Scalable Temporal Reasoning

Steffen Staab

Institute for Applied Computer Science and
Formal Description Methods (AIFB)
Karlsruhe University
D-76128 Karlsruhe, Germany
http://www.aifb.uni-karlsruhe.de/~sst

Udo Hahn

LIF Text Understanding Lab
Computational Linguistics Division
Freiburg University
D-79085 Freiburg, Germany
http://www.coling.uni-freiburg.de

Abstract

We introduce two mechanisms for scaling computations in the framework of temporal reasoning. The first one addresses abstraction at the methodological level. Operators are defined that engender flexible switching between different granularities of temporal representation structures. The second one accounts for abstractions at the interface level of a temporal reasoning engine. Various generalizations of temporal relations are introduced that approximate more fine-grained representations by abstracting away irrelevant details.

1 Introduction

Most of the challenging AI problems exhibit a computational complexity that is inherently intractable. In a seminal paper, Jerry Hobbs [1985] proposed a parsimonious reasoning mode based on different granularities of representation structures. He suggests to have a hierarchy of theories available — shallow ones that come with cheap computations, more sophisticated ones that require increasing costs. Hence, a trade-off between expressiveness and computational resources is managed by shifting between different granularity levels. The intrinsic mechanism needed in such a model are switching operators that move from one representational layer to the other.

The necessity to account for different granularities in temporal reasoning has been widely acknowledged in diverse applications such as natural language understanding [Nakhimovsky, 1988] or planning [Sathi *et al.*, 1985]. Nevertheless, granularity has often only been considered as an *addendum* to temporal reasoning schemes [Bettini *et al.*, 1997]. In contradistinction, we treat granularity as a fundamental notion, one that *underlies* the relation between the most common temporal reasoning formalisms.

We here build on the notion of *Generalized Temporal Networks (GTNs)* as introduced by Staab [1998b]. GTNs (cf. Section 3 for an overview) constitute a highly expressive constraint formalism. They subsume several widely known temporal reasoning approaches [Allen, 1983; Vilain *et al.*, 1989; Kautz and Ladkin, 1991; Dechter *et al.*, 1991; Meiri, 1996; Badaloni and Berati, 1996] and integrate them in a common framework. However, as Staab [1998b] concedes, reasoning at the most specific level of GTNs is not a panacea. What is lacking though is a mechanism that might mediate between the computation of complex temporal relations still requiring resource-intensive constraint solvers at the high end of intractability, and the computation of simpler temporal relations for which tractable constraint systems might suffice. So, our goal is to extend the GTN framework such that a cascade of constraint formalisms organized at different axes of expressiveness/computational complexity can be defined, together with operators for navigating this cascade to choose the appropriate level of constraint evaluation (cf. Section 4).

This scaling between different levels is also motivated by our natural language text understanding application [Staab and Hahn, 1997] that demands for reasoning in different temporal granularities. The potential for abstraction is not limited to the internal representation and reasoning mode, but it is also useful for accessing the system from the outside via its interface (cf. Section 5). So, the ultimate goal is to support reasoning and access by uniform abstraction mechanisms.

The latter point deserves particular attention when basic research results from the area of temporal reasoning are to move to public use, or even commercial applications. The availability of flexible, easy-to-use, explanatory adequate interfaces has already proven to be a major asset in comparably complex domains such as terminological reasoning (cf. McGuinness and Patel-Schneider [1998]). In the framework of temporal reasoning then, finding the right level of abstraction for querying is considered a crucial factor for enhanced usability of temporal reasoners. Though very expressive models [Meiri, 1996; Staab, 1998b] might be required by the actual application, they usually do not lend themselves easily to communication with the human user or with high-level system modules due to an overly fine grain size. In order to propagate flexibility to the interface level, we abstract away irrelevant details from the low-level representations at the price of only approximate, though sufficiently explicit solutions. So, providing the right level of abstraction is also a matter of adequacy of temporal reasoning, in general, and increased applicability of temporal reasoners, in particular, rather than just a detrimental loss of expressiveness at the gain of lowering computation costs.

2 A Temporal Reasoning Scenario

In order to facilitate the understanding of the concepts we introduce, we here give an everyday scheduling problem that will be used throughout this paper for illustration purposes:

(1) James is a shuttle driver for a major hotel in New York. Today's schedule posts Mr. Roget and Mr. Meyer from Paris, Mrs. Meyer from Philadelphia, and Mr. George from Sidney for transportation. The hotel's clerk told him that Mr. Roget and Mr. Meyer have tickets for different flights from Paris to NY. Mr. Roget is scheduled to arrive with the Concorde in NY at 3:00pm local time, and Mr. Meyer should arrive in NY two hours later. However, they currently try to arrange for sharing a flight which would arrive in NY at 6:00pm local time. When Mr. Meyer arrives in NY he will immediately call his wife, Mrs. Meyer, who will get the next train to NY. Hence, she will be in NY less than 4 hours after her husband has arrived. Furthermore, Mr. George's flight leaves Sidney at 12:00pm local NY time, and he has got a very long flight.
Problem: In which order must James service the guests?

