STEEL: A Spatio-Temporal Extended Event Language for Tracking Epidemic Spread from Outbreak Reports

Hervé Chaudet
Faculté de Médecine de Marseille, Laboratoire d'informatique médicale (herve.chaudet@medecine.univ-mrs.fr)

Abstract

We propose a Spatio-Temporal Extended Event Language (STEEL) for representing and reasoning about events that are described in outbreak reports. This language is an extension of the Event Calculus based on mereotopological relationships and structured conglomeration of events, in which time is replaced with spatiotemporal location. It allows representing and building aggregates of events according to the spatiotemporal location of their occurrence. In a proof a concept study, we aimed at comparing the performances of an experimental implementation in Prolog of this language with 3 human experts during a question-answering task on a trial corpus of 35 outbreak reports. This experiment showed experts' agreement with the system's responses.

INTRODUCTION

The use of emailed reports for an early and wide dissemination of epidemiological information by the Internet shows an increasing success for monitoring epidemiological events since its introduction.[1] The descriptive possibilities of these texts and their ability to deal with unattended situations make them competitive, comparing to epidemic registries, for reporting emerging infectious disease outbreaks and unusual disease patterns, including biological threats. That is why we are developing a system for automatic processing of outbreak reports with a double objective of question answering and qualitative modeling.

However, if the use of emails make easier the notification of epidemiological events, the automatic analysis and use of the transmitted information is particularly challenging, especially in a question answering task, where the system must retrieve answers rather than documents in response to a question. This requires the ability of:

- Building a representation of narrative's content that identifies epidemiological information and keeps the links with the texts. This can be made by information extraction systems that produce fact templates[2,3] or by natural language processing systems that build syntactic-semantic representations of narratives in a canonical form.[4]
- Reasoning about these representations for building the epidemic history, that is the spatiotemporal evolution of outbreak characteristics.

The experience of BIOSTORM on medical databases shows that this reasoning is highly knowledge consuming, requires particularly an explicit modeling of epidemiologic knowledge and the use of temporal and spatial abstractions for epidemic history building.[5] Moreover, lessons learned from the information extraction systems quoted above show that information in outbreak reports is event centered and that recovering the structure of outbreak scenarios is particularly difficult because of complex event structuring, inclusion relationships between events, scattering of events in texts, and information uncertainty.[6]

This paper presents our attempt for representing and reasoning about events that are described in outbreak reports such as emailed by the ProMED global electronic reporting system (http://www.promedmail.org/). After outlining the difficulties in using information from outbreak reports that have guided us, we describe the Spatio-Temporal Extended Event Language (STEEL) that we have developed for knowledge representation. In this extension, we have replaced time with spatiotemporal location and added a notion of spatiotemporal event aggregate for representing complex constructs of events. The last section of this paper succinctly describes an evaluation of its adequacy for representing the outbreak report contents by comparing the performances of this system with human experts in an *ad hoc* query situation with a trial corpus of 35 outbreak reports issued from ProMED-Mail.

USING INFORMATION FROM OUTBREAK REPORTS

Emailed outbreak reports are short news stories, which aim at reporting series of connected events that describe the spread of epidemics of infectious diseases. The term of "event" is used here for all that refers to "actions, events, motions, accomplishments and processes".[7] Figure 1 shows an example of emailed report issued from ProMED-Mail, updating information about an outbreak of dengue fever in Bangladesh.

Archive Number: 20021008.5493, Published Date: 08-OCT-2002

Bangladesh: Dengue Fever Continues to Spread
A dwindling supply of blood is exacerbating the dengue outbreak in Bangladesh. The total number of dengue-affected patients, according to the official accounts, stood at 4763 as of 16 Sep 2002. Of these, 45 have died so far. The Dengue Control Room sources said 509 persons affected with dengue virus were undergoing treatment at different hospitals across the country.
Meanwhile, it was reported that one person died from dengue hemorrhagic fever (DHF) in Magura district, while 4 others had been hospitalised in Jessore General Hospital in the last 24 hours. The Khulna City Corporation in the meantime has launched an anti-mosquito drive in the city. At least one person died in Chapai-Nawabganj district and 15 others have been hospitalised for DHF. All of them were admitted to the Rajshahi Medical College Hospital (RMCH) during the past 10 days. […]

Figure 1: Example of emailed report issued from ProMED-MAIL

The structure of outbreak reports is complex, intertwining and dispersing descriptive background information with story events throughout the narrative.[7] Three reasons may explain this:

- An emailed report may relate more than one epidemiological event, place or time.
- Its writing is highly influenced by a requirement of brevity.[8] For compactness, the story is crammed into a few complex sentences, complicating the structure of the narrative.
- It is often in a form of an update report relating the evolution of the epidemic characteristics since their last description.

The example in Figure 1 reports, in a single narrative, 10 events (8 related to patients), which concern a total of 6 spatial locations and 4 dates or time intervals, including the publication date. As illustrated, events "interlock and relate to each other in complex ways, forming inclusion relationships".[6] They frequently report aggregates of sub-events in a compact way, like in the sentence "Of them, 45 have died so far". An adequate representation of the events must capture these relationships, especially the sub-event composition.

A preliminary interview of experts in travel and tropical medicine, which are the main users of outbreak reports in our hospital, had reported that using information from outbreak reports requires reconstructing the relationships between events, their underlying temporal and spatial locations, and all descriptive background information that allows orientation in respect to person, place, time and epidemic situation.

