Uncertain Temporal Support for Medical Record Ontology

Monika Žáková, Kamil Matoušek, Olga Štěpánková
Department of Cybernetics
Czech Technical University in Prague
Technická 2, 160 00 Prague
Czech Republic
{matousek, zakova}@labe.felk.cvut.cz

Taťána Maříková
Institute of Biology and Medical Genetics
2nd Medical School, Charles University
V Úvalu 84, 150 00 Prague 5
Czech Republic
tana.marikova@lfmotol.cuni.cz

Abstract

This paper describes an intelligent approach to representing uncertainly specified time periods applicable particularly to medical records. It is based on using the operational conceptual modelling language OCML. An example of a temporal reasoning capability in data annotated by means of temporal ontology is given.

1 Introduction

Records of patients collected in hospitals or in consulting rooms of medical specialists represent valuable source of data. However most patient's records are stored in form of text files describing individual patient's status. These text files are semi-structured. They contain some structured data such as body weight, blood pressure etc., but a significant part of the report consists of notes made by the doctor. The notes are written in natural language, often in rather complex sentences. The use of natural language enables the doctor to express some uncertain and imprecise information.

In order to be able to use data for data mining and decision support and also for sharing data effectively among different medical laboratories and hospitals, it is necessary to convert as much information as possible into a structured form with added semantics. This can be achieved by annotation of medical records using ontologies and subsequent knowledge extraction. To capture as much detail from the patients' records, several ontologies are necessary.

2 Ontologies for Medical Records

We are currently developing an ontology describing the structure of medical records that are used in the Institute of Biology and Medical Genetics at Charles University in Prague. The ontology is being developed in Apollet language[1], which is based in OCML [Motta, 1999] and stored in XML format. Except for rules that are written in Lisp, the ontology can also be exported to OWL [Lacy, 2005]. The decision to use Lisp for representation of rules was taken, because implementation of reasoning in Lisp is available and there is currently no generally accepted standard for rules in OWL. The attempts to formulate such standard are still in form of a proposal [Horrocks, 2004].

The ontology follows guidelines for examinations of patients. At present it contains structure of records of examinations relevant for patients with neurofibromatosis.

In addition to the task ontology describing structure of medical records, some domain ontologies have to be used e.g. an ontology of diseases, ontology of drugs etc. In case of inherited diseases also an ontology describing family relations from genetic point of view was necessary and it was developed.

During initial experiments with annotation of medical records using MedAT (Medical Records Annotation Tool)[Žáková, 2005], it was discovered that in addition to medical domain ontologies, also ontologies capturing the imprecise and uncertain information contained in the medical records would be of a significant utility.

3 Time representation

Patients are located in space and time, while living in contexts and periods of their individual lives. Tracking changes of patients' health during time has always been a method that promised answering questions of professional physicians.

Concerning the individual expressions of time, there is a wide range of precise, imprecise or vague, uncertain dating, which cause difficulties and inaccuracy in their use. Some reasons of inaccuracy of objects dating may be e.g. the following:

- data is not available (i.e. no written resources),
- it is a subjective information given by the patient
- processes that lasted for a longer time are referred to as a single time point (e.g. feeling a pain).

The reliability of resources is also an important factor in terms of validity of temporal assertions, when patient's subjectivity influences data.

In the domain of time, statements like "in childhood", "soon after the surgery" or "during the treatment in hospital XY" are some examples. Our contribution is an attempt to deal with uncertainty in temporal assertions. We aim at suggesting a suitable and effective inference mechanism extension for ontology of medical records, which would yield sufficiently accurate localisation in time.

[1] Used in Apollo ontology editor [Eisenstadt, 1998]

4 Temporal Statement Categories

Concerning the people, the main temporal properties correspond to their lives and some events they encountered.

Major time property is the duration of a time period (e.g. *convalescence period*, *pregnancy*), which could be expressed in terms of starting and ending time points. However, time durations may be relative as well (e.g. *for three months*) and thus having no exact starting or ending time. Time properties are not often expressed in a straightforward way as they may be inherent in the data. In this case, there are many expressions with different semantics (e.g. *tomorrow*, *at the beginning of the year*, *Monday*, or *June 5th*).

Regarding the individual expressions of time, there is a wide range of precise, imprecise, or uncertain dating, which causes difficulties and further inaccuracy in any subsequent use of the data. So, assigning a value to a time property of an event does not mean that the event can simply be stick to a defined position on a timescale. Uncertain data can have either inexact position on the timescale or inexact duration. The properties of time continuity and causality lead to the existence of implicit bindings of the time events and periods, which need to be respected while inferring some conclusions.