3 The GTN Model

We here give a brief description of the GTN model, summarizing those aspects relevant for the mechanisms of scaling — the focus of this paper. A detailed explanation of GTNs and a comparison to related models is given in [Staab, 1998b].[1]

The GTN model builds on time point variables and relations that describe restrictions between two or more variables at a time. A given network of relations is tightened by propagation, and thereby, conclusions are computed until *weakly generalized path consistency* (WGPC) is reached. Relations consist of disjunctions of conjoined primitive constraints. Formally, a relation R_k comes in the following form:

(2) $R_k := \bigvee_{l=1\ldots L_k}(\bigwedge_{(v_i,v_j)\in E_k} p_{i,j,k,l})$

Hereby, $p_{i,j,k,l} = (v_i, q_{i,j,k,l}, v_j)$ are constraints between the temporal variables $v_i,\ v_j \in \mathcal{V}$, and $q_{i,j,k,l}$ are intervals such that $v_j - v_i \in q_{i,j,k,l}$. L_k denotes the number of disjunctions of the relation R_k, and E_k denotes all pairs of time point variables between which a constraint of relation R_k may hold. The set of all relations $\{R_k | k = 1 \ldots M\}$ is denoted by $\mathcal{R}$; correspondingly, its *topology* $\mathcal{E}$ is defined by $\{E_k | k = 1 \ldots M\}$.

Thus, example (1) can be modeled in the following way:

(3) a. 12:00pm: t_0
b. End of Mr. Roget's flight: t_1
c. End of Mr. Meyer's flight: t_2
d. Arrival of Mrs. Meyer in NY: t_3
e. Beginning of Mr. George's flight: t_4
f. End of Mr. George's flight: t_5
g. If Mr. Roget arrives at 3:00pm, then Mr. Meyer arrives two hours later; otherwise, they arrive together at 6:00pm: $((t_0, [3,3], t_1) \wedge (t_1, [2,2], t_2)) \vee ((t_0, [6,6], t_1) \wedge (t_1, [0,0], t_2))$
h. Mrs. Meyer arrives less than 4 hours after her husband: $(t_2, (0,4), t_3)$
i. Mr. George has a very long flight: $(t_4, [\text{"very long"}, +\infty), t_5)$
j. Mr. George's flight starts at 12:00pm: $(t_0, [0,0], t_4)$

[1] For proofs of the formal claims we make, cf. [Staab, 1998a].

In order to compose relations, propagate resulting restrictions, and determine when WGPC is reached, Staab [1998b] provides *interval mappings*, *projection* and *subsumption*. Relying on these (and some other) auxiliary means, he states a constraint propagation algorithm that enforces WGPC on GTNs. This algorithm can be applied unaltered to the scaling problems we discuss here. We will now formally introduce those auxiliary means, because the notion of scalability we develop makes direct use of them.

First, we define the notion of *projection*:

Definition 1 *The projection $\pi : 2^{\mathcal{R}} \times \mathcal{E} \mapsto \mathcal{R}$ is a binary function ($\pi_x(y) := \pi(y, x)$). $\pi_{E_g}(R_1 \wedge \ldots \wedge R_n)$ selects all constraints in $R_1 \wedge \ldots \wedge R_n$ which affect the edges in E_g. It is defined at three levels — that of simple conjoined constraints, of disjunctions of conjoined constraints, and of conjoined relations. Its input is described by referring to sets (of tuples) K_1, K_2, and K_3, respectively ($P_{k,l} := \bigwedge_{(v_i,v_j)\in E_k} p_{i,j,k,l}$):*

1. $\pi_{E_g}(\bigwedge_{(i,j,k,l)\in K_1} p_{i,j,k,l}) := \bigwedge_{(i,j,k,l)\in K_1,(v_i,v_j)\in E_g} p_{i,j,k,l}$
2. $\pi_{E_g}(\bigvee_{(k,l)\in K_2} P_{k,l}) := \bigvee_{(k,l)\in K_2} \pi_{E_g}(P_{k,l})$
3. $\pi_{E_g}(\bigwedge_{k\in K_3} R_k) := \pi_{E_g}(\bigwedge_{k\in K_3}(\bigvee_{l\in[1\ldots L_k]} P_{k,l})) = \pi_{E_g}(\bigvee_{(x_1,\ldots,x_{|K_3|}),x_i\in[1\ldots L_i]}(\bigwedge_{k\in K_3,q=x_k} P_{k,q}))$.