THE KNOWLEDGE REPRESENTATION LANGUAGE

STEEL is a typed first-order logic language that is based on the Event Calculus (henceforth EC), which was introduced by Kowalski and Sergot[9] for representing and reasoning about the occurrence of events, the properties that events initiate and terminate, and the maximal validity intervals for which these properties hold uninterruptedly. Amongst the EC's extensions that have been developed in order to enhance its expressiveness, complex patterns of actions have been explored by Cervesato and Montanari,[10] showing the ability of process constructors for packaging up related events, and the problem of event's spatio-temporal location has been addressed by Galton[11] and Bennett.[12] STEEL carries on these works, introducing a joined spatiotemporal location of event, whose properties are used for ruling the building of event aggregates.

1. Language ontology. The basic ontology of STEEL comprises 4 basic types: events, fluents, time stamps, and spatial regions.

An Event corresponds to the performance or occurrence of an action over a specified time. If actions are time independent, defining "certain useful and relevant activities that may be conducted over some time by the agents to accomplish changes of state of the world",[13] events are time dependant. Discourse elements describing events can be identified on the basis of their "dynamic verbs".[7] Events are classified according to a scheme[14] that helps in the project of locating events in time:

- *Occurrence events* are the main event subclass and correspond to the events we want to place on the time axis (e.g. "one person died from dengue").
- *Reporting events* associate the source of information with an occurrence event (e.g. "it was reported that one person died").
- *Attitude events* are similar to reporting events, but they do not guarantee the reality of the information (e.g. "has died from a disease it was feared could be Ebola fever").
- *Aspectual events* that involve aspectual verbs like start, stop, begin etc.

Events may be instantaneous or may happen over a period of time,[15] defining the notion of event duration.

Fluents are valued expressions that describe the properties of system's objects (the value of a quality or a relation), and whose interpretations change from time to time (e.g. "dwindling supply of blood"). The fluent is valid when the object under consideration gets that specific property. Fluents' states are defined according to events that can initiate or terminate them.

Time stamps. Time is a concept that cannot be easily represented,[16] and several suggestions have been proposed for natural language processing.[17,18,19] Our aim is to represent temporal entities in a convenient manner for inducing the times and ordering of events. In our ontology, time is an ordered set $(T,\prec)$ where elements of T are Shahar's time stamps,[20] which are issued from time expressions encountered in the narratives and can be placed on a time axis (e.g. "16 September 2002"). Formally, this choice allows using event name for time specification, as proposed in the New Event Calculus,[21] or as showed in a study about temporal preposition phrases.[22]

Spatial regions. Space is two-dimensional and corresponds to the set $S=\mathcal{R}x\mathcal{R}$, where $\mathcal{R}$ is the set of reals. A region is a subset of S, usually represented only by a name (e.g. "Bangladesh", "Magura district"). A point is a special kind of region (e.g. "Rajshahi Medical College Hospital"). This choice keeps a level of complexity in accordance with discourse objectives and the way spatial relations are expressed in natural language. As pointed by Asher and Vieu,[23] the mathematical conception of topological space is foreign to space as it is usually expressed in narrative texts, where the reader may use the spatial information contained in texts, even though this information does not contain any system of coordinates.

From the two last types, we define the spatiotemporal location of an event performance as a couple <t,l>, where t is a time expression and l a spatial region.

2. Language description. The basic types are used for defining 3 sets of language basic predicates.

Events and their influences on fluents. This set of modified EC predicates is described in Table 1. In a first approach, we have avoided to deal with locations simultaneously different in time and space.

The relation between t_1 and t_2 in the predicate *hap-pens*$(e,<t_1,l>,<t_2,l>)$ is formalized by the following axiom:

$$happens(e,<t_1,l>,<t_2,l>) \rightarrow t_1 \preceq t_2$$

Table 1: Description and meanings of basic language predicates related to event occurrences and their influences on fluent values.

Predicate	Meaning
$t_1 \prec t_2$	Time stamp t_1 is before time stamp t_2.
$t_1 = t_2$	Time stamp t_1 is equal to time stamp t_2.
$happens(e,<t_1,l>,<t_2,l>)$	Event e starts at spatiotemporal location $<t_1,l>$ and ends at location $<t_2,l>$ (note: $happens(e,<t,l>) \equiv_{def} happens(e,<t,l>,<t,l>)$).
initiallyTrue(f,l)	Fluent f holds from time 0 at spatial location l.
initiallyFalse(f,l)	Fluent f does not hold from time 0 at spatial location l.
initiates(e,f,<t,l>)	Fluent f starts to hold after the occurrence of event e at spatiotemporal location <t,l>.
terminates(e,f,<t,l>)	Fluent f ceases to hold after the occurrence of event e at spatiotemporal location <t,l>.
releases(e,f,<t,l>)	Fluent f is no more subject to inertia after the occurrence of event e at spatiotemporal location <t,l> (its value becomes undetermined).

Spatiotemporal relations. Following Hazarika and Cohn's mereotopological theory of space-time,[24,25] spatiotemporal relations between objects can be represented with binary relations based on the notion of connection. Two entities are spatially connected (*sp-connected*) if they share at least a spatial point, though not necessarily simultaneously (e.g. Zaïre that has been renamed as Congo Démocratique). Temporal connection (*t-connected*) of two time intervals is defined by the existence of at least a common temporal point, though not necessarily at the same place. Finally two entities are spatiotemporally connected (*st-connected*) if the closures of these entities share at least a spatiotemporal point. This *α-connected(x,y)* primitive, where α ∈ {st, sp, t}, allows defining a set of 10 others mereotopological relations that constitutes the basis of a qualitative representation language (Table 2).

