

# Recognition and normalization of temporal expressions in Serbian medical narratives<sup>1</sup>

UDC 811.163.41'322.2:61

DOI 10.18485/infodhca.2019.19.2.2

**ABSTRACT:** The temporal dimension emerges as one of the essential concepts in the field of medicine, providing a basis for the proper interpretation and understanding of medically relevant information, often recorded only in unstructured texts. Automatic processing of temporal expressions involves their identification and formalization in a language understandable to computers. This paper aims to apply the existing system for automatic processing of temporal expressions in Serbian natural language texts to medical narrative texts, to evaluate the system's efficiency in recognition and normalization of temporal expressions and to determine the degree of necessary adaptation according to the characteristics and requirements of the medical domain.

**KEYWORDS:** Natural language processing, temporal reasoning, TIMEX, TimeML, clinical records.

**PAPER SUBMITTED:** 10 October 2019

**PAPER ACCEPTED:** 06 December 2019

Jelena Jaćimović

jelena.jacimovic@stomf.bg.ac.rs

*University of Belgrade*

*School of Dental Medicine*

*Belgrade, Serbia*

## 1 Introduction

As a reflection of physical reality changes in human consciousness, or the way the human mind perceives and interprets events, time determines a person's understanding of the world. The temporal dimension is a central aspect not only of our daily lives but also of understanding the changes and problems that arise within other specific domains.

---

<sup>1</sup> This paper is part of the research conducted within the author's doctoral dissertation, defended at the Faculty of Philology, University of Belgrade.

In the field of medicine, time also emerges as one of the essential concepts (Shahar, 1999; Augusto, 2005; Zhou and Hripcsak, 2007; Reeves et al., 2013), providing a basis for the proper interpretation and understanding of medically relevant information. The particular order of symptoms, the right time of therapy administration, duration, and frequency are significant only in a specific temporal context. Moreover, during the diagnosis, it is particularly important to know the chronological order of individual symptoms or their duration.

Without temporal dimension, it is almost unlikely to adequately present clinically relevant data, nor to accurately infer and make decisions based on them. Different medical interventions occur at one or more points in time (for example, *hirurška intervencija zakazana za 29.02.2000. g. u 9 h* 'surgical intervention scheduled for 2/29/2000 at 9 am'). Furthermore, specific data, like laboratory test results or data on diagnoses and prescribed therapies, are valid only within an explicitly or implicitly defined period (such as *po dobijanju gore navedenog rezultata ordinirana je infuzija Tygacil 50 mg 10 dana* 'after receiving the above result, Tygacil 50 mg infusion administered for 10 days' or *tokom hospitalizacije ordinirana konzervativna terapija* 'conservative therapy administered during hospitalization'). Besides, relevant clinical data and recommended or performed interventions are important medical concepts, which are often temporally related (e.g. *potrebno nekoliko HD dijaliza posle implantacije grafta* 'few hemodialyses required after graft implantation'). These temporal relations need to be discovered and resolved to correctly determine the chronology of events or conditions, which enables, for instance, the successful monitoring of disease development and the effectiveness of the therapy applied.

### 1.1 The importance of automatic processing of temporal information in medical narratives

Time is crucial for the representation of information contained in medical information systems. Such information is relevant to patients, containing accurate and comprehensive data on medical conditions, diagnoses, treatment courses and outcomes. Modern advances in medical information technology have made it possible to automate most of the processes involved in providing healthcare services, facilitating the collection and storage of patient information. The real need to store these correctly temporally oriented data exists, first of all, because of the possibility of their later use in another context

to improve the quality of medical care services, diagnostic and theoretical processes in medicine.

Electronic health records contain a certain amount of information provided in a structured form. The existence of predefined data types and their relations (patient’s name or health insurance number, doctor’s visit date, etc.) enables their machine processing. However, for a more intelligent decision support system, these data types present a deficient resource, as they lack key information related to clinical treatment courses and outcomes. This crucial information, such as disease condition and its course of development, is often recorded only in unstructured texts (like physicians’ clinical notes), containing a wealth of clinically relevant information presented by complex, non-standard forms of natural language. In order to locate and extract them from medical narratives, natural language processing technologies are applied. For extracted medical concepts to be chronologically organized and used in this context, it is necessary to provide automatic extraction and interpretation of temporal information found in medical texts. The way of presenting and concluding based on time-oriented clinical data is equally important for both healthcare providers or medical professionals as well as automated systems designed for general electronic patient record search or decision support during making a diagnosis, determining therapy, or evaluating its effectiveness. For example, after the administration of „sublingual allergen-specific immunotherapy” to evaluate the clinical efficacy of treating allergic rhinitis and asthma in children, it would be useful to know whether nasal congestion and rhinorrhea symptoms were statistically significantly reduced after the first measurement, in the first six months of the therapy.

## **2 The nature of medical narrative texts and the challenges in automatic processing**

Automatic processing of temporal expressions (TEs) of medical texts involves their identification and formalization in a language understandable to computers, which is not an easy task. Besides the complex nature of time and its possible structure model (linear, branching, or circular), this problem is also affected by some difficulties arising from the domain specificity and the language used in these texts.

As one of the most significant forms of medical narrative texts, clinical notes represent the physician’s written reports of a visit and examination of a patient, intending to collect all the information necessary to detect the disease and make an accurate diagnosis. These texts contain an abundance of

temporal information and facts, listed in chronological order, from previous therapies and symptoms, through the current state, to future interventions. To enable reliable and accurate conclusions to be drawn from the collected data and the analysis, these texts must contain as much detail as possible. Their recording is time-consuming, and the doctors rarely have enough time. Since these texts are written by physicians for their colleagues and other medical professionals who have the same expert knowledge, a variety of non-standard terms and abbreviations are commonly used, providing a concise but sufficiently detailed description at the same time. Thus, in addition to the commonly used TEs, certain abbreviations of temporal meaning are often used in medical narratives to express time (for instance, *qid* from the Latin *quarter in die* – four times a day, *bid* from the Latin *bis in die* – twice a day, *post op* – postoperative, after a surgical operation). In medical narratives, like in any other natural language text, the same temporal information can be expressed in various forms (e.g. *January 3, 2007*; *03.01.07*; *01/03/2007*). An extensive review and classification of TEs found in discharge summaries, as one type of medical narrative text, are given in (Zhou et al., 2006). However, since clinical notes are written in a short time in a specific, medical language, characterized by fragmented and incomplete utterances, the lack of punctuation marks and formatting, as well as many spelling and grammatical errors, the number of forms of TEs that need to be recognized, due to their importance in interpreting clinically relevant data, becomes significantly higher (e.g. *03 01 07*; *03.01,2007*; *3.01..2007*; *0'3.01.2007*).

Another challenge in recognizing and formalizing the TEs of medical narrative texts is the fact that temporal information does not necessarily have to be explicit, but is often implicit and requires an interpretation or inferences based on general knowledge. In addition to absolute, relative TEs whose value is not explicitly stated and have to be taken from the context are also frequently used. These expressions require the value of another TE serving as an anchor for determining which particular time is meant. For example, the value of the expression *12. mart o.g.* 'March 12 this year' is relative to the date clinical note was written, while in the case of the *trećeg p.op dana* 'the third postoperative day', the calendar date of the mentioned surgery will serve as an anchor. In the statement „ranije lečena od dijabetične flegmone” 'previously treated for diabetic phlegmon' the temporal information processing system should determine if „ranije” 'previously' refers to a period of, for instance, a week or a year earlier, requiring higher levels of analysis and resolution of co-references, which is an unresolved issue that

has not yet been adequately discussed in the context of medical narrative data processing (Sun et al., 2013b).

Given the nature of medical narratives language, which is characterized by inconsistent application of grammatical and spelling rules, frequent use of abbreviations and text parts copied from other texts, it is necessary to collect and analyze as many different types of medical narrative texts as possible, to allow the development of a system that would perform high-response and precision automatic analysis. However, this is not a simple task, given the need to protect the right to the confidentiality of patients’ health information. In addition to a large amount of useful, medically relevant information, the narrative texts of this domain also contain many personal details, including a patient’s medical condition. For both ethical and legal reasons, when confidential clinical data are shared and used for research purposes, it is necessary to protect patient privacy and remove patient-specific identifiers through a process of de-identification. De-identification is focused on detecting and removing/modifying all explicit personal Protected Health Information (PHI) present in medical or other records, while still preserving all the medically relevant information about the patient. These PHI categories include names, geographic locations, elements of dates (except year), telephone and fax numbers, medical record numbers or any other unique identifying numbers, among others.