A categorization of temporal statements containing the most frequent expressions with respect to their accuracy with utility for the domain of our interest has been proposed in [Matoušek, 2004]:

1) Precise statements. The whole data is available, maximum precision is reached (e.g. *January 12, 2006, 12:30:00*).
2) Statements with higher granularity. Data is available, but not so precise. It is necessary to distinguish instants and intervals (e.g. *January 12, 2006* can be seen either as an instant of higher granularity or as a 24 hour time interval)
3) Incomplete statements. Some information is missing for precise time identification. One may intentionally use this kind of statement for recurring temporal positions – regularly repeated instants (e.g. *January 12, 12:30:00*)
4) Uncertain statements with absolute specification of uncertainty (e.g. *between February 12 and February 13, 2006*).
5) Uncertain statements with relative specification of uncertainty (e.g. *around February 12, 2000, before 90s*).
6) Statements referencing other statements with temporal properties (e.g. *after the first surgery*, *during the last in-hospital treatment*).
7) Statements with unknown or missing information (e.g. *that time I was doing …*).

We leave aside temporal statements representing relative multiplicity of recurrence (e.g. *often, rarely,* and *sometimes*). Expressions related to the current time (e.g. *yesterday*, *tomorrow*) are supposed to implicitly belong to the category 6 above.

A remark must be made that the semantics of the same temporal statement may vary depending on the context. Particularly for *around*, it may vary between time periods. For example, the statement *around the year 2001* has more uncertainty included than the statement *around last Monday*. This problem we address by introducing temporal granularity support.

5 Ontological Model of Uncertain Time

In this section we describe how theoretical framework for temporal reasoning proposed in [Matoušek, 2004] can be practically used. We describe a calendar and time system that was used in our prototype of a temporal inference engine. We also provide a short description of a temporal reasoning system implemented in OCML language. It enables users to enter historical calendar data and to carry out temporal reasoning on a knowledge base containing temporal definitions.

6 Temporal Support

6.1 Allen Relations

There are thirteen possible and mutually exclusive relations between two time intervals $I(i_1, i_2)$ and $J(j_1, j_2)$ called *Allen's interval relations* introduced in [Allen, 83]:

I precedes J	iff i_2 *before* j_1
I precedes inverse J	iff j_2 *before* i_1
I meets J	iff i_2 *equals* j_1
I meets inverse J	iff j_2 *equals* i_1
I overlaps J	iff i_1 *before* j_1 and j_1 *before* i_2 and i_2 *before* j_2
I overlaps inverse J	iff j_1 *before* i_1 and i_1 *before* j_2 and j_2 *before* i_2
I costarts J	iff i_1 *equals* j_1 and i_2 *before* j_2
I costarts inverse J	iff i_1 *equals* j_1 and j_2 *before* i_2
I during J	iff j_1 *before* i_1 and i_2 *before* j_2
I during inverse J	iff i_1 *before* j_1 and j_2 *before* i_2
I cofinishes J	iff j_1 *before* i_1 and i_2 *equals* j_2
I cofinishes inverse J	iff i_1 *before* j_1 and i_2 *equals* j_2
I equals J	iff i_1 *equals* j_1 and i_2 *equals* j_2

For graphical representation of the thirteen Allen relations see Figure 1 (moving from left to right corresponds to the passing of time).