Table 2: Definition of spatio-temporal mereological relations from the primitive *α-connected(x,y)* (where α ∈ {st, sp, t})

Relation	Predicate	Definition
x is disconnected from y	α-disconnected(x,y)	¬α-connected(x,y)
x is a part of y	α-partof(x,y)	∀z . (α-connected(z,x) → α-connected(z,y))
x is a proper part of y	α-properpart(x,y)	α-partof(x,y) ∧ ¬α-partof(y,x)
x is identical with y	α-equal(x,y)	α-partof(x,y) ∧ α-partof(y,x)
x overlaps y	α-overlap(x,y)	∀z . (α-partof(z,x) ∧ α-partof(z,y))
x is discrete from y	α-discrete(x,y)	¬α-overlap(x,y)
x partially overlaps y	α-partoverlap(x,y)	α-overlap(x,y) ∧ ¬α-partof(x,y) ∧ ¬α-partof(y,x)
x is externally connected to y	α-externconnected(x,y)	α-connected(x,y) ∧ ¬α-overlap(x,y)
x is a tangential proper part of y	α-tangproppart(x,y)	α-properpart(x,y) ∧ ∃z . (α- externconnected (z,x) ∧ α- externconnected (z,y))
x is a non tangential proper part of y	α-nontangproppart(x,y)	α-properpart(x,y) ∧ ¬∃z . (α- externconnected (z,x) ∧ α- externconnected (z,y))

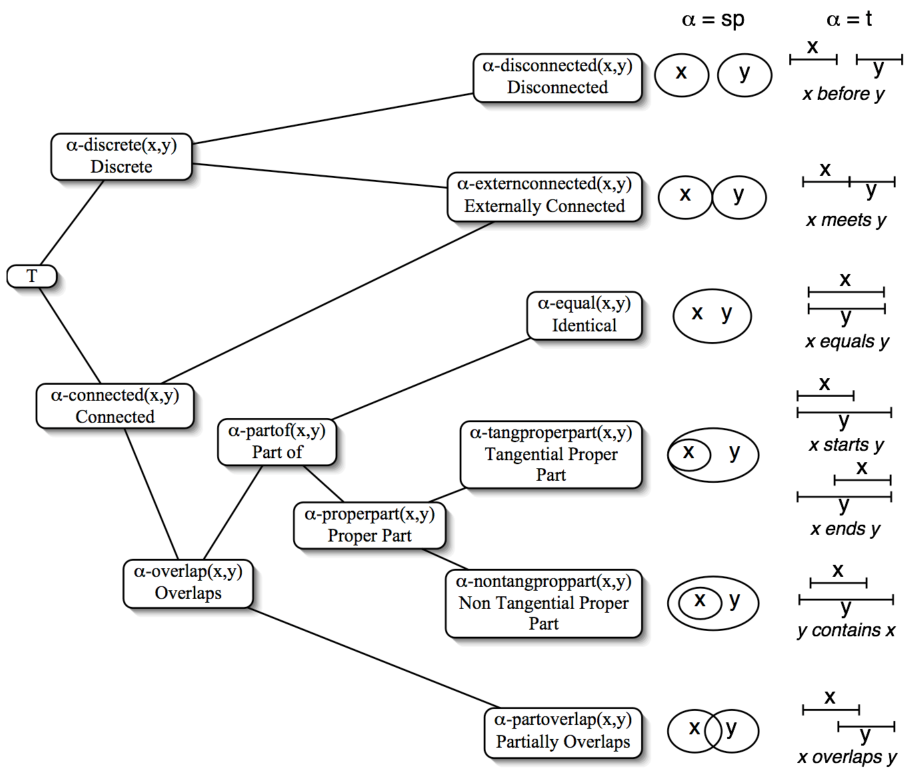

Figure 2: Subsumption lattice of basic mereotopological relations and graphical representation of their semantics for spatial and temporal domains

Figure 2 presents graphically these relations, their semantics for the (sp-) and temporal (t-) relations, the names given by Allen[17] in the last case, and shows their subsumption lattice. We have not figured the names of reciprocal relations. The six terminal relations in the lattice are provably Jointly Exhaustive and Pairwise Disjoint. Experimental results concerning the cognitive adequacy of the interval relations for spatial representation and reasoning showed that people use ordinal information similar to this calculus when representing and remembering spatial arrangements.[26]

The spatio-temporal connection of two entities can then be easily defined by the simultaneous existence of a member of the graph subtree of Figure 2:

st-connected(x,y) → t-connected(x,y) ∧ sp-connected(x,y)

These mereological relations allows introducing an axiom of spatial persistence, stating that an event happens in a region if it happens in a part of this region (e.g. if there is a case of Ebola fever in Gabon then there is a case of Ebola fever in Africa):

happens(e,<t_1,l_2>,<t_2,l_2>) ← happens(e,<t_1,l_1>, <t_2,l_1>) ∧ sp-partof(l_1,l_2)

Macro-events. Capturing the sub-event compositions that are reported in outbreak reports requires more than just representing the hierarchical relationships between events mentioned above in section 2. Representing complex pattern of events involves additional relations among events, such as sequentiality, simultaneity, iteration, or temporal delays between events. Cervesato and Montanari introduced macro-events, which are expressions defined by applying the following grammar,[10] where m is a macro-event, d and D are time expressions with d < D, and e is an event:

m ::= e (basic event)
| $m_1 ;_d^D m_2$ (sequence with delay d to D)
| $m_1 + m_2$ (alternative)
| $m_1 \parallel m_2$ (parallelism)
| m^n (n-time iteration)

We consider that a macro-event is an occurrence of a structured conglomeration of events, and is a direct subconcept of *Event*.