## 2.1 Previous work in TE medical narratives processing

Automatic processing of temporal information is one of the dominant natural language processing areas, although it could not be said to have had the same treatment in all domains of language use. Despite the progress made in the general domain characterized by the narrative style of a standard language, which is a rich source of TEs and relations examples, more active work on revealing temporal information contained in medical narratives has only been occurring for the last ten years (Lin et al., 2015). Nevertheless, the problem of representation and inference using some of the temporal aspects has attracted the attention of biomedical researchers for decades (Savova et al., 2009; Meystre et al., 2008; Augusto, 2005). Several studies aimed at processing the temporal information of medical texts were conducted even in the late 1980s (Johnson, 1987; Hirschman, 1981; Obermeier, 1985). The first system to identify words and linguistic forms that carry temporal meaning in medical texts was developed at the University of New York under the LSP (Linguistic String Project) program (Sager, 1967).

This system, designed to process standard English, is adapted to the needs of applying to medical texts - LSP-MLP (Linguistic String Project - Medical Language Processor), with the aim of identifying, analyzing and formalizing various forms of temporal information into a representation that will allow physicians to extract and summarize details on the symptoms, administration, and dosage of drugs, as well as reactions to their administration and possible side effects (Hirschman, 1981; Lyman et al., 1985). Also, one of the first systems for temporal information of medical texts analysis, known as GROK (Grammatical Representation of Objective Knowledge) (Obermeier, 1985), was developed using texts related to liver disease. The result of this system was a text representation based on knowledge of medical field key concepts, within which relevant medical events were extracted and chronologically ordered.

During the last decade of the 20th century, an increasing number of researchers have been concerned with the role of time in medicine and the temporality of medical narrative texts, publishing the achieved results both in medical and computer science journals as well as in conference proceedings (Keravnou, 1991; Goodwin and Hamilton, 1996; Combi and Shahar, 1997). The most significant system developed during that period at Columbia University in New York (Friedman et al., 1993), known as MedLEE (Medical Language Extraction and Encoding System), has been used since 1995 for the needs of the Presbyterian Hospital of Columbia in New York City. However, during this period, a large number of researchers interested in temporal resolution in the clinical domain dealt primarily with structured data in the form of explicitly encoded events (laboratory tests, doctor visits, etc.), which were dated and stored in databases. A comprehensive methodological overview of the approaches developed and the proposed standards for formalizing clinical-domain temporal information, which will enable the exchange of data among healthcare information systems, is given in (Augusto, 2005).

More recently, there has been an increasing interest of researchers in the effective use of temporal information of medical narratives and their incorporation into information extraction systems. Solving the problem of extraction of temporal information in the field of medicine, unlike the general domain, requires consideration of the system's design to be used for decision support in diagnosis, treatment determination, clinical data summarization, epidemiological studies conduction, etc. (Adlassnig et al., 2006). For example, in (Denny et al., 2010) the authors developed a system that extracts the time and status of screening tests from electronic health records for the early

detection of colorectal cancer. Similar ideas were applied to determine the status of the drugs used. To study patients’ exposure to Warfarin<sup>2</sup> at the time of hospital admission, Liu and collaborators (2011) developed a system that extracts data on the use of certain medicines at a given time from patients’ electronic health records. Concerning the automatic processing of temporal expressions of unstructured data, such as medical narratives, several papers provide a summary of the published literature (Meystre et al., 2008; Zhou and Hripcsak, 2007).

The problem of automatic processing of temporal information of general-domain narrative texts has been the focus of several international challenges, which have been addressed by a large number of researchers, primarily due to the existence of a sufficient number of publicly available annotated corpora necessary for the development of these systems. However, given the medical texts nature whose public use is not possible, more active involvement in the automatic processing of temporal information of medical texts was enabled in 2012 when for the i2b2 (Informatics for Integrating Biology and the Bedside) challenge was established the first English language clinical corpora annotated with temporal information (Sun et al., 2013a). The previous tasks of this challenge, which has been organized by the US National Center for Biomedical Informatics since 2007, were focused on automatic processing of clinical texts and problem solving (such as the deidentification of protected health information, extraction of medically relevant concepts and the relationships that exist between them, as well as identifying basic concept co-references).

The theme of the sixth i2b2 challenge, involving 18 teams from around the world, was the extraction of temporal information from clinical narrative texts such as the extraction of events (e.g. patient’s medical problems, tests, therapies), TEs and temporal relations. Considering the widespread use of the TimeML specification language for temporal annotation, which also served as the basis for the development of ISO-TimeML standard (ISO, 2009), a modified version of this guideline, currently known as THYME-TimeML (Temporal Histories of Your Medical Events), developed specifically for the clinical domain was used (Styler IV et al., 2014a,b). More recently, several studies that have been addressing the possibility of customizing the TimeML guidelines for temporal annotation of medical narratives (Savova et al., 2009; Galescu and Blaylock, 2012) and good clinical practice guide-

---

<sup>2</sup> Warfarin is an anticoagulant used to prevent heart attacks, strokes, and blood clots.

lines (Wenzina and Kaiser, 2014a,b) have achieved encouraging results. In addition to the TimeML adaptation for medical texts, some authors have also explored other options (Zhou et al., 2006; Tao et al., 2011).

The most successful i2b2 challenge systems in 2012 used already existing systems designed for recognition and normalization of TEs in newspaper texts, such as HeidelTime, SUTIME, GUTIME, NorMA systems, applying rule-based or hybrid approaches, combining hand-crafted rules and machine learning methods (Sun et al., 2013a). The analysis of the achieved results has shown that for most systems normalization of relative TEs is a more complex problem requiring further research.

Following the organized i2b2 challenge, the scientific community's interest in extracting clinical domain temporal information continues as part of the SemEval competition, held in 2015. As one of the tasks of the SemEval-2015 challenge, Clinical TempEval (Bethard et al., 2015) was a task aimed at identifying the extent and main features of TEs in clinical notes and pathology reports of the Mayo Clinic, Rochester, Minnesota. Clinical TempEval 2017 follows in the footsteps of the i2b2 2012 shared task, Clinical TempEval 2015, and Clinical TempEval 2016 in bringing timeline extraction to the clinical domain. As in past Clinical TempEvals, data were drawn from clinical notes and pathology reports for cancer patients at the Mayo Clinic. The key challenge for the Clinical TempEval 2017 task was the adaptation of the systems trained on data from colon cancer patients to make predictions on brain cancer patients.

Besides the mentioned international challenges, other systems have been developed to extract temporal information from medical narrative texts, primarily based on already existing general-purpose systems. For example, by modifying existing and creating new TARSQI rules, the Med-TTK system has been developed for use in the medical domain (Reeves et al., 2013). The normalization of recognized TEs is not covered by this system. Regarding normalization, the current problem is the normalization of the relative TEs of clinical texts (Sun et al., 2015).

For the most part, the research conducted in the field of automatic annotation of TEs of medical narrative texts relates to documents written in English. However, some efforts to adapt and develop systems for medical narratives in other languages, such as French (Hamon and Grabar, 2014), Chinese (Xiaoja et al., 2011), Swedish (Velupillai, 2014), Bulgarian (Angelova and Boytcheva, 2011; Boytcheva and Angelova, 2012; Boytcheva et al., 2012) have also been noted.

## 2.2 The aim of this paper

The automatic processing of temporal information of Serbian medical narrative texts has not been the subject of research so far. Considering the importance of time in interpreting medically relevant information, it would be useful to develop a system for the automatic processing of TEs of medical narrative texts, which would influence the development of other applications of automatic processing of medical texts, and consequently improve the quality of healthcare services, diagnostic and theoretical processes in medicine. Because medical domain narratives are stylistically very different from general natural language texts, natural language processing technologies developed for other domains cannot simply be applied to clinical domain texts, but require some modifications and adjustments.

Therefore, the aim of this paper is to apply the existing system for automatic processing of TEs in Serbian natural language texts to medical narrative texts, to evaluate the system’s efficiency in recognition and normalization of TEs and to determine the degree of necessary adaptation according to the characteristics and requirements of the medical domain. Some *ad-hoc* solutions and preliminary evaluation of the system’s performance conducted in this paper will serve as a basis for further research on the automatic processing of Serbian language in the medical field.

## 3 Recognition and normalization of TEs in the corpus of medical narratives

The automatic processing of TEs involves two stages. The first phase - recognition of TEs - concerns the identification of those fragments of text that carry temporal meaning and which represent the full range of TEs present in the texts, as well as the determination of their type. The second phase of TEs annotation is the normalization of their values, expressed explicitly or implicitly in the text.

