためのモデルに temporal relation を使って、マルチメディアデータ提示のスケジューリングを実現した。そこには、マルチメディアデータベースにおける時間に関する質問処理の視点が無い。

Allen[5]、Chandra 等 [7] では、期間(duration)を全く取り上げていない。Little 等 [6] では、音楽演奏を開始する時点の計算や、temporal relation による制約を与えるために期間を用いている。しかし、本論文での特徴は期間を導入したことにある。それは、前章で述べたように、質問処理するためのキーとして期間を直接用いることが可能であり、この点が他の論文と異なる特徴である。

期間に関する2項 temporal relation としては、 $>$ 、 $=$ 、 $<$ の3種類がある。従って、期間を導入することによって計10種類の2項 temporal relation に基づいて演算系を定義し、質問言語を与えなければならない。この問題の形式化については、別稿で述べたい。

4 医療症例データベース

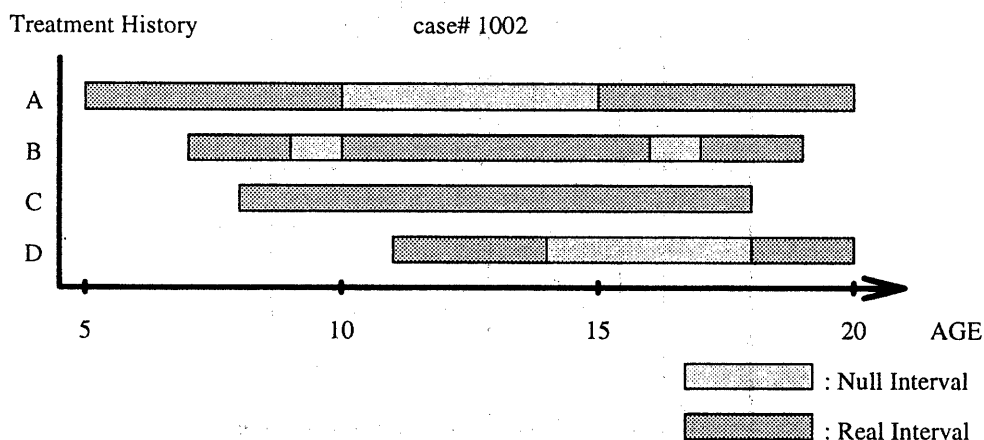


図 5: 症例の治療の履歴の例

この章では、上記のモデルを医療症例データベースに適用する場合について検討する。

最初に医療症例データベースで取り扱う治療の履歴について述べる。図5は、症例の治療の履歴の例である。縦軸は、治療の種類(A,B,C,D)を表し、横軸は、症例の年齢(5歳から20歳まで)を表している。

このような治療の履歴に対して、次のような質問が考えられる。

- 治療の時点の質問の例：10歳の時、治療Bを行なった症例は？
- 治療の期間の質問の例：5年間、治療Aを行なった症例は？

以上のように、医療症例データベースでは、治療を行なった時点をキーとした検索の他に、治療期間をキーとした検索が重要である。これらは、前述の時間軸データモデルの枠内ですべて処理できる。

4.1 Null Interval の導入

症例データベースでは、実際に治療を行なった区間の他に、治療を休止した区間も重要な意味を持つ場合がある。例えば、治療 A を行ない、5 年間その治療を休止した症例は？ という検索である。つまり、ある治療を行なって、その治療後の様子を伺うことである。このような要求に対して、我々は、Real Interval と Null Interval を導入する。Real Interval とは、実際に治療を行なった区間であり、Null Interval とは、治療を休止した区間である。この様子を図 5 に示す。Real Interval と Null Interval の定義は以下で与える。

4.2 Composite Interval の導入

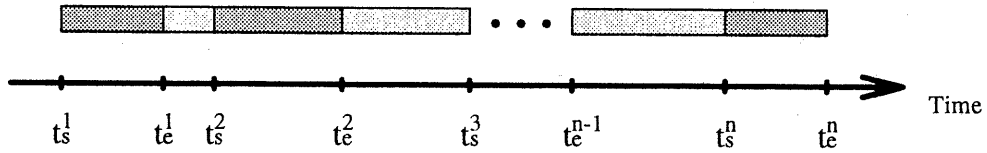


図 6: Composite Interval の例

Null Interval を導入すると共に、Composite Interval も導入する必要がある。Composite Interval とは、Real Interval と Null Interval を組み合わせた区間である。図 6 は、Composite Interval の例である。