Each macro-event instance is defined by an instance of a macro-event structure (MES), where S is a formula obtained by applying recursively the grammar.

A resulting MES is a tree in which the leaves are event subconcepts. A MES can be used for defining subclasses of events, and if m is a macro-event and MES_m its structure, then $MEClass_m \doteq MacroEvent \sqcap MES_m$.

Table 3 presents how we have expressed these macro-events in predicate form. These rules give a

Table 3: Definition of macro-events in first order logic.

Event	Predicate	Definition
$m_1 ;_d^D m_2$	$seqevent(m_1,m_2,d,D)$	$happens(m,\langle t_1,l\rangle,\langle t_2,l\rangle) \wedge meventdef(m, seqevent(m_1,m_2,d,D)) \leftrightarrow \exists\, t_3, t_4 .\ (happens(m_1,\langle t_1,l\rangle,\langle t_3,l\rangle) \wedge happens(m_2,\langle t_4,l\rangle,\langle t_2,l\rangle) \wedge t_3+d \preceq t_4 \preceq t_3+D)$
$m_1 + m_2$	$altevent(m_1,m_2)$	$happens(m,\langle t_1,l\rangle,\langle t_2,l\rangle) \wedge meventdef(m, altevent(m_1,m_2)) \leftrightarrow happens(m_1,\langle t_1,l\rangle,\langle t_2,l\rangle) \vee happens(m_2,\langle t_1,l\rangle,\langle t_2,l\rangle)$
$m_1 \parallel m_2$	$parevent(m_1,m_2)$	$happens(m, \langle t_1,l\rangle,\langle t_2,l\rangle) \wedge meventdef(m, parevent(m_1,m_2)) \leftrightarrow \exists\, t_3, t_4, t_5, t_6 .\ (happens(m_1, \langle t_3,l\rangle,\langle t_4,l\rangle) \wedge happens(m_2, \langle t_5,l\rangle,\langle t_6,l\rangle) \wedge t_1=min(t_3,t_5) \wedge t_2=max(t_4,t_6))$
m^n	$iterevent(m,n)$	$happens(m, \langle t_1,l\rangle,\langle t_2,l\rangle) \wedge meventdef(m, iterevent(E,n)) \leftrightarrow \exists\, t_3, t_4 .\ (happens(m_1, \langle t_1,l\rangle,\langle t_3,l\rangle) \wedge happens(m_2, \langle t_4,l\rangle,\langle t_2,l\rangle) \wedge meventdef(m_1, iterevent(E,n-1)) \wedge E(m_2) \wedge t_3 \preceq t_4)$

logical framework for building the representation of complex events with coherent time boundaries The relation between a macro-event m and its MES is given by a predicate *meventdef(m, MES_m)*.
Macro-events can substitute plain events in the EC defined above, in particular in the predicates *initiates* and *terminates*, allowing properties to be started and ended by generic macro-events.

3. Continuity reasoning toolbox. History of epidemic spread must be reconstructed from the spatiotemporal connections of events and fluents that are described through partial observations reported in emailed texts. This task needs in particular to capture the notions of spatiotemporal continuity of fluents for determining the maximal validity intervals (henceforth MVI) and the structures of macro-events described over several texts (evolution of the number of deaths, of new cases…).

Table 4: Description and meanings of basic language predicates related to event influence on fluent value persistence.

Predicate	Meaning
$clipped(\langle t_1,l\rangle,f, \langle t_2,l\rangle)$	Fluent f is terminated between time t_1 and time t_2 at spatial location l.
$declipped(\langle t_1,l\rangle,f, \langle t_2,l\rangle)$	Fluent f is initiated between time t_1 and time t_2 at spatial location l.
$holdsAt(f,\langle t,l\rangle)$	Fluent f holds at spatiotemporal location <t,l>.

Fluent persistence. Events' influences on the persistence of fluent values with respect to time and spatial location are expressed with a set of 6 axioms describing the semantics of 3 basic predicates that are presented in Table 4:

$$clipped(\langle t_1,l_1\rangle,f,\langle t_4,l_1\rangle) \leftrightarrow \exists\, e,t_2,t_3,t_5,l_2 .\ happens(e,\langle t_2,l_2\rangle,\langle t_3,l_2\rangle) \wedge (terminates(e,f,\langle t_5,l_1\rangle) \vee releases(e,f,\langle t_5,l_1\rangle)) \wedge t_2 \prec t_5 \prec t_3 \wedge t_1 \prec t_5 \prec t_4 \wedge sp\text{-}partof(l_1,l_2)$$