Similar to TEs of standard Serbian, recognizing the TEs of medical narrative texts involves identifying and determining the range of linguistic expressions of absolute (e.g. *27. juna 2003. godine* ‘June 27th 2003’, *15:35 časova* ‘15:35 o’clock’, *proleća 2015.* ‘spring 2015’, *tri nedelje* ‘three weeks’) or relative (e.g. *16. maja* ‘May 16th’, *sutra uveče* ‘tomorrow evening’, *sledećeg meseca* ‘next month’, *nekoliko godina* ‘a few years’) temporal meaning, represented in medical texts by various formal units. Based on the TimeML guideline, the basic semantic types of TEs that need to be recognized are

those that indicate a temporal location in the form of calendar date (DATE) or time of a day (TIME), duration (DURATION), and frequency (SET). The TEs' semantic classes QUANTIFIER and PREPOSTEXP, as suggested by the THYME annotation guideline, will not be taken into account at this time in identifying TEs because it is about applying an existing system rather than upgrading it.

The process of TEs' normalization refers to the interpretation of the recognized TE values, in a standardized form that complies with the ISO 8601 standard. All identified TEs are marked by inserting a <TIMEX3> tag with the following defined attributes:

- the attribute **type**, specifying the semantic class of the recognized expression (e.g. DATE, TIME, DURATION, SET, DATE.PERIOD);
- the attribute **temporalFunction**, indicating the absolute or relative value of the identified TE (e.g. false, true);
- the attribute **value**, representing the finite values of TEs in a standardized form (e.g. 2007-01-03T19:30, XXXX-12-23, P4M);
- the attribute **mod**, used to represent the meaning of quantified or modified TEs (e.g. START, MID, END, LESS\_THAN, APPROX);
- the attribute **valueFromFunction**, used in the case of relative TEs, indicating the function that must be performed to calculate the absolute value of the identified expression (e.g. -1W, +1D);
- the attributes **quant** and **freq**, used to complete information about the meaning of TEs that indicate the frequency at which TE regularly reoccurs (e.g. **quant**="EVERY", **freq**="6X").

Several examples below illustrate the required semantic classes of TEs, as well as how to tag the results (Example 1).

*Example 1.*

(a)

23.12.2011. 'December 23, 2011'

```
<TIMEX3 type="DATE" temporalFunction="false" val="2011-12-23">
23.12.2011.</TIMEX3>
```

*martu prošle godine* 'March last year'

```
<TIMEX3 type="DATE" temporalFunction="true" val="XXXX-03"
valueFromFunction="-1Y">martu prošle godine</TIMEX3>
```

(b)

07.10.2012. *godine oko 16 h* 'October 7, 2012 around 4 p.m.'

```
<TIMEX3 type="TIME" temporalFunction="false"
val="2012-10-07T16"
mod="APPROX">07.10.2012. godine oko 16 h</TIMEX3>
```

*ujutru* 'the morning'

```
<TIMEX3 type="TIME" temporalFunction="true"
val="TMO">ujutru</TIMEX3>
```

(c)

*više od godinu dana* 'more than a year'

```
<TIMEX3 type="DURATION" temporalFunction="false" val="P1Y"
mod="MORE_THAN">više od godinu dana</TIMEX3>
```

*više godina* 'more years'

```
<TIMEX3 type="DURATION" temporalFunction="true" val="PXY">
više godina</TIMEX3>
```

(d)

*6 puta nedeljno* 'six times a week'

```
<TIMEX3 type="SET" temporalFunction="false" val="P1W"
freq="6X">6 puta nedeljno</TIMEX3>
```

*svaki dan* 'every day'

```
<TIMEX3 type="SET" temporalFunction="false" val="P1D"
quant="EVERY">svaki dan</TIMEX3>
```

(e)

*od 29.09.2009. do 03.10.2009.* 'from 29.09.2009. to 03.10.2009.'

```
OD <TIMEX3 type="DATE.PERIOD">
<TIMEX3 type="DATE" temporalFunction="false"
val="2009-09-29">29.09.2009.</TIMEX3>
DO <TIMEX3 type="DATE" temporalFunction="false"
val="2009-10-03">03.10.2009.</TIMEX3>
</TIMEX3>
```

*10-14 dana* '10-14 days'

```
<TIMEX3 type="DURATION.PERIOD">
<TIMEX3 type="DURATION" temporalFunction="false"
val="P10D">10</TIMEX3>
```

-  
<TIMEX3 type='DURATION' temporalFunction='false' val='P14D'>14  
dana</TIMEX3>  
</TIMEX3>

### 3.1 Finite-State methodology

The existing system for the automatic processing of TEs of Serbian newspaper texts is applied to medical narrative texts to identify and normalize the TEs contained in this document type. This system is a part of the Serbian system for named entity recognition (Krstev et al., 2014). It is a rule-based system that relies on lexical resources and handles both absolute and relative time. Its role is recognizing temporal expressions in unstructured texts and re-interpreting their temporal semantics in a standard format, according to the TimeML annotation guidelines, specified in (ISO, 2009).

Resources for natural language processing of Serbian are being developed using the finite-state methodology as introduced by Maurice Gross and LADL (Laboratoire d'Automatique Documentaire et Linguistique) laboratory (Gross, 1993). For the development and application of these resources the Unitex corpus processing system<sup>3</sup> was used (Paumier, 2016). The processing of TEs is carried out on a text having undergone sentence segmentation and part-of-speech tagging and morphological analysis. On the input text, general-purpose lexical resources (electronic dictionaries and dictionary finite-state transducers (FSTs)) are applied to tag text with lemmas, grammatical categories, and semantic features. After a text is being tagged this way, the system for TE recognition is applied.

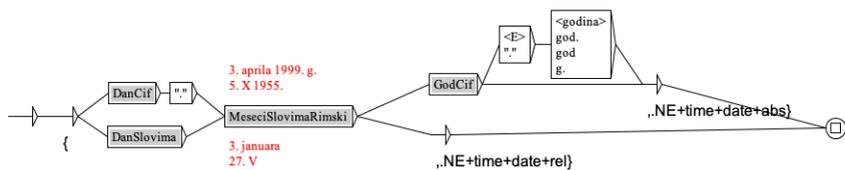
The Serbian language temporal recognition system is based on the transducer cascade - CasSys (Friburger, 2002; Friburger and Maurel, 2004), which is integrated into the Unitex system. Cascade is a simple and effective way of organizing FSTs that may greatly increase the precision and speed of the system, as well as containment of ambiguity. The cascade for TE recognition currently consists of 16 transducers, whose role is to identify the expression, as well as to determine the range and type of each expression detected, following the TimeML scheme (DATE, TIME, DURATION and SET). The created graphs of this extensive grammar in the form of a series of transducers are designed to identify expressions that indicate calendar dates, times of

---

<sup>3</sup> <https://unitexgramlab.org>

day, duration and frequency of reoccurring times. The entire process (from the pre-processing of the text, through the creation of recognition rules, to the extraction of temporal information) is performed through the Unitex graphical user interface and the working environment.

Each transduction is defined by a set of patterns. For the most frequent alternative forms of dates and times represented in Serbian, corresponding FSTs were built and applied to text to recognize patterns described in the input alphabet. When the pattern was matched, the output alphabet specified the action to be taken. For instance, FST *Datum* in Figure 1 recognizes some possible date patterns that consist of a day (written using digits or letters) followed by a month (written in letters or Roman numerals) followed by a year (written using digits), as well as incomplete date expressions in which year is omitted. The output contains the TE described in the input and with the addition of a lexical tag that can be used in a subsequent FSTs. Semantic markers associated with recognized expressions provide useful information primarily regarding the type of the named entity (+time), as well as temporal expression (+date, +hour,<sup>4</sup> +duration, +set), as given in Example 2. Additional information concerning the type of TE is provided by semantic markers +abs and +rel used to indicate absolute and relative TEs, respectively. Lexical tags produced by FSTs, even though the most convenient for the use of subsequent FSTs in the cascade, are not useful for other applications and at the end are converted to the XML tags (Example 2).



**Figure 1.** One path from FST *Datum* that recognizes complete and incomplete date expressions

<sup>4</sup> Corresponds to TimeML type TIME.

Example 2.