An example for π is given in (4) which incorporates the information given in (3g):

(4) $\pi_{\{(t_1,t_2)\}}(\ ((t_0,[3,3],t_1) \wedge (t_1,[2,2],t_2)) \vee ((t_0,[6,6],t_1) \wedge (t_1,[0,0],t_2))\ ) = (t_1,[2,2],t_2) \vee (t_1,[0,0],t_2)$

Second, the GTN model allows for different interval structures from which the intervals $q_{i,j,k,l}$ may be taken.

Definition 2 *An interval structure $\mathcal{I}$ is a quadruple $(I, D, \circ, \cap)$. I is a set of (semi-)intervals on a domain D. I is closed under the composition and intersection operators, $\circ$ and $\cap$, respectively.*

Most common are the structures $\mathcal{I}_{\mathbb{Q}} = (I_{\mathbb{Q}}, \mathbb{Q}, +, \cap)$ and $\mathcal{I}_O = (\{(-\infty,0), (-\infty,0], (-\infty,+\infty), [0,+\infty), (0,+\infty), [0,0]\}, \mathbb{Q}, +, \cap)$, where $\mathbb{Q}$ denotes the rational numbers, $I_{\mathbb{Q}}$ all intervals on the rationals, $+$ interval addition, and $\cap$ set intersection.

In order to compose constraints from different interval structures, *interval mappings* are established that communicate restrictions.

Definition 3 *Interval mappings are functions $\mu_{\mathcal{I}_r,\mathcal{I}_s} : I_r \mapsto I_s$ from one interval structure, $\mathcal{I}_r$, to another one, $\mathcal{I}_s$, such that the following properties are fulfilled: $\forall i,j,k,l : (v_i, q_{i,j,k,l}, v_j) \Rightarrow (v_i, \mu_{\mathcal{I}_r,\mathcal{I}_s}(\mu_{\mathcal{I}_s,\mathcal{I}_r}(q_{i,j,k,l})), v_j)$. If $D_r = D_s$, it is also required that $\forall i,j,k,l : (v_i, q_{i,j,k,l}, v_j) \Rightarrow (v_i, \mu_{\mathcal{I}_r,\mathcal{I}_s}(q_{i,j,k,l}), v_j)$.*

For instance, the resulting quantitative constraints in (4) are mapped onto a common ordinal one by $\mu_{\mathcal{I}_{\mathbb{Q}},\mathcal{I}_O}$:

(5) $(t_1, \mu_{\mathcal{I}_{\mathbb{Q}},\mathcal{I}_O}([2,2]), t_2) \vee (t_1, \mu_{\mathcal{I}_{\mathbb{Q}},\mathcal{I}_O}([0,0]), t_2) = (t_1, (0,+\infty), t_2) \vee (t_1, [0,0], t_2) = (t_1, [0,+\infty), t_2)$

Third, *subsumption* is used in order to compare models of GTN relations.

Definition 4 *A relation $R_1 := \{P_{1,l} | l = 1 \ldots L_1\}$ subsumes a relation $R_2 := \{P_{2,l} | l = 1 \ldots L_2\}$ ($R_1 \trianglerighteq R_2, R_2 \trianglelefteq R_1$) iff $\bigcup_{l=1\ldots L_2} P'_{2,l} \setminus \bigcup_{l=1\ldots L_1} P'_{1,l} = \emptyset$, where*

$$P'_{k,l} := \bigtimes_{i,j=1\ldots N, i<j} \left(\begin{array}{ll} q_{i,j,k,l}, & \text{iff } (v_i, v_j) \in R_k \\ D, & \text{otherwise} \end{array} \right).$$

Lemma 5 *If $R_1 \trianglerighteq R_2$, then every model for the relation R_2 that assigns values to time point variables in $\mathcal{V}$ is also a model for the relation R_1.*

For instance, the result in (4) subsumes the input given to $\pi_{\{(t_1,t_2)\}}$, the result in (5) subsumes the result of (4), and due to the transitivity of subsumption the relation in (5) subsumes the one in (3g).