$$declipped(\langle t_1,l_1\rangle,f,\langle t_4,l_1\rangle) \leftrightarrow \exists\, e,t_2,t_3,t_5,l_2 .\ happens(e,\langle t_2,l_2\rangle,\langle t_3,l_2\rangle) \wedge (initiates(e,f,\langle t_5,l_1\rangle) \vee releases(e,f,\langle t_5,l_1\rangle)) \wedge t_2 \prec t_5 \prec t_3 \wedge t_1 \prec t_5 \prec t_4 \wedge sp\text{-}partof(l_1,l_2)$$

$$holdsAt(f,\langle t,l\rangle) \leftarrow initiallyTrue(f,l) \wedge \neg clipped(\langle 0,l\rangle,f,\langle t,l\rangle)$$

$$holdsAt(f,\langle t,l_1\rangle) \leftarrow \exists\, e,t_1,t_2,t_3,l_2 .\ happens(e,\langle t_1,l_2\rangle,\langle t_2,l_2\rangle) \wedge initiates(e,f,\langle t_3,l_1\rangle) \wedge t_1 \prec t_3 \prec t_2 \wedge t_3 \prec t \wedge \neg clipped(\langle t_3,l_1\rangle,f,\langle t,l_1\rangle) \wedge sp\text{-}partof(l_1,l_2)$$

$$\neg holdsAt(f,\langle t,l\rangle) \leftarrow initiallyFalse(f,l) \wedge \neg declipped(\langle 0,l\rangle,f,\langle t,l\rangle)$$

$$\neg holdsAt(f,\langle t,l_1\rangle) \leftarrow \exists\, e,t_1,t_2,t_3,l_2 .\ happens(e,\langle t_1,l_2\rangle,\langle t_2,l_2\rangle) \wedge terminates(e,f,\langle t_3, l_1\rangle) \wedge t_1 \prec t_3 \prec t_2 \wedge t_3 \prec t \wedge \neg declipped(\langle t_3, l_1\rangle,f,\langle t, l_1\rangle) \wedge sp\text{-}partof(l_1,l_2)$$

Event aggregation and spatiotemporal continuity. For the purpose of constructing spatiotemporally located event aggregates, we introduced a constructor, written $\bigwedge_h$, which combines the events of two *happens* predicates for building the structure of the resulting macro-event, depending on their spatiotemporal and ontological relationships.
Let e_1 and e_2 be two instances respectively of $E_1 \sqsubseteq$ *Event* and $E_2 \sqsubseteq$ *Event*. *Happens*$(e_1,\langle t_1,l_1\rangle,\langle t_1',l_1\rangle)$ and *happens*$(e_2,\langle t_2,l_2\rangle,\langle t_2',l_2\rangle)$ are their representations in the knowledge base. The time interval during which each event occurs can be respectively defined as $[t_1,t_1']=d_1$ and $[t_2,t_2']=d_2$.
Two main cases of constructor's behavior must be considered depending on the ontological relationships between the two events.
If:

- $E_1=E_2$ with disjoint instances, that is:
  - The instances are differents: $e_1 \neq e_2$
  - Or, in the case of macro-events, the interpretations of events' classes are disjoint: $MEClass_{e1}^{\mathcal{I}} \cap MEClass_{e2}^{\mathcal{I}} = \emptyset$
- Or $E_2=Macroevent \sqcap iterevent(E_1,*)$

Then the instances are related to a same event and the macro-event constructor proceeds to its iteration.
In all other cases, the structure of the resulting event involves sequentiality or parallelism, depending on the st-relationships between the events. Table 5 summarizes the constructor's results.

Table 5: Results of the macro-event constructor ∧h(happens(e1,<t1,l1>,<t1',l1>), happens(e2,<t2,l2>,<t2',l2>))

Case: E_1=E_2=E or E_2=Macroevent ⊓ iterevent(E_1=E,*)	
sp relationships	*Results*
sp-partof(l',l) with ($l'=l_1 \wedge l=l_2$) or ($l'=l_2 \wedge l=l_1$)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m,iterevent(E,*)) \wedge t=min(t_1,t_2) \wedge t'=max(t_1',t_2')$
sp-partoverlap (l_1,l_2) ∨ sp-discrete(l_1,l_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m,iterevent(E,*)) \wedge t=min(t_1,t_2) \wedge t'=max(t_1',t_2') \wedge l=(l_1 \cup l_2)$
Otherwise	
sp relationship: case sp-equal(l_1,l_2)	
t relationships	*Results*
t-discrete (d_1,d_2)	$happens(m,<t_1,l>,<t_2',l>) \wedge meventdef(m,seqevent(e_1,e_2,d,d)) \wedge d=t_2-t_1' \wedge l_1=l_2=l$
equals(d_1,d_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m,parevent(e_1,e_2)) \wedge t_1=t_2=t \wedge t_1'=t_2'=t' \wedge l_1=l_2=l$
t-properpart(d_1,d_2)	$happens(m,<t_2,l>,<t_2',l>) \wedge meventdef(m, parevent(e_1,e_2)) \wedge l_1=l_2=l$
overlaps(d_1,d_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m, parevent(e_1,e_2)) \wedge t=min(t_1,t_2) \wedge t'=max(t_1',t_2') \wedge l_1=l_2=l$
sp relationship: case sp-discrete(l_1,l_2) ∨ sp-partoverlap(l_1,l_2)	
t relationships	*Results*
t-discrete (d_1,d_2)	$happens(m,<t_1,l>,<t_2',l>) \wedge meventdef(m,seqevent(e_1,e_2,d,d)) \wedge d=t_2-t_1' \wedge l=(l_1 \cup l_2)$
equals(d_1,d_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m,parevent(e_1,e_2)) \wedge t_1=t_2=t \wedge t_1'=t_2'=t' \wedge l=(l_1 \cup l_2)$
t-properpart(d_1,d_2)	$happens(m,<t_2,l>,<t_2',l>) \wedge meventdef(m, parevent(e_1,e_2)) \wedge l=(l_1 \cup l_2)$
overlaps(d_1,d_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m, parevent(e_1,e_2)) \wedge t=min(t_1,t_2) \wedge t'=max(t_1',t_2') \wedge l=(l_1 \cup l_2)$
sp relationship: case sp-properpart(l_1,l_2)	
t relationships	*Results*
t-discrete (d_1,d_2)	$happens(m,<t_1,l>,<t_2',l>) \wedge meventdef(m,seqevent(e_1,e_2,d,d)) \wedge d=t_2-t_1'$
equals(d_1,d_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m,parevent(e_1,e_2)) \wedge t_1=t_2=t \wedge t_1'=t_2'=t'$
t-properpart(d_1,d_2)	$happens(m,<t_2,l>,<t_2',l>) \wedge meventdef(m, parevent(e_1,e_2))$
overlaps(d_1,d_2)	$happens(m,<t,l>,<t',l>) \wedge meventdef(m, parevent(e_1,e_2)) \wedge t=min(t_1,t_2) \wedge t'=max(t_1',t_2')$