23.12.2011.

```
{23.12.2011.,.NE+time+date+abs}
```

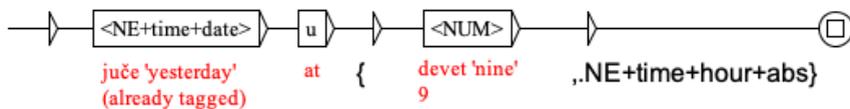
```
<time.date.abs>23.12.2011.</time.date.abs>
```

3. januara 'January 3rd'

```
{3. januara,.NE+time+date+rel}
```

```
<time.date.rel>3. januara</time.date.rel>
```

Applied to a text in a predefined order, FSTs first track down patterns of dates, times, and durations that can be retrieved with a high degree of certainty, while the retrieval of more complex TEs is postponed (e.g. temporal ranges, conjoined expressions, combinations of calendar dates and times-of-day). Furthermore, there are a lot of appearances of numerals that do not necessarily have to refer to the time of a day, and the cascade provides the right context for disambiguation. For instance, numerals that occur together with some already tagged dates could be reliable indicators of some time patterns after which words *čas* or *sat* 'hour' do not appear (Figure 2).



**Figure 2.** A simplified path in an FST that a numeral occurring after already marked date tags as time

The system responsible for the normalization of recognized TEs was also developed using the finite state methodology. As well as for TEs recognition, the Unitex software tool was used to create a collection of FSTs describing normalization rules and their application. Unlike the TEs recognition in which rules given by finite transducers are performed in a cascade order, whereby one transducer uses the results of previously applied ones, in the phase of TEs normalization rules are applied in the form of the FSTs local grammar to interpret the value of recognized expressions. Since the interpretation of TE value depends directly on its type, the local grammar aimed for normalization consists of four major transducers, each corresponding to

existing TE semantic classes. For instance, the FST given in Figure 3 illustrates the normalization of commonly used forms of relative calendar dates. The complete methodology used for recognition and normalization of TEs

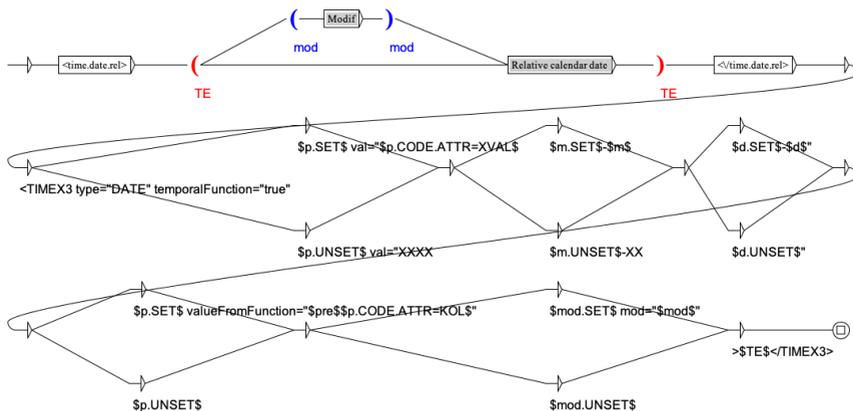


Figure 3. Transducer *Date relative* output

in Serbian natural language texts is described in detail in (Jaćimović, 2013, 2016).

### 3.2 Serbian medical narrative corpus

To evaluate the performance of the system in the automatic processing of TEs of medical narrative texts, a corpus of 150 randomly selected documents was used, consisting of discharge lists (n=100) and physician reports (n=50) of two teaching units of the School of Dental Medicine, University of Belgrade. The texts used for evaluation were previously automatically de-identified (Jaćimović et al., 2015), replacing personal and location names, as well as other health protected information with fictitious data, while TEs were not changed.<sup>5</sup> The selected texts were not used in the development

<sup>5</sup> Since these texts represent material that has not been fully de-identified, the medical records have been used with the consent of the School of Dental Medicine, University of Belgrade.

of the system and present entirely new material, suitable for preliminary evaluation.

Discharge lists and reports are unstructured, natural-language texts typed by physicians at the conclusion of hospital treatment or after multiple patient visits, and contain patient medical histories, current physical status, prescribed medication, laboratory test results, diagnostic findings, recommendations on the discharge of patients and other information relating to the patient's health. The dimensions of the corpus used, measured by the number of words and sentences, are presented in Table 1. Sentence segmenta-

Corpus	Number of words	Number of sentences
Discharge lists	23.175	2.167
Physician reports	13.373	1.219
Total	36.529	3.386

**Table 1.** The dimensions of the corpus of medical narrative texts used

tion was performed automatically (as described in 3.1), so the given numbers are considered only as approximate values, due to possible mistakes made during the text segmentation. The corpus used to evaluate the system, like all other medical narrative texts, is characterized by unfinished sentences, lack of punctuation marks, and an unusually large number of spelling and typographical errors, much larger than, for example, in the newspaper texts. Given the circumstances in which these texts originate, it is quite expected that many errors will occur, as one of the essential characteristics of medical narratives. Unlike newspaper texts written in standard Serbian, in which only 3-4% of words are unknown after applying electronic dictionaries, medical narrative texts contain slightly more than 20% of unrecognized words. This is understandable since these are specific domain texts characterized by terminology not yet included in the standard Serbian electronic dictionaries.

## 4 Implementation results and system performance evaluation

The Serbian system for automatic processing of TEs has been evaluated on the corpus of medical narratives previously described in Subsection 3.2.

Since the application of any automated information retrieval and extraction system results in some errors, it is necessary to determine their extent and type before drawing conclusions and conducting further research. Possible error types of an automated temporal processing system can occur concerning both recognition and normalization of the recognized TEs’ values. First of all, in the recognition phase, errors occur in the form of missing slots, which are not and should have been recognized as a TE; then, incorrect slots, which represent expressions incorrectly identified as a TE; lastly, errors in the form of a wrongly defined range or type of TE. During the normalization of TE values, it is possible to misdefine or omit the attribute values used for the interpretation of recognized TE values. Source of errors may be the result of errors in the text itself or omitted or insufficiently defined rules of the grammar used. The extensive list of error types and their examples will be given in Subsection 4.1.

Automatically recognized and normalized values are controlled by adding the `proveraP` ‘checkP’ and `proveraN` ‘checkN’ attributes (explained in detail below) within the `<TIMEX3>` tag to define the type and source of the error (Table 2). All unrecognized expressions are assigned `<TIMEX3>` tags and a `proveraP` attribute whose value is `MISS`. The examination was performed by

Value	Meaning
OK	Correctly recognized/normalized
UOK	The TE range or some of the attributes are omitted or not correctly defined
UOK/E	The TE range or some of the attributes are not correctly defined or missed due to a text error
NOK	All attributes incorrectly defined ( <code>NOKp</code> – recognition, <code>NOKn</code> – normalization)
MISS	Missed TE
MISS/E	Missed TE due to a text error

**Table 2.** The general values of the attributes used for evaluation (`proveraP` and `proveraN`)

one annotator (with experience in evaluating the named entity extraction system) based on a comprehensive instruction manual, with the control of the applied system’s author. For all correctly recognized or normalized TEs,

the value OK is assigned to the attributes `proveraP` and `proveraN` (Example 3).

*Example 3.*

(a)

*1978 GODINE*

```
<TIMEX3 proveraP='OK' proveraN='OK' type='DATE'  
temporalFunction='false' val='1978'>1978 GODINE</TIMEX3>
```

*prošlog meseca* 'last month'

```
<TIMEX3 proveraP='OK' proveraN='OK' type='DATE'  
temporalFunction='true' val='XXX-XX' valueFromFunction='-1M'>  
prošlog meseca</TIMEX3>
```

(b)

*oko 18 h* 'around 6 p.m.'

```
<TIMEX3 proveraP='OK' proveraN='OK' type='TIME'  
temporalFunction='false' val='T18' mod='APPROX'>oko 18  
h</TIMEX3>
```

(c)

*više od tri dana* 'more than three days'

```
<TIMEX3 proveraP='OK' proveraN='OK' type='DURATION'  
temporalFunction='false' val='P3D' mod='MORE_THAN'>više od  
tri dana</TIMEX3>
```

(d)

*svake godine* 'every year'

```
<TIMEX3 proveraP='OK' proveraN='OK' type='SET'  
temporalFunction='false' val='P1Y' quant='EVERY'>svake  
godine</TIMEX3>
```

In situations where none of the attributes assigned at the recognition or normalization stage are correctly defined, the value NOK is assigned to the attributes used for evaluation. Given the structure of the system, which does not allow normalization of values of temporal expressions that were not previously recognized, missed expressions due to a text error can only occur with the recognition of TEs when MISS/E value is assigned to the attribute `proveraP`.