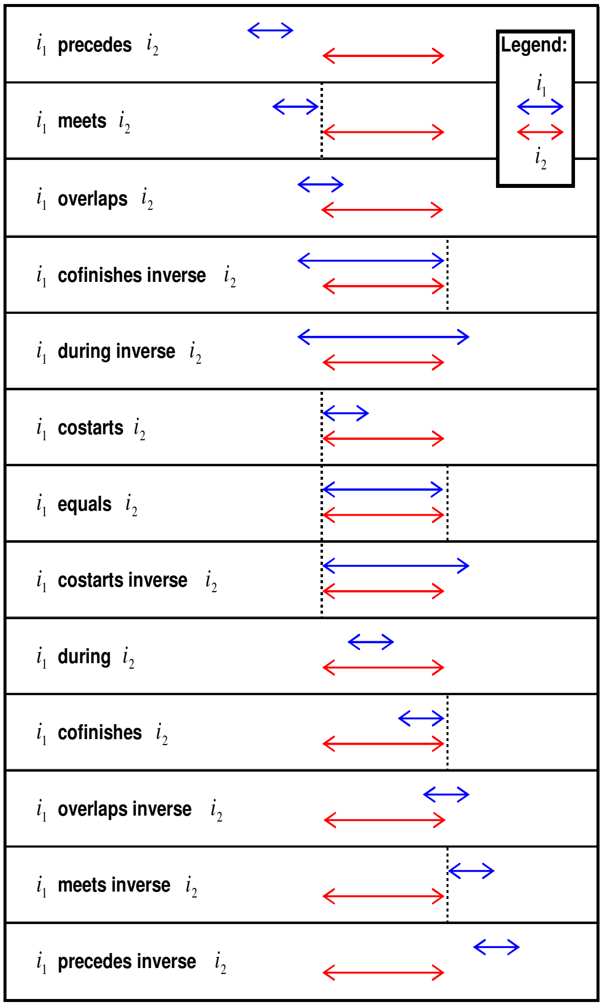

Figure 1: Thirteen Allen's temporal relations of two time intervals

6.2 Time Granularity

Time granularity is a concept known and used across information and knowledge based systems. For example in *data warehouses* [Mat01], which store data in a specific multidimensional form suitable for ad-hoc data analysis, consider data granularities on various dimensions including the time granularities, along which coarser or finer data views can be defined. However, in ontological knowledge modelling of time instants and periods it is often neglected.

Time Granularity is the level of detail in which the time is considered (measured). The finest temporal scale defines the *finest granularity*.

Different time statements can refer to different time granularities. E.g. "May, 12, 2003" is a time statement with day granularity, while "In 2002" has the year granularity. With coarser time granularities their granularity values are defined, too. Time granularity defines its own unit scale for time positions.

Granularity temporal scale is a temporal scale, which orders all granularity values within a specified granularity.

Time Point with Granularity is a generalized time point. Its location can span over multiple (simple) time points in the finest temporal scale.

So a time point with granularity is represented by the granularity value and its time granularity. Its *location* is its representing time interval.

On the other hand, there exists another function which assigns to a time point with specified granularity its numerical *temporal position on its granularity temporal scale*.

Time points with granularity, as well as simple time points, do not posses the property of duration. Their temporal property is to represent rather instants than intervals, but from a coarser perspective than simple time points.

Granularity values are mapped into the finest temporal scale via *granularity representation function,* which transforms the given value of granularity into the *representing time interval* at the finest temporal scale.

Let us show an example of a relation between two time points with granularity (t^g). This is done as an extension of Allen's interval relationships for two time points with different granularity:

t_1^{g1} *precedes* t_2^{g2} iff *End(* t_1^{g1} *) before Start(* t_2^{g2} *)*

For time points with granularity common relations used to compare the finest time points were presented [Matoušek, 2004].

However, unlike simple time points, these relations naturally do not cover all possible mutual locations of two time points with granularity. This is caused by the encapsulated uncertainty, which is involved when using higher time granularities.

E.g. Time point with granularity can fall into the representing time interval of another point with coarser granularity.

In order to support statements on higher granularity durations (e.g. *"three months"*) we introduce:

Time quantity with granularity is a measurable non-negative amount of time, which can be expressed in terms of granularity temporal scale representing the distance of two time points with that granularity.

Time quantity with granularity can be obtained e.g. as difference of two time points with the same granularity:

Time intervals with granularity have their *duration* specified using time quantities with granularity. They *may* have specified time points with granularity as its endpoints. If their location with granularity is known, then their *duration with granularity* can be calculated.

6.3 Uncertainty of Temporal Position

So far we have discussed the temporal representation of precise temporal statements either on the finest temporal scale or influenced by uncertainty using time granularity. Now we will go further into representation of other kinds of time uncertainty.

As we have seen in the discussion on uncertainty types (see section 6.3), there are various types of indeterminate expressions: *now*, *(some time) before*, *(some* *time) after*, *at the break of a decade*, *during*, or even *unknown*. *Relative time points* need not have specified location. *Recurring* temporal entities may have more locations (*Monday*; *January*; *12 pm*; *Saturday 3 am* etc.). They represent sets of time points with granularity. *Relative time interval* may have both its endpoints' locations unspecified.

We introduce *Time Uncertainty* as a concept, whose values represent individual uncertainty types.

Uncertain Time Point is a time point, which has its (uncertain) location and possibly spanning over multiple time points.

Time uncertainty values appear explicitly as properties of uncertain time points.