Macro-event occurrence and maximum validity intervals of properties. A macro-event m, which structure is MES_m, has occurred over a spatiotemporal interval [$<t,l>,<t',l>$], written *meo(m,t,t',l)*, iff: $\exists\ t_1,\ t_2,\ l_1\ .\ meventdef(m,MES_m) \wedge happens(m,<t_1,l_1>,<t_2,l_1>) \wedge t \preceq t_1 \preceq t_2 \preceq t' \wedge sp\text{-}partof(l_1,l)$. The macro-event occurrence may be not explicitly present in the knowledge base, and determined recursively using the result definitions of the macro-event constructor $\bigwedge_h$.

The MVI of a property or a fluent p, written mvi(p,t,t',l), is the maximal spatiotemporal interval [$<t,l>,<t',l>$] over which p holds uninterruptedly. This can be written as:

$$
\begin{aligned}
mvi(p,t,t',l) \leftrightarrow\ & t \prec t' \wedge \\
& (initiallyTrue(p,l) \vee (happens(e_1,<t_1,l_1>,<t_2,l_1>) \wedge \\
& initiates(e_1,p,<t,l>) \wedge t_1 \prec t \prec t_2 \wedge sp\text{-}partof(l,l_1)) \wedge \\
& happens(e_2,<t_3,l_2>,<t_4,l_2>) \wedge \\
& (terminates(e_2,p,<t',l>) \vee releases(e_2,p,<t',l>)) \wedge \\
& t_3 \prec t' \prec t_4 \wedge sp\text{-}partof(l,l_2) \wedge \\
& \neg clipped(<t,l>,p,<t',l>)
\end{aligned}
$$

STEEL is able to determine the MVI, to check the truth of MVIs or macro-event occurrences, and to process Boolean combinations of MVI and macro-event occurrence verifications.

EVALUATION

As proof of concept, we have studied the adequacy of STEEL for representing outbreak report contents by comparing the performances of an experimental implementation of this language in Prolog with human experts in a query situation.

STEEL was implemented in SWI-Prolog (University of Amsterdam, http://www.swi-prolog.org/). The axioms and definitions from the previous section of this paper were transcribed into prolog rules. The whole language kernel is a module of about 200 rules. The trial knowledge base was built from a trial corpus of 35 emailed outbreak reports issued from the ProMED mail list and describing an outbreak of Ebola fever in Gabon from December 2001 until May 2002. The size of this trial corpus was 8105 words, 213 sentences. From this corpus, 224 events and 328 objects have been extracted with a simple annotation tool built ad hoc, which solicited the annotator for inferring the spatiotemporal locations of events when they were not directly specified in the text. Amongst these events, 148 were macro-events. The annotator responses were either a location or a relational (i.e. precedence, inclusion…) expression involving a location. Figure 3 shows an excerpt of the resulting knowledge base that is about 2000 prolog rules long.

Source
X-ProMED-Id: 20011205.2950 Date: 2001-12-05 Subject: PRO/EDR> Viral hemorrhagic fever, suspected - Gabon Source: WHO Disease Outbreaks Report, Wed 5 Dec 2001 [edited] On Tue 4 Dec 2001, WHO received reports of 7 deaths in an outbreak of suspected viral haemorrhagic fever in Ogooue Ivindo Province in the northeastern part of the country.
Representation
ist(happens(reports(proMED,system,proMedMail20012950), [[2001,12,05],_]),system). … ist(happens(reports(_,WHO,[proMedMail20012950|1]),[[2001, 12,04],_]),proMedMail20012950). ist(happens(event1,[time1,'Ogooué Ivindo Province'], [time2,'Ogooué Ivindo Province']), [proMedMail20012950|1]). ist(agent(event1,isPossibly('viral haemorrhagic fever',_)), [proMedMail20012950|1]). ist(sp-partof(northEastPart('Gabon'),'Ogooué Ivindo Province'), [proMedMail20012950|1]). … instance(event1,macroevent). … meventdef(event1,itervent(death,7)).