For incorrectly defined or omitted values of the <TIME3> tag, value UOK is assigned to the attribute used for evaluation (**proveraP** or **proveraN**), with an indication mark which attribute is incorrectly specified. A detailed view of the <TIME3> attribute marks used for evaluation purposes is illustrated in the Table 3. For instance, if the TE type is incorrect, the recognition

Attribute	Mark
type	t
range	o
value	v
mod	m
valueFromFunction/function	f
quant	q
freq/frequency	u

Table 3. <TIME3> attribute marks

attribute **proveraP** will be assigned the value UOKt, while in the case of the misdefined **mod** and **valueFromFunction** attributes, the **proveraN** attribute will be assigned a UOKmf value. If an incorrectly defined range or a type of recognized expression is the result of a text error, the **proveraP** attribute will be assigned UOKo/E, or UOKt/E value, respectively.

The possible attribute values for **proveraP** are as follows:

- OK – correctly determined type and precisely defined full range of TE;
- UOK – some of the attributes are incorrect;
  - UOKt (the expression type is not correctly determined, but the full range is accurately defined);
  - UOKt/E (the expression type is not correctly determined due to a text error, but the full range is accurately defined);
  - UOKo (the type of expression is accurately specified, but the full range is not correctly defined);
  - UOKo/E (the expression type is accurately specified but the full range is not correctly defined due to a text error);
- NOKp – incorrectly defined both the type and range of the TE;
- NOK – an expression that is incorrectly defined and tagged as a TE (the recognized expression is not TE);

- MISS – missed expression (not recognized, although it should have been);
- MISS/E – the expression was not recognized due to a text error.

The possible values for the `proveraN` attribute are the following (used `<TIMEX3>` attribute marks are explained in Table 3):

- OK – normalization is done correctly;
- UOK – some of the attributes is incorrect;
  - UOKv (attribute `value` is incorrect);
  - UOKm (attribute `mod` is incorrect);
  - UOKf (attribute `valueFromFunction` is incorrect);
  - UOKmf (attributes `mod` and `valueFromFunction` are incorrect);
  - UOKvf (attributes `value` and `valueFromFunction` are incorrect);
  - UOKvm (attributes `value` and `mod` are incorrect);
  - UOKq (attribute `quant` is incorrect);
  - UOKu (attribute `freq` is incorrect);
  - UOKvq (attributes `value` and `quant` are incorrect);
  - UOKvu (attributes `value` and `freq` are incorrect);
- NOKn – all normalization attributes are incorrectly defined.

The results of evaluation show that out of the total number of existing TEs (884), 772 (87.3%) TEs indicating the date, time, duration and frequency occur in the corpus of medical narrative texts, and 112 (12.7%) expressions that indicate periods (Table 4). In this corpus, 613 (69.3%) abso-

<b>TIMEX3</b>	<b>abs</b>	<b>rel</b>	<b>Total</b>
DATE	402	40	442
TIME	43	37	80
DURATION	168	22	190
SET			60
Total	613	99	772
DATE.PERIOD			109
DURATION.PERIOD			3
Total	613	99	884

**Table 4.** The proportion of the TE types in the corpus of medical narrative texts

lute TEs were found that convey the meaning of a point in time and duration

(DATE, TIME, and DURATION), while 99 (11.2%) belong to the group of relative expressions. Regarding the expressions indicating the frequency of occurrences in time, a total of 60 were identified (6.8%), which is expected for the narrative texts of the medical domain.

During the verification of TEs defined by the TimeML guideline, identification and tagging of expressions characteristic of medical narrative texts (indicating a specific period associated with an event, based on the THYME instruction) were manually performed for later research. Thus, in the corpus used, 26 terms were found belonging to the newly defined PREPOSTEXP class (Example 4).

*Example 4.*

(a)

*Postoperativno sprovedena RT.* 'Postoperatively conducted RT.'  
<TIMEX3 type='PREPOSTEXP'>Postoperativno</TIMEX3> sprovedena  
RT.

(b)

*Rezultat trombektomije dobar, ali trećeg p.op dana dolazi do ponovne tromboze.*

'Thrombectomy result is good, but on the third p.op day thrombosis recurs.'

Rezultat trombektomije dobar, ali <TIMEX3 type='PREPOSTEXP'  
val='POD8'>trećeg p.op dana</TIMEX3> dolazi do ponovne  
tromboze.

(c)

*... otpušten je iz bolnice 10. dana kao oporavljen ...*

'... discharged from the hospital on 10<sup>th</sup> day as recovered ...'

... otpušten je iz bolnice <TIMEX3 type='PREPOSTEXP'  
val='HD10'>10. dana</TIMEX3> kao oporavljen ...

(d)

*Međutim intraoperativno su nađeni nepovoljni uslovi ...*

'However adverse conditions were found intraoperatively ...'

Međutim <TIMEX3 type='PREPOSTEXP'>intraoperativno</TIMEX3> su  
nađeni nepovoljni uslovi ...

(e)

*Preoperativno korigovana kardiološka terapija.*

'Preoperatively corrected cardiac therapy.'

<TIMEX3 type="PREPOSTEXP">Preoperativno</TIMEX3> korigovana kardiološka terapija.

The evaluation of the system's success in recognizing and normalizing the TEs of Serbian medical narrative texts was performed based on standard measures for assessing the performance of the information retrieval and extraction system. The basic evaluation unit was a complete TE. The precision ( $p$ ) of the system was observed, representing the ratio of the total number of correctly recognized TEs (OK) to the total number of recognized expressions (M) (Formula 1);

$$p = \frac{OK}{M} = \frac{OK}{OK + UOK + NOK} \quad (1)$$

next, response ( $r$ ), as the ratio of the total number of correctly recognized TEs (OK) to the total number of existing expressions in the text (N) (Formula 2), and  $F_1$  measure, as a common measure that reflects the harmonic mean of response and precision and always has a value that weighs less than the value of the response or precision achieved (Formula 3).

$$r = \frac{OK}{N} = \frac{OK}{OK + MISS} \quad (2)$$

$$F_1 = 2 \cdot \frac{pr}{p + r} \quad (3)$$

The evaluation used a balanced  $F_1$  measure, which favors neither response nor precision.

Regarding performance in recognizing TEs (determining range and type), the system achieved slightly higher accuracy (94%) than response (90%), with an overall  $F_1$  measure of 92% (Table 5). The best results are obtained when identifying and determining the range and type of expressions that indicate a period or represent a sequence of multiple TEs that are in a particular relation. Errors made in this process are solely the result of typing errors in the text, or inconsistent application of spelling rules.

The system achieved the highest precision in determining the range and type of expressions indicating frequency, while at the same time the response in identifying this class of expressions was the lowest, with an achieved  $F_1$  measure of 67%. A much lower response rate (86%) than the achieved accuracy (96%) also occurs in expressions that indicate duration, which, given the source of errors, may be explained by omissions in defining rules for identification of TEs. Regarding expressions with the meaning of a point in time

	N	M	OK	UOK <sub>t</sub>	UOK <sub>o</sub>	UOK <sub>o</sub> /E	MISS	MISS/E	NOK	p	r	F <sub>1</sub>
DATE	442	419	393	0	17	9	14	7	0	0.94	0.95	0.94
TIME	80	80	67	1	5	5	2	0	2	0.84	0.97	0.90
DURATION	190	165	158	3	2	2	21	4	0	0.96	0.86	0.91
SET	60	30	30	0	0	0	28	2	0	1	0.5	0.67
<b>Total</b>	<b>772</b>	<b>694</b>	<b>648</b>	<b>4</b>	<b>24</b>	<b>16</b>	<b>65</b>	<b>13</b>	<b>2</b>	<b>0.93</b>	<b>0.89</b>	<b>0.91</b>
DATE.PERIOD	109	105	104	0	0	1	0	4	0	0.99	0.96	0.98
DURATION.PERIOD	3	3	3	0	0	0	0	0	0	1	1	1
<b>Total</b>	<b>884</b>	<b>802</b>	<b>755</b>	<b>4</b>	<b>24</b>	<b>17</b>	<b>65</b>	<b>17</b>	<b>2</b>	<b>0.94</b>	<b>0.90</b>	<b>0.92</b>

**Table 5.** Data on the results of the evaluation and the achieved system performance in recognizing the TEs of medical narrative texts based on existing semantic expression classes

i.e. calendar dates and times of the day, a better response (95% and 97%, respectively) to the achieved accuracy (94 % and 84 %) clearly indicates the need to modify the existing rules, with the aim of correct tagging the identified TEs range. Detailed information on the performance of the system in recognizing TEs is given in the Table 5, while a detailed error analysis will be conducted below (4.1). Column N indicates the total number of TEs that exist in the corpus of medical texts, while column M contains the total number of TEs identified by the system.