The *location* of an uncertain time point is not given but *constrained*. However, in order to use the relative time uncertainty ranges, we use a reference position, which yields the *positioning context* of the time point.

The amount of uncertainty is given by a temporal entity *Range of uncertainty* represented on the finest temporal scale, which includes all possible temporal locations of the time point.

Time uncertainty types, which we consider, fall into time intervals. Other approaches may consider e.g. also uncertainty in recurring temporal entities.

The properties of time uncertainty are *FromTimePoint, ToTimePoint, BeforeRelTime* and *AfterRelTime*. The supported representations of time uncertainty are:

For *absolute* specification of a *time range of uncertainty,* given by specified time points, we introduce uncertainty as a property pair of two time points: *FromTimePoint* and *ToTimePoint*.

For *relative time range of uncertainty*, there is another pair of relative temporal position properties and their respective granularities: *BeforeRelTime, AfterRelTime, BeforeGranularity* and *AfterGranularity*.

For such uncertainty values the approach calculates the *representing time interval* at the finest temporal scale.

Uncertain location references the uncertain time point into a specific location on a temporal scale.

To clarify the usage of this time uncertainty representation, let us give here some examples:

Temporal statements like "*around the year 2002*", "*after 3rd of June 1990*", will use in its representation the *BeforeRelTime* and *AfterRelTime* together with uncertain location properties. On the other hand if an event happened in some time between "*The first examination* and *the last manifestation of a specific symptom*", then *FromTimePoint* and *ToTimePoint* will be used.

For uncertain time points we cannot easily define common relations used to compare the finest time points like before, equals, after. We can only express their sufficient conditions corresponding to certainty or necessary conditions corresponding to possibility.

These possible and necessary bounds fulfill the rule that if one time point is *certainly before* another, then also the former is *possibly before* the latter.

Of course, an analogous property holds for the remaining pairs: <*certainly equal, possibly equal*> and <*certainly after, possibly after*>.

There is naturally a mapping of time points with granularity into their representation as uncertain time points. It is not possible to simply express by a general formula all time points with specified granularity as time points with a concrete uncertainty, because individual granularity values are mapped into the finest granularity, which need not be defined analytically by a formula.

6.4 Combining Time Uncertainty and Granularity

As there is natural to combine uncertainty and granularity specifications, we have to allow uncertain time points with granularity, which have specified both granularity and uncertainty. The representing time intervals in this case can be obtained as well as in the simpler cases.

An example of the possibility expression which is valid follows:

$$^{u1}t_1^{g1} \text{ possibly equals } {^{u2}t_2^{g2}}$$
$$\text{iff } not(\,Start(\,^{u2}t_2^{g2}\,)\ after\ End(\,^{u1}t_1^{g1}\,)\,)$$
$$and\ (not\ Start(\,^{u1}t_1^{g1}\,)\ after\ End(\,^{u2}t_2^{g2}\,))$$

For uncertain time points with granularity common relations used to compare the finest time points i.e. *before* and *after* expressed again in terms of their sufficient and necessary conditions can be easily derived e.g.:

$$^{u1}t_1^{g1} \text{ certainly before } {^{u2}t_2^{g2}} \quad \text{iff } {^{u1}t_1^{g1}} \text{ certainly precedes } {^{u2}t_2^{g2}}$$

The possible and necessary bounds of common comparison relations fulfil the rule that if one time point is *certainly after* another, then also the former is *possibly after* the latter. The same property holds for the pairs *certainly equal, possibly equal* and *certainly after, possibly after.* This can be easily proven as there are just the thirteen mentioned possible and mutually exclusive relations of two time points' representing intervals.