Finally, before proceeding with the topic proper of this paper, we give the definition of an unambiguous GTN, a UGTN:

Definition 6 *A UGTN is a GTN, where $\forall k = 1...M : L_k = 1$.*

4 Abstraction at the Reasoning Level – Switching Between Levels of Granularity

The GTN model is the foundation for switching between different levels of reasoning. GTNs based on intervals from the reals, $\mathcal{I}_{\mathbb{Q}}$, bring about a very fine-grained level of temporal reasoning as a general frame of reference. Switching to coarser models is possible relative to at least two dimensions: First, the dimension of interval structures of different granularities permits such changes, e.g., abstractions from quantitative constraints to qualitative, i.e., ordinal ones, like $\mathcal{I}_O$, or granularity changes between days, weeks and months as they have been described by Bettini *et al.* [1997]. Second, one may consider disjunctions of conjoined constraints as already too sophisticated a level of representation. From such a level, it is, nevertheless, possible to move into a sparser theory, by using abstraction on the propositional level as investigated by Giunchiglia and Walsh [1992].

Figure 1 is an excerpt of a network heterarchy of temporal reasoning schemes (with arrows pointing from less towards more expressive formalisms). $GTN(\mathcal{I}_{\mathbb{Q}})$ and $GTN(\mathcal{I}_O)$ denote GTNs based only on the interval structures $\mathcal{I}_{\mathbb{Q}}$ and $\mathcal{I}_O$, respectively. STP and TCSP stand for non-disjunctive and disjunctive quantitative constraint systems, respectively, as described by Dechter *et al.* [1991].[2] The term *integration* stands for the integration of TCSPs with Allen's model [Meiri, 1996]. TCSP-LPC (cf. Schwalb and Dechter [1997]) is not really a representation schema on its own. Viewed from a representational perspective, it is equivalent to TCSPs, but it propagates only a limited number of disjunctions in each step such that propagation, as a whole, remains polynomial in the number of relations.

This heterarchy mirrors the well-known trade-off between expressiveness and efficiency. Determining consistency is NP-hard in all formalisms, except for the point algebra and for STP networks (cf. Vilain *et al.* [1989], Dechter *et al.* [1991]). However, even approximating constraint propagation algorithms can be very expensive when large ranges are embodied in the network.

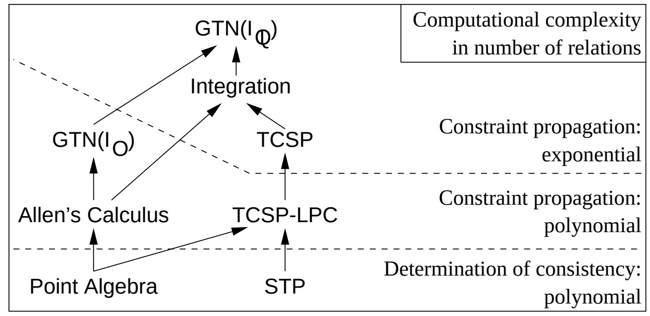

Figure 1: Expressiveness of Reasoning Schemes

We attempt to deal with this complexity bottleneck by providing smooth shifts among different levels of expressiveness. Following Hobbs's strategy that "idealization allows simplifications into tractable local theories", our proposal approximates given information by "simpler" one. These shifts are performed by two families of operators already introduced in the previous section: The first one, π_{E_g}, takes interdependent constraints as input and disregards their relationships, e.g., as with the disjunction in (4). The second one, $\mu_{\mathcal{I}_r,\mathcal{I}_s}$, allows switching between different interval granularities, as, e.g., illustrated by collapsing information in (5).

As can be read off from the diagram, both idealizations abstract from networks composed of detailed representations, with expensive constraint processing, to coarser representations, which allow for more efficient reasoning. Hence, expressiveness is traded off against efficiency. Disregarding structural interdependencies, e.g., allows the projection of $GTN(\mathcal{I}_O)$ information into an efficiently solvable point algebra. A coarser level of quantities, and thus a small overall range, is directly reflected by a tighter worst-case bound for constraint propagation (cf. Theorem 13 in [Staab, 1998b]).

The soundness of both abstraction operators is ensured by Definition 3 for operators μ and by Lemma 7 for operators π:

Lemma 7 *For all $E \in \mathcal{E}$ and $\{R_1, \ldots, R_m\} \in 2^{\mathcal{R}}$: The constraints given by $R_1 \wedge \ldots \wedge R_m$ entail the constraints given by $\pi_E(\{R_1 \wedge \ldots \wedge R_m\})$.*

Let us now illustrate the use of these abstraction mechanisms by considering the temporal reasoning problem given in (1). In order to retrieve qualitative ordering information, such as determining arrival orderings, it is often desirable to move down the heterarchy from GTN($\mathcal{I}_{\mathbb{Q}}$) to a point algebra. This is done for two relevant pieces of knowledge, *viz.* for (3g) by the combination of (4) with (5), and for (3h) by (6).