The results achieved with respect to the success of the TEs normalization process are very good (Table 6). The values of all correctly recognized

	N	M	OK	UOK <sub>v</sub>	MISS	MISS/E	NOK	p	r	F <sub>1</sub>
DATE	419	419	419	0	0	0	0	1	1	1
TIME	80	80	78	2	0	0	0	0.98	1	0.99
DURATION	165	165	165	0	0	0	0	1	1	1
SET	30	30	30	0	0	0	0	1	1	1
<b>Total</b>	<b>694</b>	<b>694</b>	<b>692</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0.997</b>	<b>1</b>	<b>1</b>

**Table 6.** Data on evaluation results and system performance achieved in normalizing TEs of medical narrative texts based on existing semantic expression classes

TEs are accurately normalized (the achieved response is 100%), assigning all the appropriate attributes necessary for their interpretation. Since for the expressions indicating the time of day, due to a wrongly defined range, incor-

rect values of the attribute `value` were entered, the precision in performing normalization is 99.7%. Column N shows the total number of TEs identified by the system in the corpus of medical texts (for which the normalization process is expected to be performed), while column M contains the total number of TEs normalized by the system.

#### 4.1 Error analysis

Errors that occurred while applying the system, as well as its success in recognizing and normalizing TEs concerning existing semantic classes of TEs, are presented within Tables 5 and 6. The errors observed and the precision, response, and  $F_1$  measure achieved in determining the range and assigning the necessary attributes, without considering the class to which the TE belongs, are illustrated by the data in the Table 7. Column N shows

	N	M	OK	UOK	UOK <sub>o</sub> /E	MISS	MISS/E	NOK	$p$	$r$	$F_1$
range	884	802	761	24	17	65	17	0	0.95	0.90	0.93
type	802	802	796	4	0	0	0	2	0.99	1	1
temporalFunction	694	694	694	0	0	0	0	0	1	1	1
value	694	694	692	2	0	0	0	0	0.997	1	1
mod	46	46	46	0	0	0	0	0	1	1	1
valueFromFunction	11	11	11	0	0	0	0	0	1	1	1
quant	12	12	12	0	0	0	0	0	1	1	1
freq	4	4	4	0	0	0	0	0	1	1	1

**Table 7.** Data on evaluation results and system performance achieved in determining the range and attributes of TEs

the total number of TEs that exist in the corpus of medical texts for which range determination and attribute assignment were expected, while column M contains the total number of TEs whose range and attributes were determined by the system. Considering the achieved  $F_1$  measure, the biggest problem in the automatic processing of TEs, if not the only one, was identified already in the first step when recognizing and determining the range of expressions.

Expressions not identified and tagged by the system during processing make 9% of the total number of existing TEs in the corpus used (N). Almost a third of the missed expressions ( $n=28$ ) are those expressing the frequency of the prescribed therapies (npr. *na **drugi dan*** 'on the second day', *infuzijom 25000 j / **24 h*** 'infusion of 25000 U/24 hr', *0,8 ml na **12 h*** '0,8 ml in 12 hrs', etc.). Omissions related to expressions that indicate calendar dates

and durations are primarily the results of the application of rules, which, for the sake of greater system precision, did not allow the identification of numerical terms that may be extremely ambiguous or unusual in the context of newspaper texts (Example 5).

*Example 5.*

(a)

*Navodi da je 1995 a potom 96 operisala karcinom*

'She states that in 1995 and then in 96 she had cancer surgery'

*Navodi da je*

<TIMEX3 proveraP="MISS" type="DATE">1995</TIMEX3>

*a potom*

<TIMEX3 proveraP="MISS" type="DATE">96</TIMEX3> operisala  
karcinom

(b)

*U 7 god. operisao ...* 'In 7 yr. had surgery ...'

U <TIMEX3 proveraP="MISS" type="DURATION">7 god.</TIMEX3>  
operisao ...

(c)

*Pušač unazad oko 20 g.* 'Smoker for about 20 y.'

Pušač unazad <TIMEX3 proveraP="MISS" type="DURATION">oko 20  
g.</TIMEX3>

Omissions in the identification of TEs resulting from typographical errors make almost 17% of the total number of missed expressions. Some cases are illustrated with the Example 6.

*Example 6.*

*emesec dana* 'one omonth'

<TIMEX3 proveraP="MISS/E" type="DURATION">emesec dana</TIMEX3>

<TIMEX3 proveraP="MISS/E" type="DATE">18.o6.2009.</TIMEX3>

<TIMEX2 proveraP="MISS/E" type="DATE">26 07 2013</TIMEX3>

*sptembru o.g.* 'sptember this yr.'

<TIMEX3 proveraP="MISS/E" type="DATE">sptembru o.g.</TIMEX3>

*svaakodneвно* 'daaily'

```
<TIMEX3 proveraP="MISS/E" type="SET">svaakodneвно</TIMEX3>
```

( *sati* '( hours'

```
<TIMEX3 proveraP="MISS/E" type="SET">( sati</TIMEX3>
```

An analysis of errors that indicate an incorrectly defined range of the identified TE shows that 58.5% of them are indeed system flaws (Example 7), while the remaining 41.5% are errors due to existing spelling and typographical errors in the text (Example 8).

*Example 7.*

(a)

*početkom septembra o.g.* 'early September th.yr.'

```
<TIMEX3 proveraP="UOKo" proveraN="OK" type="DATE"
temporalFunction="true" val="XXXX-09" mod="START">početkom
septembra</TIMEX3> o.g.
```

(b)

*...povređen 22.12.2011. godine oko 01h 30 min.*

'...injured 22.12.2011. around 01h 30 min.'

```
...povređen <TIMEX3 proveraP="UOKo" proveraN="OK"
type="TIME" temporalFunction="false" val="2011-12-22T01"
mod="APPROX">22.12.2011. godine oko 01h </TIMEX3> 30 min.
```

(c)

*godinu i po dana* 'year and a half'

```
<TIMEX3 proveraP="UOKo" proveraN="OK" type="DURATION"
temporalFunction="false" val="P1.5Y">godinu i po</TIMEX3> dana
```

*Example 8.*

(a)

*...po podnevnim časovima istog dana* 'in the after noon same day'

```
...po podnevnim časovima <TIMEX3 proveraP="UOKo" proveraN="OK"
type="DATE" temporalFunction="true" val="XXXX-XX-XX"
valueFromFunction="=1D">istog dana</TIMEX3>
```

(b)

*0'6.12.2012.* 'December 6th, 2012'

```
0'<TIMEX3 proveraP="UOKo" proveraN="OK" type="DATE"
temporalFunction="false" val="2012-12-06">6.12.2012.</TIMEX3>
```

Since TEs in the context of medical narrative texts are a special PHI category that is subject to secrecy, it is very significant to have the best possible response when identifying them, as well as to achieve high precision in range determination in order not to reveal some information. With this in mind, the results of the error analysis in determining the range of identified TEs show that in most cases the omissions made do not adversely affect the determination of the type of expression and the normalization of its value, which is extremely important for the de-identification process i.e. hiding actual dates and moving them for a randomly selected interval. For example, in determining the range of calendar dates, the most frequently omitted part is related to the element of the year expressed by the abbreviated form of the relative expression *o.g. (this year)* (Example 7.a). The identified part of the expression that has to be hidden *early September* is correctly defined as the relative calendar date, with correctly assigned values of the **value** and **mod** attributes.

Errors in determining the type of recognized TE account for only 0.7% of the total errors. In most cases, these are expressions that indicate the frequency of prescribed therapy, and they are recognized as expressions with a meaning of duration (Example 9).

*Example 9.*

```
... cefalosporinski preparat na 8 sati '...cephalosporin on 8 hours'
... cefalosporinski preparat na <TIME3 proveraP='UOKt'
proveraN='OK' type='DURATION' temporalFunction='false'
val='P8H'>8 sati</TIME3>
```

Regardless of the wrong expression type, the form and value of the **value** attribute correspond to the defined expression form that indicates the frequency. Since the expression implies a repetition rate every 8 hours, the missing item is the **freq** attribute, whose value would be **EVERY**. Because the TimeML guideline suggests the **freq** attribute use only if the repetition rate is explicitly stated, the omitted attribute was not treated as an error this time.

Detecting and correcting errors in the form of *intruders*, or expressions that are misidentified as TEs, is especially important in the processing of medical narrative texts, particularly if the misspelled expression is an expression indicating some medically relevant information. The error analysis after the application of the system revealed only two such incorrectly identified expressions (Example 10).