ここで、Composite Interval を定義するために、Real Interval と Null Interval を定義する。Real Interval I は、

$$I_1 = \langle t_s^1, t_e^1 \rangle, I_2 = \langle t_s^2, t_e^2 \rangle, \dots, I_n = \langle t_s^n, t_e^n \rangle.$$

である。そして、Null Interval NI は、Real Interval と Real Interval の間に必ず存在することから、次のように定義できる。

$$NI_1 = \langle t_e^1, t_s^2 \rangle, NI_2 = \langle t_e^2, t_s^3 \rangle, \dots, NI_{n-1} = \langle t_e^{n-1}, t_s^n \rangle.$$

従って、Composite Interval CI は、((((I_1 meets NI_1) meets I_2) meets NI_2) $\dots$ meets NI_{n-1}) meets I_n) で定義される区間である。また、Composite Interval の期間 D_{CI} は、次の式で与えられる。

$$D_{CI} = |CI| = |I_1| + |NI_1| + |I_2| + |NI_2| + \dots + |NI_{n-1}| + |I_n| = \sum_{i=1}^n |I_i| + \sum_{j=1}^{n-1} |NI_j|.$$

図 5 の治療 A の Composite Interval は、5 歳から 10 歳まで治療を行なって、10 歳から 15 歳まで休み、15 歳から 20 歳まで再び治療を行なったことを示している。Composite Interval の導入によって、実際に行なった治療 A の合計の期間の計算 (治療のすべての Real Interval の合計期間) や、治療 A の合計の期間の計算 (治療のすべての Null Interval と Real Interval の合計期間) が可能である。

一般に、ある治療は Composite Interval で表される。図 5 のように 1 患者が複数の治療 (A ~ D) を同時に受けている場合には、その履歴は Composite Interval の集合 {Composite Interval} で表

される。複数の治療の履歴について同時に質問する場合には、各々の治療が互いに独立しているならば、個々の治療毎に質問処理をした後、結果を and、or 等で演算すれば良い。しかし、治療間での時間的制約、例えば、ある治療の Null Interval の時に、別の治療ができる、すなわち、その治療の Real Interval となるなど、治療間の 2 項関係、さらには、多項関係が存在する場合には、質問処理が簡単ではない。これらの多項関係は、Little 等 [6] の定義したものに比べはるかに複雑である。{Composite Interval} をどのように統一的に処理すべきかは今後の課題である。

5 おわりに

本論文では、事象の時間を区間とそのスカラー値の期間からの構成として捉え、その区間を時間軸データモデルを用いて表現する方法を提案し、その有効性を述べた。

また、医療症例データベースにおける時間の要求を分析し、Null Interval と Composite Interval の必要性について述べた。また、症例の履歴全体を表す Set of Composite Intervals の重要性についても述べた。

今後の課題として、時間軸データモデルを医療症例データベースに適用し、Null Interval、Composite Interval と Set of Composite Intervals を取り扱う代数系 (Medical Record Algebra) を考案しなければならない。また、オブジェクト指向データベースにこのモデルを導入して実現する方法も検討しなければならない。

参考文献

- [1] R. Snodgrass and I. Ahn. "Temporal Databases". *IEEE COMPUTER*, 19(9):35-42, 1986.
- [2] R. Snodgrass. "The Temporal Query Language TQuel". *ACM Trans. on Databases Systems*, 12(2):247-298, June 1987.
- [3] G. Wu and U. Dayal. "A Uniform Model for temporal Object-Oriented databases". In *Proc., Eighth IEEE Data Engineering Conf.*, February 1992.
- [4] W. Käfer and H. Schöning. "Realizing a Temporal Complex-Object Data Model". In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 266-275, San Diego, California, June 1992.
- [5] J.F. Allen. "Maintaining Knowledge about Temporal Intervals". *Commun. ACM*, 26:832-843, Nov. 1983.
- [6] T.D.C. Little and A. Ghafoor. "Interval-Based Conceptual Models for Time-Dependent Multimedia Data". *IEEE Trans. on Knowledge and Data Engineering*, 5(4):551-563, Aug. 1993.
- [7] R. Chandra, A. Segev, and M. Stonebraker. "Implementing Calendars and Temporal Rules in Next Generation Databases". In *Proc., Tenth International Conference on Data Engineering*, pages 264-273, Feb 1994.