(6) $\mu_{\mathcal{I}_{\mathbb{Q}},\mathcal{I}_O}((t_2, (0,4), t_3)) = (t_2, (0,+\infty), t_3) \Leftrightarrow t_2 < t_3$

From (5) and (6) we may easily read off that Mr. Roget, Mr. Meyer and Mrs. Meyer arrive just in this order, the men may even arrive simultaneously.

[2] In the formal framework of GTNs, for TCSPs we require $\forall k : |E_k| = 1$, and for STPs we assume $\forall k : |E_k| = 1 \wedge L_k = 1$.

Given that we have neglected knowledge about durations, we do not know how Mr. George's arrival is ordered with respect to the other ones. What is needed is reasoning at the level of TCSPs — on the one hand:

(7) $\pi_{\{(t_0,t_1)\}}(\; ((t_0,[3,3],t_1) \wedge (t_1,[2,2],t_2)) \vee$
$((t_0,[6,6],t_1) \wedge (t_1,[0,0],t_2)) \;) =$
$(t_0,[3,3],t_1) \vee (t_0,[6,6],t_1)$

From (4) and (7) we derive:

(8) $(t_0,[5,5],t_2) \vee (t_0,[3,3],t_2) \vee (t_0,[8,8],t_2) \vee$
$(t_0,[6,6],t_2)$

From (8) and (3h) we conclude:

(9) $(t_0,(5,9),t_3) \vee (t_0,(3,7),t_3) \vee (t_0,(8,12),t_3) \vee$
$(t_0,(6,10),t_3) \;\; = \;\; (t_0,(3,12),t_3)$

On the other hand, one needs to account for background knowledge about the duration of flights. Assuming an interval structure (like the ones in Clementini *et al.* [1997]) referring to flights of "short", "medium", "long", and "very long" time extension, a common grounding between "very long" and hour units may be that "very long flights" take at least 15 hours. With this information and with (9), one may conclude, finally, that Mr. George will arrive last.

Though for most temporal reasoning mechanisms the two families of abstraction operators, π and μ, play the major role, one may think of alternative operators, too. For instance, Schwalb and Dechter [1997] encountered the TCSP fragmentation problem by restricting propagation to (almost) convex constraints. An operator τ that abstracts from general non-convex relations into a limited number of convex disjunctions may render constraint propagation more efficient. However, the disadvantage remains that the resulting network does not have a similarly relevant status as, say, an interval algebra, for which path consistency has been determined.

With these abstraction operators, problems stated at one level are mostly lifted onto a coarser level of reasoning. However, going the other way round may also prove fruitful. Reasoning at the coarser level is cheap and all the conclusions that are inferred at the coarser level also hold at a finer grain size. Very often the switch backwards is given by the identity operation[3], e.g., between GTNs based on $\mathcal{I}_O$ and $\mathcal{I}_{\mathbb{Q}}$.

5 Abstraction at the Interface Level — Generalizing by Approximating Relations

Increased expressiveness and the application of powerful abstraction mechanisms that mediate between different precision levels of reasoning may actually aggravate the application of a temporal reasoning system. While thirteen qualitative interval relations in Allen's calculus or disjunctions of interval constraints in TCSPs may already pose non-trivial problems for a human to deal with, GTN relations have an even more complicated structure. Thus, GTN relations are often too unwieldy to be used in a temporal query language or by a module of a larger system, though an application may actually require their use. For instance, a text understanding and generation system dealing with the scheduling problem as given in (1) may need to account for complex propositions such as (3g). This means that high-level conceptual representation structures, e.g., *"a very long flight"* or *"X arrives after Y"*, that are typically employed by such a system must be translated to low-level GTN expressions when in-depth temporal reasoning is required.

[3] One notable exception arises when granularity levels are not directly comparable, e.g., month *vs.* week (cf. Bettini *et al.* [1997]).

To bridge the conceptual distance, we here introduce an interface level that abstracts from unnecessary details and, hence, generalizes to the *relevant* distinctions that need to be made. In doing so, we provide definitions of abstracting relations that are used to move from the interface level down to the reasoning level – and in the reverse direction. Switching from the interface to the reasoning level, e.g., when posing a query to a temporal reasoning system, one simply has to expand the definition of the abstracting relation. Table 1 shows some examples of such "macro" definitions. Switching back, i.e., outputting an abstracted relation to the interface level, e.g., as an answer to a query posed by a "naive" user, requires reasonable criteria for the selection of those interface relations that are best suited to abstract from a given low-level relation. We here define two notions of "best approximations":

Definition 8 *Let a set of abstracting relations be given by* $R_1^a, \ldots, R_n^a$.