*Example 10.*

...koštani stepenik u predelu rama leve orbite na oko 7 sati  
 '...bone step in the region of the left orbit frame at **around 7 o'clock**'  
 ...koštani stepenik u predelu rama leve orbite na  
 <TIMEX3 proveraP="NOK" proveraN="OK" type="TIME"  
 temporalFunction="false" val="T07" mod="APPROX">oko 7  
 sati</TIMEX3>

For example, in the field of maxillofacial surgery to diagnose a zygomatic fracture, an expression is used, which by analogy with the clockwise describes and localizes the occurrence of a bone step on the orbital frame (*at 5 o'clock position*).

Regarding other attributes defined by the TimeML guideline, the system made only two errors in determining the value of the most important attribute in the normalization process – value attribute. Both errors are the result of a wrongly defined range of TE (example 11), based on which the normalization of the value of the recognized expression was performed.

*Example 11.*

Detralex tbl 2 ujutro 'Detralex tbl 2 in the morning'  
 Detralex tbl <TIMEX3 proveraP="UOKo" proveraN="UOKv"  
 type="TIME" temporalFunction="false" val="T02">2  
 ujutro</TIMEX3>

The number that was mistakenly identified as the absolute time of day and included in the range of expression is an indication of the amount of therapy prescribed, which should be tagged as a QUANTIFIER type according to the THYME manual.

Based on the calculated system performance as a traditional  $F_1$  measure, the system error rate is calculated as  $E=1-F$ . However, since the importance of intruders and missed expressions has been minimized in the calculation of the harmonic response and precision (in that way the total error rate can never be greater than the error rate of any type of error separately), evaluation is also performed using the *Slot Error Rate* (SER). As a simple error measure, the SER equally weights different types of error directly, enabling the comparison of all systems against the fixed base. The SER is equal to the sum of all three types of errors (missed – MISS, intruders – NOK and non-matched – UOK) – divided by the total number of expressions existing in the reference corpus (N) (Formula 4).

$$SER = \frac{MISS + NOK + UOK}{N} = \frac{MISS + NOK + UOK}{OK + UOK + MISS} \quad (4)$$

Based on the data in Tables 8, 9 and 10, it is clear that the error rate calculated based on the *SER* is about 50% higher than the error rate presented by the  $F_1$  measure. The obtained *SER* values more emphasize the

	OK	UOK	MISS	NOK	$F_1$	<i>E</i>	<i>SER</i>
DATE	393	26	21	0	0.94	0.06	0.11
TIME	67	11	2	2	0.90	0.10	0.19
DURATION	158	7	25	0	0.91	0.09	0.17
SET	30	0	30	0	0.67	0.33	0.5
<b>Total</b>	648	44	78	2	0.91	0.09	<b>0.16</b>
DATE.PERIOD	104	1	4	0	0.98	0.02	0.05
DURATION.PERIOD	3	0	0	0	1	0	0
<b>Total</b>	755	45	82	2	0.92	0.08	<b>0.15</b>

**Table 8.** Values of the achieved  $F_1$  measure and *E* and *SER* error measures in recognizing the TEs of medical narrative texts based on existing semantic expression classes

	OK	UOK	MISS	NOK	$F_1$	<i>E</i>	<i>SER</i>
DATE	419	0	0	0	1	0	0
TIME	78	2	0	0	0.99	0.01	0.03
DURATION	165	0	0	0	1	0	0
SET	30	0	0	0	1	0	0
<b>Total</b>	692	2	0	0	1	0.001	<b>0.003</b>

**Table 9.** Values of the achieved  $F_1$  measure and *E* and *SER* error measures in normalizing the TEs of medical narrative texts based on existing semantic expression classes

system’s errors in the process of recognizing and normalizing the TEs of medical narrative texts.

Since medical narrative texts differ stylistically from general domain texts (Sun et al., 2013b) and that one-third of these texts consists of incomplete

	OK	UOK	UOK/E	MISS	MISS/E	NOK	$F_1$	$E$	$SER$
range	761	24	17	65	17	0	0.93	0.07	<b>0.14</b>
type	796	4	0	0	0	2	1	0.004	<b>0.008</b>
temporal- Function	694	0	0	0	0	0	1	0	0
value	692	2	0	0	0	0	1	0.001	<b>0.003</b>
mod	46	0	0	0	0	0	1	0	0
value- From- Function	11	0	0	0	0	0	1	0	0
quant	12	0	0	0	0	0	1	0	0
freq	4	0	0	0	0	0	1	0	0

**Table 10.** Values of the achieved  $F_1$  measure and  $E$  and  $SER$  error measures in determining the range and attributes of TEs

sentences that are difficult to identify with conventional language parsers (Savova et al., 2009), automatic processing of medical narrative texts requires a slightly different approach. Given that in the corpus of medical texts used, 26.4% of the total number of errors occur due to inconsistent application of grammatical and spelling rules, as well as the existence of more typographical errors, the rules used to recognize TEs need to be modified in order to improve both the precision and response of the automatic processing process.

Semantic classes of TEs that occur in newspaper texts also exist in medical narrative texts, the most common being calendar dates and expressions indicating duration. However, unlike the general domain, the use of expressions that indicate frequency is more common within medical texts, notably the frequency of the use of medication therapies. In the corpus used, even 7.7% of expressions of this type were identified, which is an unusually high percentage compared to their representation in newspaper texts. Although a large number of expressions (for which recognition rules were created) were not found in the corpus of medical texts, a particular class of TEs (PRE-POSTEXP), specific only to medical domain texts, was identified.

## 5 Conclusion

The system for recognition and normalization of Serbian language TEs, developed on the examples from newspaper articles, has been applied to

medical domain texts to determine the success of its application in this field. The results of the evaluation indicate that the created system, without any prior preparation and adaptation of rules to this specific domain, performs automatic recognition and normalization of TEs of medical narrative texts of the Serbian language, with high precision (94%). Although the achieved precision in automatic processing of TEs is lower than the precision achieved in the general domain (99%) (Jaćimović, 2016), it indicates that the system can be successfully applied in this domain as well, by adjusting the recognition rules based on the analysis of errors identified by the evaluation. Regarding the achieved response, the system applied to medical narrative texts performs better results with respect to the general domain (90% and 80%, respectively), which is understandable given the nature of medical texts that, compared to newspaper texts, contain a limited number of TE forms.

The error analysis revealed some omissions, which were essentially the result of the absence of appropriate rules, or the application of the existing rules, not allowing the identification of some expressions due to the higher precision of the system. The modularity of the system allows us to easily refine the existing and incorporate new rules for identifying those forms that are characteristic of medical texts, which will affect the improvement of both precision and response. Regarding the normalization of recognized expressions, as in the case of newspaper texts, the identified errors indicate that this process depends entirely on the success of the recognition (correct determination of the TEs range and type) since the values of all correctly identified and tagged expressions were normalized accurately. Given that medical narrative texts belong to a group of texts characterized by the existence of a large number of spelling and typographical errors, it would be useful to implement the existing text correction system (Stanković et al., 2011) in the further work.

## References

- Adlassnig, Klaus-Peter, Carlo Combi, Amar K Das, Elpida T Keravnou and Giuseppe Pozzi. “Temporal representation and reasoning in medicine: research directions and challenges”. *Artificial intelligence in medicine* Vol. 38, no. 2 (2006): 101–113
- Angelova, Galia and Svetla Boytcheva. “Towards temporal segmentation of patient history in discharge letters”. In *Proceedings of the Second Workshop on Biomedical Natural Language Processing*, 2011, 49–54

- Augusto, Juan Carlos. "Temporal Reasoning for Decision Support in Medicine". *Artificial intelligence in medicine* Vol. 33, no. 1 (2005): 1–24
- Bethard, Steven, Leon Derczynski, Guergana Savova, Guergana Savova, James Pustejovsky et al. "Semeval-2015 task 6: Clinical tempeval". *Proc. SemEval* (2015)
- Boytcheva, Svetla and Galia Angelova. "A workbench for temporal event information extraction from patient records". In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, 48–58. Springer, 2012
- Boytcheva, Svetla, Galia Angelova and Ivelina Nikolova. "Automatic analysis of patient history episodes in Bulgarian hospital discharge letters". In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 77–81. Association for Computational Linguistics, 2012
- Combi, Carlo and Yuval Shahar. "Temporal reasoning and temporal data maintenance in medicine: issues and challenges". *Computers in biology and medicine* Vol. 27, no. 5 (1997): 353–368
- Denny, Joshua C, Josh F Peterson, Neesha N Choma, Hua Xu, Randolph A Miller et al. "Extracting timing and status descriptors for colonoscopy testing from electronic medical records". *Journal of the American Medical Informatics Association* Vol. 17, no. 4 (2010): 383–388
- Friburger, N. and D. Maurel. "Finite-state transducer cascades to extract named entities in texts". *Theoretical Computer Science* Vol. 313, no. 1 (2004): 93–104
- Friburger, Nathalie. "Reconnaissance automatique des noms propres: application à la classification automatique de textes journalistiques". PhD. thesis, Tours, 2002
- Friedman, Carol, James J Cimino and Stephen B Johnson. "A conceptual model for clinical radiology reports.". In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 829. American Medical Informatics Association, 1993
- Galescu, Lucian and Nate Blaylock. "A corpus of clinical narratives annotated with temporal information". In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 715–720. ACM, 2012
- Goodwin, Scott D and Howard J Hamilton. "It's about time: an introduction to the special issue on temporal representation and reasoning". *Computational Intelligence* Vol. 12, no. 3 (1996): 357–358
- Gross, Maurice. "Local grammars and their representation by finite automata". *Data, Description, Discourse. Papers on the English Language in honour of John McH Sinclair* (1993): 26–38