7. Combining Time Uncertainty and Granularity

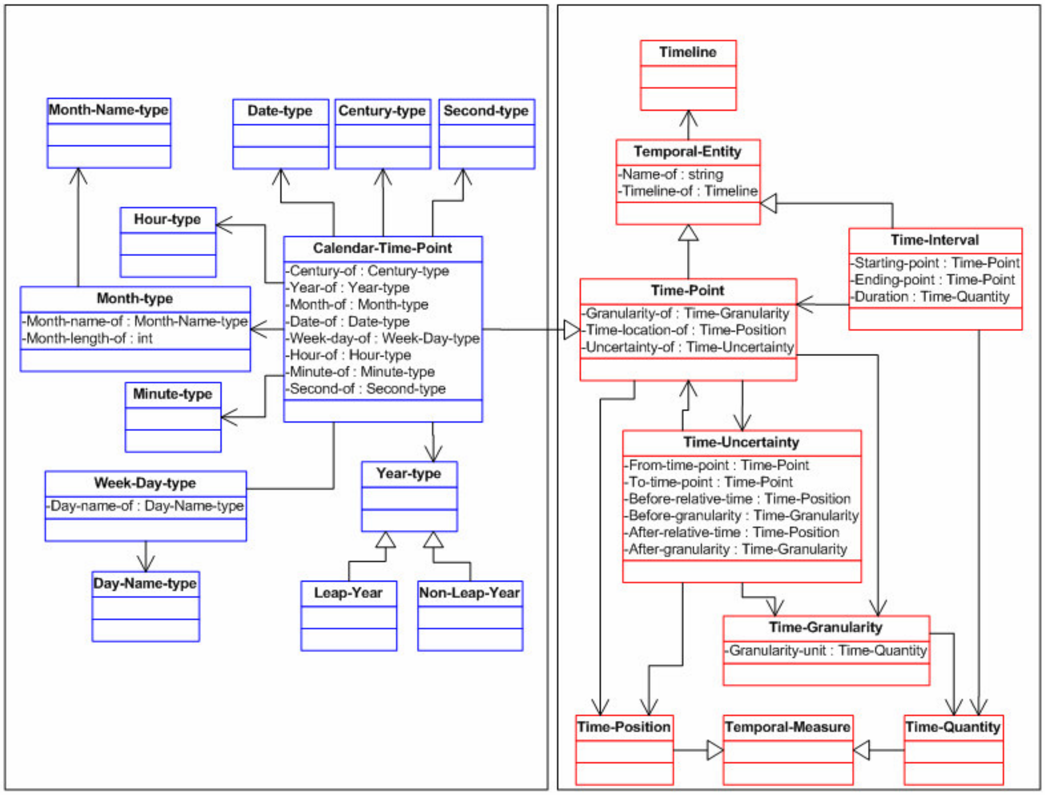

Figure 2. Basic class structure

Our temporal reasoning engine builds on the inference capabilities of OCML. Its primary temporal coordinate system (*temporal scale*) is chosen to be that one internally built-in to the *Common LISP* language. It uses the temporal scale with zero point equal to 1.1.1900 0:00:00 UTC and the shortest interval of one second. However, the extent of the LISP dates is limited. Thus, in order to offer calendar dates and times in an ancient history, it was necessary to introduce functions that decode and encode broader range of time positions.

The basic class structure of the time ontology in our temporal inference engine is depicted using *unified modelling language* in the Figure 2. Classes in the right frame correspond to the definitions in our theoretical framework for reasoning in the time domain. The main components of our temporal model are subclasses of *temporal-entity* – *time-point* and *time-interval*. The property *timeline-of* of *temporal-entity* is a sort of stereotype, which enables distinguishing different kinds of temporal entities. Any instance of temporal entity can have any number of timelines assigned to it. Query results can be constrained by including the name of a timeline that is the subject of interest in the query submitted to the temporal inference engine. Thus, timelines define a kind of namespaces. Beside classes, the model contains basic time point and time interval relations, rules, and functions that handle both granularity and uncertainty introduced as part of the theoretical framework.

7.1 Calendar Dates and Time

In order to support calendar date and time specification, we introduce a subclass of *time-point*, denoted as *calendar-time-point* with the system of classes around it in the left frame of the Figure 1. The slots *century-of*, *year-of*, *month-of*, *date-of*, *week-day-of*, *hour-of*, *minute-of*, and *second-of* enable representation of individual parts of calendar date and time. The prototype inference engine provides internal routines, which are used to recalculate these slots into the internal representation of time locations and conversely back from the time location into these slots. When filling out only the slots with the respective information available or applicable, incomplete date specifications or those on a coarser granularity level can be used.

Some general constraints should always be satisfied, when working with temporal entities. One example is the property of *transitivity* of functions *before* and *equals*.

