

УДК 004.896

**Ю. Н. Крапивный**, канд. физ.-мат. наук,  
**Г. В. Касаткина**

## ПРОБЛЕМЫ И МЕТОДЫ ИЗВЛЕЧЕНИЯ ТЕМПОРАЛЬНЫХ ЗНАНИЙ ИЗ ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

*Представлен анализ подходов автоматизированной интеллектуальной обработки текстов на естественном языке с целью извлечения информации о событиях и временных связях между ними. Описывается семантика основных понятий предметной области, этапы извлечения временной информации, выявляются наиболее перспективные направления исследований в этой области.*

**Ключевые слова:** извлечение знаний, обработка естественного языка, временные зависимости, классификация событий, нормализация временных указателей, временные аннотации, правила временного вывода, временные логики, онтологии событий, лексический вид события, категория вида, топологии времени, извлечение зависимостей

**Yu. N. Krapivny**, associate Professor,  
**G. V. Kasatkina**

## PROBLEMS AND METHODS OF TEMPORAL KNOWLEDGE EXTRACTION FROM NATURAL LANGUAGE TEXTS

*This survey analyzes various approaches to automatic natural language text processing and temporal and event data extraction as well as temporal relationship extraction. After the basic concepts are introduced, we describe different stages of information extraction and elaborate on the achieved results and the most promising areas of research.*

**Keywords:** knowledge extraction, temporal relationships, natural language processing, event categorization, timestamp normalization, temporal annotations, temporal reasoning, tense logics, event ontologies, lexical aspect, temporal topologies, relationship extraction

**Ю. М. Крапивний**, канд. фіз.-мат. наук,  
**Г. В. Касаткіна**

## ПРОБЛЕМИ І МЕТОДИ ВИБИРАННЯ ТЕМПОРАЛЬНИХ ЗНАНЬ ІЗ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ

*Представлено аналіз підходів автоматизованої інтелектуальної обробки текстів на природній мові з метою вибирання інформації про події та темпоральні зв'язки між ними. Описується семантика основних понять предметної області, етапи витягу темпоральної інформації, виявляються найбільш перспективні напрямки досліджень у цій області.*

**Ключові слова:** вибирання знань, обробка природної мови, темпоральні залежності, класифікація подій, нормалізація темпоральних маркерів, темпоральні анотації, правила темпорального виводу, темпоральні логіки, онтології подій, категорія виду, темпоральні топології, вибирання залежностей

### 1. Введение

Работа посвящена анализу существующих подходов автоматизированной интеллектуальной обработки текстов с целью извлечения, представления и обработки информации о временной связи событий, содержащейся в повествовательных текстах на естественном языке (ЕЯ) (примером таких текстов могут послужить новости, отчеты, истории болезней).

Актуальность работы обусловлена тем, что в настоящее время существует огромное количество текстов в электронном виде, которые представляют собой неисчерпаемый источник знаний о мире.

Однако обработать такой объем информации без помощи автоматизированных средств практически невозможно.

Система, которая могла бы выполнять поиск событий в определенных временных рамках или формировать на основе документа краткую схему основных событий в тексте, значительно облегчила бы извлечение и формализацию знаний в ЕЯ-текстах. Возможность создавать графы событий, связанных причинно-следственными или временными зависимостями, на основе группы документов позволила бы получить полную картину того или иного исторического события, его причин и последствий, или отразить общее положение дел на некоторый момент времени. Кроме того, в последнее время сис-

© Крапивный Ю.Н., Касаткина Г.В., 2013

темы извлечения временной информации используются для сбора статистики на основе отчетов: в медицине это истории болезни [1], в производственных отраслях – обработка отчетов о поломках [2]. Полученная статистика выглядит как цепочка событий или состояний и позволяет проследить, какие действия привели к поломке, как развивалось заболевание или какое лечение наиболее эффективно.

В русскоязычных изданиях вопросы формализации времени в основном рассматриваются как этап разработки систем автоматизированного управления ([3] и др.) и лишь несколько работ ([4] и др.) посвящено именно обработке текстов. В основном публикации по теме представлены в англоязычных изданиях начиная с 50-х годов прошлого столетия, когда вопрос извлечения и представления временной информации решали на стыке лингвистики и логики ([5] и др.). Начиная с 90-х годов прошлого столетия исследователи стали также применять статистические методы и методы машинного обучения ([6] и др.). На сегодняшний день существуют разработки (например, [14]), которые извлекают временные зависимости из текстов на ЕЯ и представляют их в формальном виде; также предложены некоторые алгоритмы для согласования этих зависимостей, базирующиеся на формальных теориях времени, таких, например, как интервальная теория времени [20].