Figure 3: Excerpt of the trial knowledge base.

For the experimentation we have built from the trial corpus a test set of 18 questions covering a spectrum of question types. Two examples of questions follow:

- What is the number of new cases of Ebola fever in Gabon between 2001-12-29 and 2002-1-6?
- What names of cities and villages located in Ogooué-Ivindo Province are mentioned in the outbreak reports?

These questions have been addressed in logic formalism to the system, which found a response in every case. The CPU time required for a response was 273±98 ms on a PowerPC G4 under Darwin/MacOSX at 1.2 Ghz with 512 Ko of L2 cache and a 167 Mhz bus.

Then we gave in a booklet the trial document collection to 3 experts in tropical and/or travel medicine that are usual users of ProMED-Mail. In a first test we have asked them to answer to the same questions. In a second test we have given them the system's responses, without indicating their origin, and asked them if these responses were satisfactory or not.

Each expert took between 75 and 90 minutes for completing the experiment. The expert-expert and expert-system answer accordance in the first test is summarized in Table 6. The number of different answers stands at 3 in all cases, except one case of total concordance between an expert and the system. Testing the homogeneity of accordance distribution shows that experts make no distinction between the system and another expert (Fisher's exact probability test on a 2x6 contingency table, p-value=0.4676). After merging the expert's responses, we found that a total of 4 questions have got at least one expert's response different from the system. We have considered this merging as the maximal discordance, and we have tested it with a binomial sign test on the number of agreements versus the number of discordances. The critical probability P_{H0}(number of agreements ≥ 14)=0.015442 confirms the correctness of the system's responses according to the experts. This conclusion is confirmed by the results of the second part of the experience, which reports that all system's responses are appropriate for all experts.

Table 6: Expert-expert and expert-system accordance / discordance ratios for the trial set of 18 questions.

	System	Expert 2	Expert 3
All experts	4 / 14	-	-
Expert 1	0 / 14	-	-
Expert 2	3 / 15	3 / 15	-
Expert 3	3 / 15	3 / 15	3 / 15

CONCLUDING REMARKS AND FUTURE WORK

In this paper we introduce an extension of the Event Calculus suited to the representation of information extracted from outbreak reports within the framework of automatic following of epidemic spread.

The adequacy of Event Calculus to narratives' representation,[27] the structural characteristics of these texts, especially their centering on spatiotemporally located events, have pushed us to use this formalism rather than those issued from the Situation Calculus,[27,28] or from action-based models.[29,30,31,32,33] In its original version or its extensions, this calculus is unable to easily represent the joint spatiotemporal location of events or the events' aggregation with respect to their occurrence location (required for merging information from several reports). Others models based on chronicles have been developed for representing spatiotemporal situation described in narratives,[34] but in an objective of situation recognition and not of modeling and information summarization as in our case.

Our extension allows a representation that is very close to the narrative, centered on event occurrences and keeping the event relationship network. It also allows to group together several report contents for building a global description of an outbreak occurrence, considering an epidemic as a complex event, which results from the aggregation of the events that are reported (spatiotemporal abstraction mechanism).

However, if this representation seems to be adequate for the purpose of representing information extracted from outbreak reports, its ontology (a modeling language ontology) cannot be considered as those of epidemiology (a domain ontology), although it may or must be a part of. Further work in this direction must be made to get a system that is able to proceed to an automatic modeling of outbreaks.

The system developed in this work cannot be considered as a complete information extraction/representation system. It only focuses on the final step of a complete natural language processing system that remains to be completed. However, we expect that STEEL would help us solving the problem of location and event identification[6,14,35] during natural language processing of outbreak reports.

ACKNOWLEDGMENTS

This research was supported in part by the French Ministry of Research and Technology (ACI "Télémédecine et Santé" n° 2000/120), by a grant from the Délégation Générale à l'Armement of the French Defense (DGA n° 00.34.030.00.470.75.65) and by the ADIMI Association for Medical Informatics.

References

[1] Woodall J. Official versus unofficial outbreak reporting through the Internet. International Journal of Medical Informatics 1997;47:31-4.

[2] Damianos L, Day D, Hirschman L et al. Real users, real data, real problems: the MiTAP system for monitoring bio events. In: Proceedings of BTR2002: Unified Science & Technology for Reducing Biological Threats & Countering Terrorism; The University of New Mexico, Albuquerque, New Mexico; March 2002.

[3] Grishman R, Huttunen S, Yangarber R. Information extraction for enhanced access to disease outbreak reports. Journal of Biomedical Informatics 2002;35:236–46.

[4] Roux M, Ledoray V. Understanding of medico-technical reports. Artificial Intelligence in Medicine 2000;18:149–72.

[5] Buckeridge DL, Graham J, O'Connor MJ, Choy MK, Tu SW, Musen MA. Knowledge-based bioterrorism surveillance. Proc AMIA Symp. 2002;:76-80.

[6] Huttunen S, Yangarber R, Grishman R. Diversity of scenarios in information extraction. In: Proceedings of 3rd International Conference on Language Resources and Evaluation (LREC 2002); Las Palmas, Canary Islands, Spain; May 2002.

[7] Schokkenbroek C. News Stories - Structure, time and evaluation. Time and Society 1999; 8(1):59-98.

[8] Bell A. The discourse structure of news stories. In: A Bell and P Garrett, editors. Approches to Media Discourse. Oxford: Blackwell; 1998. p. 64-104.