A relation R_i^a *is a smallest upper approximation of a relation* R *with regard to* $R_1^a, \ldots, R_n^a$*, iff* $R_i^a \trianglerighteq R$ *and there is no* $R_j^a, i \neq j$ *such that* $R_i^a \trianglerighteq R_j^a \trianglerighteq R$*.*

A relation R_i^a *is a greatest lower approximation of a relation* R *with regard to* $R_1^a, \ldots, R_n^a$*, iff* $R_i^a \trianglelefteq R$ *and there is no* $R_j^a, i \neq j$ *such that* $R_i^a \trianglelefteq R_j^a \trianglelefteq R$*.*

This definition may yield several smallest upper and greatest lower approximations. A unique upper approximation is given by the conjunction of the best upper bounds, while a unique lower approximation is given by the disjunction of the best lower bounds.

We do not present an algorithm here for computing the

A	Interval A **is between** interval B and interval C $((A_b,(-\infty,0),B_e) \wedge (A_e,(0,+\infty),C_b)) \vee$ $((A_b,(-\infty,0),C_e) \wedge (A_e,(0,+\infty),B_b))$
B	Interval A **is at least** n **units disjoint** from B $(A_e,[n,+\infty),B_b) \vee (A_b,(-\infty,-n],B_e)$
C	**If** time point a **before** time point b **then** time point c **before** time point d $((a,(0,+\infty),b) \wedge (c,(0,+\infty),d)) \vee (a,(-\infty,0],b)$
D	Time points a,b,c appear **in this order**. $(a,(0,+\infty),b) \wedge (b,(0,+\infty),c)$
E	Time point a **is between** time point b and time point c $((a,(-\infty,0),b) \wedge (a,(0,+\infty),c)) \vee$ $((a,(0,+\infty),b) \wedge (a,(-\infty,0),c))$
F	Time point a **being at least** d **after** time point b **correlates** with time point a **being at least** d **after** c $(b,[d,+\infty),a) \wedge (c,[d,+\infty),a)$

Table 1: A Sample of Abstracting Relations

approximating relations, since its appropriateness depends heavily on the abstracting relations being given and the temporal reasoning system being used. Two obvious problems may illustrate these interdependencies: First, for abstracting relations with quantitative parameters the proper instantiation of free parameters with actual values in the corresponding relation allows for redundant variation. Symmetric relations like "*time point* t_1 *is at most 1 unit away from* t_2" require particular care, since the equivalent "*time point* t_2 *is at most 1 unit away from* t_1" does not yield any new information. Second, additional constraints are needed to control proper instantiation of an abstracting relation. For instance, the definition A in Table 1 should be supplemented by the restriction that A_b and A_e really form an interval. Though, in principle, all pairs of time points may determine an interval that one could talk about, in practice, this generality should be avoided. Another additional constraint considered plausible for all abstracting relations is the unique name assumption which prevents, e.g., the unification of the three variables a, b, c in the abstracting relation F from Table 1.

Let us now illustrate our notion of generalization with two examples. Assume we want to mine the GTN resulting from (3) for interesting complex rules. For our first example, we are interested in temporal rules on how the arrival time of Mr. Roget influences the schedule of Mrs. Meyer appearing after him. Then, we add an unconstrained relation R_z to the GTN with $E_z := \{(t_0, t_1), (t_1, t_3)\}$. Composing (in the way defined by Staab [1998b]) the relation given in (3g) with the one from (3h) and projecting the result onto R_z yields:

(10) $((t_0, [3,3], t_1) \wedge (t_1, (2,6), t_3)) \vee$
$((t_0, [6,6], t_1) \wedge (t_1, (0,4), t_3))$

Generalizing this relation, obviously, only the abstracting relations D, E and F may apply (cf. Table 1), since the other ones require intervals instead of time points (e.g., A and B) or a different number of time points (*viz.* four as in C). Approximating "from above", abstracting relation F does not generalize (10) at all, while "*Time points* t_0, t_1, t_3 *appear in this order*" is the best generalization, since it is more specific than the corresponding instantiation of E. An approximation "from below" fails, because none of the abstracting relations is more specific than the relation in example (10).

Correspondingly, we may ask how Mr. George's arrival correlates with those of Mr. Roget and Mr. Meyer. Given that we have only qualitative information about the length of Mr. George's flight, it seems most appropriate to reason entirely on a qualitative interval structure. For the sake of brevity, we may here ignore many of the intricating presuppositions involved in algebraic operations on qualitative durations (cf. Clementini *et al.* [1997]) and simply present the result derived from the corresponding inference process:

(11) $((t_1, [\text{"medium"}, +\infty), t_5) \wedge (t_2, [\text{"medium"}, +\infty), t_5))$

This result is generalized ("from above *and* below") by "*Time point* t_5 *being at least a medium time after* t_1 *correlates with* t_5 *being at least a medium time after time point* t_2".