**Цель настоящей работы** состоит в обозначении и анализе круга проблем, которые необходимо решить для успешного извлечения временных отношений из ЕЯ-текстов, представить различные способы формализации и решения этих проблем, описать некоторые существующие автоматизированные системы извлечения временных связей и выявить наиболее перспективные направления исследований в этой области.

Извлечение временных зависимостей в ЕЯ-текстах можно разбить на три этапа:

1) извлечение временных примитивов (сущностей): *событий* и *временных указателей*;

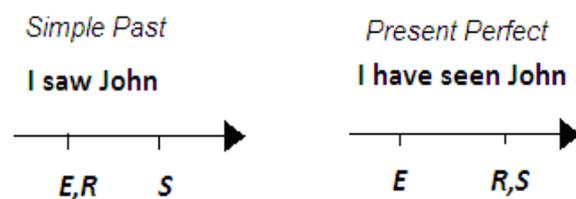
2) формализация полученной информации в рамках явной или неявной модели времени;

3) установление временных связей между извлеченными из текста сущностями.

## 2. Извлечение событий

### 2.1. Формализация грамматических категорий

Основными грамматическими категориями, выражающими время в ЕЯ, являются грамматическое **время (tense)** и **вид (aspect)** глагола. Первая категория позволяет расположить событие относительно момента речи (раньше, позже, одновременно); вторая уточняет, в какой временной категории рассматривается событие – как длящееся, совершенное, имеющее последствия и т.д. История попыток формализации грамматических значений начинается с Рейхенбаха (Reichenbach) [5], который выразил видовременные формы глаголов через соотношение трех точек на временной оси: **момента речи (S-speech time)**, **времени события (E-event time)** и **точки отсчета (R-reference point)** (см. рисунок).



Пример формализации грамматического времени по Рейхенбаху

Другой вариант формализации был предложен Артуром Прайором (Prior) [7], который построил на основе обычной пропозициональной логики **логику для выражения грамматического времени (tense logic)** – модальную логику, использующую, помимо классических, четыре дополнительных квантора:

- P* – в некоторый момент в прошлом;
- F* – в некоторый момент в будущем;
- G* – в любой момент в будущем;
- H* – в любой момент в прошлом.

Минимальная логика времени Прайора *Kt* опирается на систему аксиом таких, например, как:  $\varphi \rightarrow H F \varphi$  (для любого  $\varphi$ , существующего на данный момент, в прошлом всегда было верно утверждение: «В некоторый момент  $\varphi$  будет существовать»).

На основании приведенных операторов Прайор разработал четыре варианта временной логики разной степени выразительности. Среди недостатков этих логик – отсутствие оператора для обозначения настоящего времени, а также невозможность выразить грамматическую категорию вида: прогрессивные и перфектные формы глаголов.

## 2.2. Онтология событий

Хотя грамматические категории заключают в себе значительную часть временной информации, без учета лексических значений слов в предложении или тексте общая картина будет неполной и неверной.

Лексическую категорию, описывающую характер протекания ситуации, процесса, события и т.п. в реальном мире, называют **лексическим аспектом** (lexical aspect, Aktionsart) [8]. Если время глагола располагает его относительно времени речи, а вид глагола указывает точку зрения наблюдателя, то лексический аспект характеризует внутреннюю структуру описываемой ситуации. В зависимости от лексического аспекта, глаголы (а точнее, группы глаголов в предложениях) делят на несколько групп [9], две самые крупные из которых описывают **состояния (states)** и **события (events)**.

Глаголы, описывающие события, в свою очередь, делятся на **процессы** (processes, activities), **завершенные действия** (accomplishments), **достижения** (achievements) и **семельфактивы** (semelfactives).

Описание внутренней структуры, компонентов и временных характеристик разных классов событий называют **онтологией событий**. Попытки формализовать правила изменения класса глагола в зависимости от его аргументов, а также контекста привели к созданию **аспектуального исчисления** Даути (Dowty) [10], **алгебры событий** Баха (Bach) [11], **теории подструктуры событий** Пустейовского (Pustejovsky) [12] и другим теориям.