8 Simple Practical Examples

Let us show how the temporal inference engine can be used. We shall model the life and illness periods of a fictive person, John Newman with patient id 100000. At the beginning the important time points have to be defined.

```
(def-instance John-Newman-birth Calendar-Time-point
  ( (date-of 14) (month-of 5) (year-of 1916)
    (granularity-of day-granularity)))
(def-instance John-Newman-illness-start Calendar-Time-point
  ( (date-of 26) (month-of 8) (year-of 1946)
    (granularity-of day-granularity)))
(def-instance John-Newman-illness-end Calendar-Time-point
  ( (date-of 29) (month-of 11) (year-of 1978)
    (granularity-of day-granularity)))
```

The following time intervals are related to the specified time points:

```
(def-instance Illness-John-Newman- Time-interval
  ( (starting-point John-Newman-illness-start)
    (ending-point John-Newman-illness-end)))
(def-instance Life-John-Newman Time-interval
  ( (starting-point John-Newman-illness-start)
    (ending-point John-Newman-death)))
```

If a statement is inaccurate, we may need to create an instance of time uncertainty. Here, we model the moment John Newman got injured, which happened sometime around the year 1970. It is possible to include uncertainty parameterisation as in this case using *param-around-unc*:

```
(def-instance param-around-unc time-parameter((value-of 10)))

(def-instance Around-a-Year Time-Uncertainty
  ( (Before-relative-time param-around-unc)
    (Before-granularity year-granularity)
    (After-relative-time param-around-unc)
    (After-granularity year-granularity)))
(def-instance John-Newman-Injury Calendar-Time-point
  ( (year-of 1970) (granularity-of year-granularity)
    (uncertainty-of around-a-year)))
```

8.1 Inference Examples

Having defined the necessary facts, we may be interested in particular results using the inference engine. For example, having consulted data concerning all different examinations performed, the following query will retrieve the time points corresponding to the individual examinations between the first and the second surgery:

```
(ocml-eval
  (findall ?a
    (and (timeline-of ?a Examinations)
         (and (before first-surgery ?a)
         (after second-surgery ?a)))))
```

The returned result then finds e.g. a neurological examination:

```
(NEUROL-EX-100000-040911)
```

8.2 Terminology used in Medical Practice

The first type of concepts added to the time ontology were periods of life as recognized by the doctors e.g. childhood, infancy, juniors, adults, seniors.

Another category of time points include events referencing another time point e.g. with respect to the previous examination, after the first surgery.

The third category consists of time instants in with respect to patient's life e.g. in 8 months, in 42. week.

The fourth type of time specification that can be encountered are time specifications in natural language that can vary both in granularity and accuracy e.g. this year, in this time, in the long term, for a couple of days.

The last category of statements includes statements using the proposition since and a time interval e.g. since infancy, since fall 1998 to indicate that a phenomenon started at some time point within the specified interval and continues up to the present.

Conclusions

Current results of our research are giving useful means for machine representation of temporal aspects in medical patient's records including the uncertain facts. Using the proposed solution the existing medical record ontology was extended and the first example results have been shown.

In future the ontology should be further extended to cover fully the variety of record kinds of the patients treated at the Czech Institute of Medical Genetics.

An effort is currently given to equip Apollo ontology editor with suitable input/output filters from/to OWL format and other ontology representations.

In temporal framework, further research topics include the possibility to operate with recurrent time events (e.g. *every 4 months*).

Acknowledgement

The research is supported by the research grant of the Czech Ministry of Education, Youth and Sport: "Transdisciplinary Research in the Area of Biomedical Engineering II", No. MSM6840770012 and the grant of the Czech University Development Fund (FRVŠ) No. 33-05023.

References

[Allen, 1983] J. F. Allen. Maintaining Knowledge about Temporal Intervals. Communications of the ACM, 26(11):832-843, 1983.

[Žáková, 2005] Monika Žáková, Olga Štěpánková, Taťána Maříková: 'MedAT: Medical Resources Annotation Tool.' In Proceedings of EMBEC 2005, Prague, Czech Republic, to appear.

[Eisenstadt, 1998] M. Eisenstadt, and T. Vincent, *The Knowledge Web – Learning and Collaborating on the Net*, Kogan Page, London, 1998.

[Horrocks, 2004] Ian Horrocks, Peter F. Patel-Schneider: 'A Proposal for an OWL Rules Language', International WWW Konference 2004, New York, USA, available online September 10.

[Lacy, 2005] Lee W. Lacy: 'OWL: Representing Information Using the Web Ontology Language', Trafford, Victoria, Canada, 2005.

[Matoušek, 2004] Matoušek, K. - Uhlíř, J.: On Representing Uncertain Historical Time. In Proceedings of DEXA 2004. Los Alamitos: IEEE Computer Society Press, 2004, s. 105-109. ISBN 0-7695-2195-9.

[Motta, 1999] Enrico Motta: `Reusable Components for Knowledge Modeling'. IOS Press, 1999.