- Hamon, Thierry and Natalia Grabar. “Tuning HeidelTime for identifying time expressions in clinical texts in English and French”. *EACL 2014* (2014): 101–105
- Hirschman, L. “Retrieving time information from natural language texts”. In *Information retrieval research*, Oddy, RN, SE Robertson, CJ Van Rijsbergen and Williams P, 154–171. Butterworths, London, 1981
- ISO. “ISO/DIS 24617-1 Language Resources Management - Semantic Annotation Framework (SemAF) - Part 1: Time and Events (SemAF-Time, ISO-TimeML)”, Standard, International Organization for Standardization. Geneva, Switzerland, 2009
- Jaćimović, Jelena. “Automatic Processing of Temporal Expressions in Serbian Natural Language Texts”. In *35th Anniversary of Computational Linguistics in Serbia*, 57–69, 2013
- Jaćimović, Jelena. “Automatsko prepoznavanje i normalizacija vremenskih izraza u nestrukturiranim novinskim i medicinskim tekstovima na srpskom jeziku”. PhD. thesis, Univerzitet u Beogradu, Filološki fakultet, Beograd, 2016
- Jaćimović, Jelena, Cvetana Krstev and Drago Jelovac. “A rule-based system for automatic de-identification of medical narrative texts”. *Informatica* Vol. 39, no. 1 (2015): 43–51
- Johnson, S. “Temporal information in medical narrative”. In *Medical Language Processing: Computer Management of Narrative Data*, Sager, N, C Friedman and MS Lyman, Reading, 175–94. MA: Addison-Wesley, 1987
- Keravnou, Elpida. “Medical temporal reasoning”. *Artificial Intelligence in Medicine* Vol. 3, no. 6 (1991): 289–290
- Krstev, Cvetana, Ivan Obradović, Miloš Utvić and Duško Vitas. “A system for Named Entity Recognition Based on Local Grammars”. *Journal of Logic and Computation* Vol. 24, no. 2 (2014): 473–489
- Lin, Ching-Heng, Nai-Yuan Wu, Wei-Shao Lai and Der-Ming Liou. “Comparison of a semi-automatic annotation tool and a natural language processing application for the generation of clinical statement entries”. *Journal of the American Medical Informatics Association* Vol. 22, no. 1 (2015): 132–142
- Liu, Mei, Min Jiang, Vivian K Kawai, Charles M Stein, Dan M Roden et al. “Modeling drug exposure data in electronic medical records: an application to warfarin”. In *AMIA annual symposium proceedings*, Vol. 2011, 815–823. American Medical Informatics Association, 2011
- Lyman, M, N Sager, C Friedman and E Chi. “Computer-structured narrative in ambulatory care: its use in longitudinal review of clinical data”. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 82–86. American Medical Informatics Association, 1985

- Meystre, Stéphane M, Guergana K Savova, Karin C Kipper-Schuler, John F Hurdle et al. “Extracting Information from Textual Documents in the Electronic Health Record: a Review of Recent Research”. *Yearb Med Inform* Vol. 35 (2008): 128–44
- Obermeier, Klaus K. “Temporal inferences in medical texts”. In *Proceedings of the 23rd annual meeting on Association for Computational Linguistics*, 9–17. Association for Computational Linguistics, 1985
- Paumier, Sébastien. *Unitex 3.1 User manual*, 2016, <http://igm.univ-mlv.fr/~unitex/UnitexManual3.1.pdf>
- Reeves, Ruth M, Ferdo R Ong, Michael E Matheny, Joshua C Denny, Dominik Aronsky et al. “Detecting temporal expressions in medical narratives”. *International journal of medical informatics* Vol. 82, no. 2 (2013): 118–127
- Sager, Naomi. “Syntactic analysis of natural language”, *Advances in computers* Vol. 8 (1967), no. 153–188: 35
- Savova, G, S Bethard, W Styler, J Martin, M Palmer et al.. “Towards temporal relation discovery from the clinical narrative”, In *AMIA... Annual Symposium proceedings/AMIA Symposium. AMIA Symposium*, 568–572, 2009
- Shahar, Yuval. “Timing Is Everything: Temporal Reasoning and Temporal Data Maintenance in Medicine”. In *Artificial Intelligence in Medicine*, Horn, Werner, Yuval Shahar, Greger Lindberg, Steen Andreassen and Jeremy Wyatt, *Lecture Notes in Computer Science*, Vol. 1620, 30–46. Springer Berlin Heidelberg, 1999
- Stanković, Ranka, Ivan Obradović, Cvetana Krstev and Duško Vitas. “Production of Morphological Dictionaries of Multi-Word Units Using a Multipurpose Tool”. In *Proceedings of the Computational Linguistics-Applications Conference*, Jassem, K, PW Fuglewicz, M Piasecki and A Przepiórkowski, Jachranka, 77–84. Poland: Polish Information Processing Society, 2011
- Styler IV, William F, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan et al. “Temporal annotation in the clinical domain”. *Transactions of the Association for Computational Linguistics* Vol. 2 (2014a): 143–154
- Styler IV, William F, Guergana Savova, Martha Palmer, James Pustejovsky, Tim O’Gorman et al. *THYME annotation guidelines*, 2014b
- Sun, Weiyi, Anna Rumshisky and Ozlem Uzuner. “Evaluating Temporal Relations in Clinical Text: 2012 i2b2 Challenge.”, *Journal of the American Medical Informatics Association* Vol. 20, no. 5 (2013a): 806–813

- Sun, Weiyi, Anna Rumshisky and Ozlem Uzuner. “Temporal Reasoning over Clinical Text: the State of the Art”, *Journal of the American Medical Informatics Association* Vol. 20, no. 5 (2013b): 814–819
- Sun, Weiyi, Anna Rumshisky and Ozlem Uzuner. “Normalization of relative and incomplete temporal expressions in clinical narratives”. *Journal of the American Medical Informatics Association* Vol. 22, no. 5 (2015): 1001–1008
- Tao, Cui, Harold R Solbrig and Christopher G Chute. “CNTRO 2.0: a harmonized semantic web ontology for temporal relation inferencing in clinical narratives”. *AMIA Summits Transl Sci Proc* Vol. 2011 (2011): 64–8
- Velupillai, Sumithra. “Temporal expressions in swedish medical text—a pilot study”. In *Proceedings of BioNLP, 2014*, 88–92
- Wenzina, Reinhardt and Katharina Kaiser. “Towards the Application of TimeML in Clinical Guidelines”. In *Modellierung im Gesundheitswesen. Tagungsband des Workshops im Rahmen der Modellierung 2014*, ICB-Research Report, 2014a, 37–48
- Wenzina, Reinhardt and Katharina Kaiser. “Using TimeML to support the modeling of computerized clinical guidelines.”. In *MIE, 2014b*, 8–12
- Xiaojia, Zhou, Li Haomin, Lu Xudong and Duan Huilong. “Temporal expression recognition and temporal relationship extraction from Chinese narrative medical records”. In *Bioinformatics and Biomedical Engineering, (iCBBE) 2011 5th International Conference on*, IEEE, 2011, 1–4
- Zhou, Li and George Hripcsak. “Temporal reasoning with medical data—a review with emphasis on medical natural language processing”. *Journal of biomedical informatics* Vol. 40, no. 2 (2007): 183–202
- Zhou, Li, Genevieve B Melton, Simon Parsons and George Hripcsak. “A Temporal Constraint Structure for Extracting Temporal Information from Clinical Narrative”. *Journal of biomedical informatics* Vol. 39, no. 4 (2006): 424–439