## 3. Извлечение временных отношений. Теории времени

Под **временными отношениями** (связями, зависимостями) будем понимать взаимное расположение двух или более событий во времени. Выявить в тексте временные зависимости между событиями и временными

указателями достаточно сложно, и главным затруднением является то, что временные связи зачастую не выражаются в тексте явно, а только подразумеваются. При автоматическом извлечении временных связей, как правило, анализируют **времена и виды глаголов, наречий, временные указатели, предлоги и союзы**, указывающие на временные зависимости.

Формализованное представление информации о времени чаще всего имеет вид временных зависимостей или отношений между временными примитивами. Состав и структура примитивов определяются заданной **онтологией времени**: это могут быть точки (моменты) и/или интервалы разных видов. Набор возможных зависимостей между примитивами также определяется онтологией; зависимости указывают на время появления, существования и окончания примитивов относительно друг друга. **Топология времени** задает область интерпретации примитивов и ее свойства (упорядоченность, ограниченность/неограниченность, дискретность/непрерывность и т.д.) и задается посредством множества аксиом. Совокупность онтологий и топологий времени образует **теорию времени**, которая характеризуется **выразительностью и полнотой системы рассуждений** [17].

Современные методы представления временных зависимостей можно разбить на два основных класса – основанные на моделировании изменений во времени и основанные на явном моделировании времени. В первом случае примитивами будут такие сущности как **действия и состояния**, описывающие изменения системы. Среди таких способов моделирования можно назвать, например, исчисление ситуаций МакКарти (McCarthy) [18] и исчисление событий Ковальского и Сергота (Kowalski, Sergot) [19]. Однако подобные методы налагают существенные ограничения на представление мгновенных событий и требуют полной определенности взаиморасположения любых двух событий/ситуаций.

Второй класс методов обладает большими выразительными возможностями и основывается на явном представлении времени с учетом его свойств. Примитивами считаются

**точки** либо **интервалы**. В основном это различные временные логики, где время учитывается либо путем усложнения интерпретации (модальные логики), либо расширения синтаксиса (модификации логики первого порядка).

Создаваемая модель времени должна удовлетворять интуитивным представлениям человека о течении времени [17]. Эти представления выражаются рядом **аксиом**, таких например, как

**иррефлексивности**

$$\forall i: \neg(i < i'), \quad (1)$$

**асимметричности**

$$\forall i, i': i < i' \Rightarrow \neg(i' < i), \quad (2)$$

**транзитивности:**

$$\forall i, i', i'': i < i' \wedge i' < i'' \Rightarrow i < i''. \quad (3)$$

Здесь символ “<” означает отношение предшествования,  $i, i', i'', i'''$  – моменты времени.

На основании этих аксиом можно задать **точечную теорию времени** – непрерывную или дискретную, с линейным или ветвящимся временем. Синтаксически полных теорий среди этого множества две. Обе строятся на основе линейной модели времени, но в первой время будет непрерывно, во второй – дискретно.

Однако в точечных теориях трудно выразить длительность события или состояния; поэтому появились теории на основе временных интервалов. Самую известную **интервальную теорию времени** предложил Аллен (Allen) [20]. Примитивами в этой теории считаются временные промежутки – интервалы – и набор из тринадцати базовых отношений между интервалами: равенство, предшествование, соприкосновение и др.

Интервальная теория Аллена  $\langle P, AR \rangle$  опирается на следующие **аксиомы**:

$$1) \forall p \in P, r \in AR: \exists p'(r(p, p')) \quad (4)$$

(для любого интервала и любого отношения из AR существует другой интервал, связанный с первым данным отношением);

$$2) \forall p, p' \in P, \forall r, r' \in AR: r(p, p') \Rightarrow \neg r'(p, p') \quad (5)$$

(отношения из AR являются взаимоисключающими);

3) таблицу транзитивности Аллена (композиции отношений).