Conceptualizations at the interface level are of particular value for combining single evidence and generalizing it. In our text understanding application, e.g., we represent graded information like *"hard disk A is faster than hard disk B"* by GTN relations (cf. Staab and Hahn [1997]). Most of these relations can be handled by a comparatively inexpensive representation formalism. However, we also have to deal with much more complex utterances like *"up from a block size of 32 KB the data throughput decreases from 800 KB/s to less than 600 KB/s"*, which require more expressive representations, and, hence, costly reasoning. By flexibly assigning reasoning tasks to the least expensive representation level the entire understanding process might still be executed within feasible bounds. When just few of the represented GTN relations are complex, which is the case most of the time, reasoning at the finer levels remains feasible. Only if complicated GTN relations abound, one must resort to reasoning at coarser levels as an approximation — and eventually to an abstracting interface level that makes generalizations accessible to the user instead of a myriad of tiny bits of detail.

6 Related Work

The scalability of temporal reasoning, as a *static* notion, is already inherent to the hierarchy of calculi discussed in Section 4. It derives from the fact that these constraint systems stand for different levels of expressiveness. As the arrows in Fig. 1 indicate there are rather limited calculi (e.g., point algebra [Vilain *et al.*, 1989]), ones with increased expressiveness (e.g., Allen's calculus [Allen, 1983] or TCSPs [Dechter *et al.*, 1991]) and fairly general ones (such as integration models for Allen's calculus with metrical reasoning [Kautz and Ladkin, 1991; Meiri, 1996; Badaloni and Berati, 1996]). As a framework for our research, we have introduced a very general model, *viz.* GTNs, whose expressiveness exceeds that of all previously mentioned calculi. So this hierarchy lays down the foundations for formalizing temporal constraints at different levels of granularity. Weaker constraint systems may be an appropriate choice for applications which require less specific constraints and offer on their bonus side the tractability for certain reasoning algorithms.

Using this trade-off between expressiveness and computational complexity in a strategic manner leads to the idea to navigate this graph of different levels of expressiveness on demand — depending on the needs of the particular application. The idea to *dynamically* shift between less expressive and computationally cheaper systems and more expressive though computationally more expensive ones during run-time is the starting point of our work, and has been on the research agenda for quite a long time (cf. Hobbs [1985], Sathi *et al.* [1985], Nakhimovsky [1988]). This flexible manouvering between granularities as a principle method rather than as an impeding side conditions constitutes the main difference between our approach and common reasoning systems that implement several metric systems.

For instance, Bettini *et al.* [1997] have extended STP networks in order to represent interval structures from a large range of granularity levels. Thereby, they have even included non-contiguous structures (e.g., business days). As an approximating reasoning algorithm they propagate constraints in parallel networks of single granularities. Operators that map constraints between granularities communicate between the different networks. However, propositional abstraction, such as defined by our operator π is neglected in their approach as well as in other temporal reasoning systems.

This negligence may even be a drawback with regard to performance issues. Approaches for efficient temporal reasoning use, e.g., approximating propagation mechanisms [Schwalb and Dechter, 1997] or heuristics that optimize the search process [Stergiou and Koubarakis, 1998]. Though our proposal still lacks comparable empirical evidence, we can guarantee the determination of criteria important for the inferencing task (e.g., consistency for point algebra, path consistency for qualitative relations) in polynomial time, when granularities are switched to compute coarser results first, and refinements at more precise levels are postponed to subsequent rounds. Optimized schemes like those in [Schwalb and Dechter, 1997; Stergiou and Koubarakis, 1998] may still not terminate and, if they are terminated from outside due to exhausted time budgets (as set up by anytime devices, cf. Russell and Zilberstein [1991]), the network cannot be asserted to be in a similarly well-defined state as a cascade of GTNs at different granularities.

A complementary proposal has been made by Euzenat [1995], who permits to represent seemingly contradictory information at different levels of granularity, e.g., at some given level one may perceive that two intervals meet, while at a finer level one may recognize that the first is just a tiny bit before the second. His abstraction operators reflect how perception may change by switching between different levels.

7 Conclusion

In this paper, we defined operators for transient moves between different abstraction levels of temporal reasoning. Operators that were originally devised for constraint propagation, *viz.* $\pi, \mu, \trianglerighteq$, are used in our scalability framework for switching smoothly between different granularities as far as the reasoning proper and communication via interfaces are concerned. Furthermore, we proposed macro definitions to achieve abstraction by approximating relations at the interface level, a research issue that to the best of our knowledge has not been dealt with so far.