[9] Kowalski R, Sergot M. A logic-based calculus of events. New Generation Computing 1986;4:67–95.

[10] Cervesato I, Montanari A. A calculus of macro-events: progress report. In: A Trudel, SD Goodwin (eds), 7th International Workshop on Temporal Representation and Reasoning (TIME'00); Cape Breton, Nova Scotia, Canada; 7-9 July 2000. IEEE Computer Society Press; 2000. p. 47-58.

[11] Galton A. Toward an integrated logic of space, time, and motion. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93); Chambéry, France; 28 August-3 September 1993. p. 1550-5.

[12] Bennett B. Space, time, matter and things. In: Proceedings of the international conference on Formal Ontology in Information Systems (FOIS'2001); Ogunquit, Maine, USA. New York, NY, USA: ACM Press; 2001. p. 105-16.

[13] Knight B, Peng T, Ma J. Reasoning about change over time: actions, events, and their effects. In: The Fourth Symposium on Logical Formalizations of Commonsense Reasoning (CS98); Queen Mary and Westfield College, London, UK; 7-9 January 1998. p. 183-97.

[14] Setzer A, Gaizauskas R. Annotating events and temporal information in newswire texts. In: Proceedings of the Second International Conference On Language Resources And Evaluation (LREC-2000); Athens, Greece; 31 May- 2 June 2000. p. 1287-93.

[15] Shanahan M. Solving the frame problem. Cambridge, Massachussets: MIT Press; 1997.

[16] Combi C, Shahar Y. Temporal reasoning and temporal data maintenance in medicine: issues and challenges. Computers in Biology and Medicine 1997;27(5):353-68.

[17] Allen JF. Towards a general theory of action and time. Artificial Intelligence 1984;23:123-54.

[18] Galton A. A critical examination of Allen's theory of action and time. Artificial Intelligence 1990;42:159-88

[19] Steedman M. Temporality. In: J Van Bentham and A ter Meulen, editors, Handbook of logic and language. Amsterdam: Elsevier;1997. p. 895-938.

[20] Shahar Y. A framework for knowledge-based temporal abstraction. Artificial Intelligence 1997;90:79-133.

[21] Sadri F, Kowalski R. Variants of the Event Calculus. In: L Stirling Ed),Proceedings of the International Conference on Logic Programming,

Kanagawa, Japan, June 1995, The MIT Press, p. 67-81.
[22] Pratt I, Francez N. Temporal prepositions and temporal generalized quantifiers. Linguistic and Philosophy 2001;24(2):187-222.
[23] Asher N, Vieu L. Toward a geometry of common sense: a semantic and complete axiomatization of mereotopology. In: Proceedings of IJCAI'95; Montréal, Québec, Canada; 20-25 August 1995. p. 846-52.
[24] Hazarika SM, Cohn AG. Qualitative spatio-temporal continuity. In: D Montello, editors. Proceedings of COSIT'01; Morro Bay, California, USA; September 2001. Volume 2205 of Lecture Notes in Computer Sciences. Berlin: Springer-Verlag; 2001. p. 92-107.
[25] Cohn AG, Hazarika SM. Continuous transitions in mereotopology. In: Commonsense-2001: 5th Symposium on Logical Formalizations of Commonsense reasoning; New York, USA; 2001. p. 71-80.
[26] Knauff M. The cognitive adequacy of Allen's interval calculus for qualitative spatial representation and reasoning. Spatial Cognition and Computation 1999;1:261-90.
[27] Cervesato I, Franceschet M, Montanari A. A guided tour through some extensions of the event calculus. Computational Intelligence 2000;16(2):307-47.
[28] Miller R, Shanahan M. Narratives in the Situation Calculus. Journal of Logic and Computation 1994; 4(5 - Special Issue on Actions and Processes):513-30.
[29] Miller R. Situation Calculus specifications for Event Calculus logic pograms. In: Proceedings of the 3rd International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR'95), 1995, Springer Verlag (LNAI 928), p. 217-30.
[30] Kakas A, Miller R. A simple declarative language for describing narratives with actions. Journal of Logic Programming 1997;31 (1-3) (Special Issue on Reasoning about Action and Change):157-200.
[31] Kakas AC, Miller R, Toni F. ε-RES : reasoning about actions, events and observations. In : T. Eiter, W. Faber, M. Truszczynski (Eds), 6th Int. Conf. LPNMR 2001, Vienna, Austria, 17-19 September 17-19 2001, p. 254-66.
[32] Baral C, Gelfond M. Reasoning about effects of concurrent actions. Journal of Logic Programming 1997; 31:85-118.
[33] Baral C, Son TC, Tuan L. A transition function based characterization of actions with delayed and continuous effects. In: D Fensel, F Giunchiglia, DL McGuinness, MAe Williams, Proceedings of the Eights International Conference on Principles and Knowledge Representation and Reasoning (KR-02), Toulouse, France, 22-25 April 2002, p. 291-302.
[34] Sansonnet JP, Gerard S. A spatio-temporal model for the representation of situations described in narrative texts. In: D Christodoulakis (eds), Natural Language Processing - NLP 2000, Second International Conference, Patras, Greece, 2-4 June 2000, Springer, Lecture Notes in Computer Science 1835, p. 176-84.
[35] Filatova E, Hovy E. Assigning Time-Stamps to Event-Clauses. In: Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing, ACL-2001, Toulouse, France, 6-11 July, p. 88-95