The major open issue is then *when* to bring *what* level of abstraction into play. In our opinion, there is no general solution to this problem. In the research environment we work in, a natural language text understanding system, the appropriate choice of adequate abstraction levels often comes with the author's choice of specific linguistic expressions occurring in the text, their corresponding semantic interpretation and the progression of the text (cf. Matsushita *et al.* [1998]). Having fixed such a starting point, we proceed from the cheapest level possible and turn to more expressive and expensive levels only when this is needed for proper text understanding.

Acknowledgements. S. Staab was a member of the Graduate Program *Human and Machine Intelligence* at Freiburg University, which was funded by DFG.

References

[Allen, 1983] J.F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[Badaloni and Berati, 1996] S. Badaloni and M. Berati. Hybrid temporal reasoning for planning and scheduling. In *Proc. of TIME-96*. IEEE Computer Society Press, 1996.

[Bettini *et al.*, 1997] C. Bettini, X.S. Wang, and S.S. Jajodia. Satisfiability of quantitative temporal constraints with multiple granularities. In *Proc. of CP-97*, pages 435–445. Springer, 1997.

[Clementini *et al.*, 1997] E. Clementini, P. Di Felice, and D. Hernández. Qualitative representation of positional information. *Artificial Intelligence*, 95(2):317–356, 1997.

[Dechter *et al.*, 1991] R. Dechter, I. Meiri, and J. Pearl. Temporal constraint networks. *Artificial Intelligence*, 49(1-3):61–95, 1991.

[Euzenat, 1995] J. Euzenat. An algebraic approach for granularity in qualitative space and time representation. In *Proc. of IJCAI-95*, pages 849–900, 1995.

[Giunchiglia and Walsh, 1992] F. Giunchiglia and T. Walsh. A theory of abstraction. *Artificial Intelligence*, 56(2-3):323–390, 1992.

[Hobbs, 1985] J.R. Hobbs. Granularity. In *Proc. of IJCAI-85*, pages 432–435, 1985.

[Kautz and Ladkin, 1991] H.A. Kautz and P.B. Ladkin. Integrating metric and qualitative temporal reasoning. In *Proc. of AAAI-91*, pages 241–246, 1991.

[Matsushita *et al.*, 1998] M. Matsushita, M. Ohta, and T. Iida. A visualization method of time expressions using starting/ending point plane. In *Proc. of TIME-98*, pages 162–168. IEEE Computer Society Press, 1998.

[McGuinness and Patel-Schneider, 1998] D. McGuinness and P. Patel-Schneider. Usability issues in knowledge representation systems. In *Proc. of AAAI-98*, pages 608–614, 1998.

[Meiri, 1996] I. Meiri. Combining qualitative and quantitative constraints in temporal reasoning. *Artificial Intelligence*, 87(1-2):343–385, 1996.

[Nakhimovsky, 1988] A. Nakhimovsky. Aspect, aspectual class, and the temporal structure of narrative. *Computational Linguistics*, 14(2):29–43, 1988.

[Russell and Zilberstein, 1991] S. Russell and S. Zilberstein. Composing real-time systems. In *Proc. of IJCAI-91*, pages 212–217, 1991.

[Sathi *et al.*, 1985] A. Sathi, M.S. Fox, and M. Greenberg. Representation of activity knowledge for project management. *IEEE Transactions on PAMI*, 7(5):531–552, 1985.

[Schwalb and Dechter, 1997] E. Schwalb and R. Dechter. Processing disjunctions in temporal constraint networks. *Artificial Intelligence*, 93(1-2):29–61, 1997.

[Staab and Hahn, 1997] S. Staab and U. Hahn. "Tall", "good", "high" — compared to what? In *Proc. of IJCAI-97*, pages 996–1001, 1997.

[Staab, 1998a] S. Staab. *Grading Knowledge — Extracting Degree Information from Texts*. PhD thesis, Freiburg University, 1998.

[Staab, 1998b] S. Staab. On non-binary temporal relations. In *Proc. of ECAI-98*, pages 567–571, 1998.

[Stergiou and Koubarakis, 1998] K. Stergiou and M. Koubarakis. Backtracking algorithms for disjunctions of temporal constraints. In *Proc. of AAAI-98*, pages 248–253, 1998.

[Vilain *et al.*, 1989] M. Vilain, H. Kautz, and P. van Beek. Constraint propagation algorithms for temporal reasoning: A revised report. In D.S. Weld and J. de Kleer, editors, *Readings in Qualitative Reasoning about Physics*, pages 373–381. 1989.