Теорию Аллена можно расширить, определив моменты времени как точки соприкосновения интервалов (через отношение MEETS). Тогда модель времени приобретет вид  $\langle I, P, Limits, AR \rangle$  со следующими дополнительными аксиомами (символ  $\parallel$  обозначает отношение MEETS):

$$\forall p, p' \in P: p \parallel p' \Rightarrow \exists i(Limits(i, p) \wedge Limits(i, p')) \quad (6)$$

$$\forall i \in I, \exists p, p' \in P: p \parallel p' \wedge Limits(i, p) \wedge Limits(i, p').$$

И, наконец, третьим вариантом моделирования времени является представление времени не через временные, а **событийные примитивы**, т.е. состояния и действия. В этом случае теория времени имеет вид  $\langle E, <, O \rangle$ , где E – непустое множество событий, < – отношение предшествования, а O – отношение пересечения. Эта теория описывается набором аксиом, основные из них:

**несимметричности предшествования**

$$e < e' \Rightarrow \neg(e' < e), \quad (7)$$

**транзитивности предшествования**

$$e < e' \wedge e' < e'' \Rightarrow e < e'', \quad (8)$$

**симметричности пересечения**

$$e O e' \Rightarrow e' O e. \quad (9)$$

Теории времени, основанные на событиях, отличаются от точечных и концептуальных теорий, но имеют много общих черт. Фактически события являются аналогами точек либо интервалов. Основное различие состоит в том, что если две точки или два интервала, занимающие одно пространство на временной оси, являются идентичными, то для событий это неверно.

**4. Интеллектуальные системы анализа текстов ЕЯ**

**4.1. Извлечение событий**

Работа любой автоматизированной системы оперирования временными связями начинается с **выделения релевантных сегментов текста**, т.е. сегментов, которые содержат основную информацию о событиях, временных указателях и связях.

В 2010 г. исследовательская группа SIGLEX ([www.clres.com/siglex.html](http://www.clres.com/siglex.html)), являющаяся частью Ассоциации вычислительной лингвистики ([www.aclweb.org](http://www.aclweb.org)), провела конкурс SemEval на лучшую систему семантического анализа. Конкурс включал задание **TempEval-2** [13] на извлечение событий и временных указателей, а также

установление между ними временных связей. Учитывая то, что конкурс SemEval, насколько нам известно, является пока что единственным, включающим в себя отдельное задание по работе с временными зависимостями в ЕЯ, можно полагать, что работы, представленные на конкурс, дают достаточно полную картину достижений в этой области.

Из работ, представленных на конкурсе TempEval, только шесть выполнили задание по извлечению событий из текста: **Edinburgh-LTG** [14], **JU\_CSE\_TEMP** [24], **TIPSem** и **TIPSem-B** [23], **TRIPS** и **TRIOS** [22].

Система **TIPSem** отличается от системы **TIPSem-B** только тем, что использует механизм извлечения семантических ролей, тогда как **TIPSem-B** задействует модуль машинного обучения с учителем по методу CRF (Conditional Random Fields, [15]). Обучающей выборкой служил модифицированный корпус TimeBank, предоставляемый всем участникам конкурса. В системе **TIPSem** обучение проводилось на двух уровнях – синтаксическом и семантическом.

В системе **Edinburgh-LTG** для извлечения событий использовался набор заданных вручную правил, а также модуль машинного обучения, построенный на основе метода максимальной энтропии (C&C Maximum Entropy Tagger [16]).

Система **TRIPS** в ее исходном варианте работала с типами событий, отличными от описанных в стандарте TimeML. Система **TRIOS** представляет собой расширенную версию анализатора **TRIPS**, с дополнительным классификатором на основе логико-марковской сети.

#### 4.2. Извлечение и нормализация временных маркеров

Временными маркерами или указателями называют сегменты текста, которые явно указывают на то, когда имела место некоторая ситуация. Согласно спецификации TimeML [21] (ISO-стандарт разметки временных указателей в тексте), существует четыре типа временных указателей: **даты** (dates), **времена/временные метки** (times), **промежутки** (durations), **множества** (sets). Вычисление абсолютного значения относительных временных указателей необходимо

для того, чтобы наиболее точно локализовать во времени события, извлеченные из текста. Этот процесс называют нормализацией, суть которого состоит в том, чтобы привести указатели к некоторому стандартному виду (**нормализовать**). Так система автоматической обработки должна относить маркеры «Halloween 2012», «October 31, 2012», «9/31/12», «two days ago» (в тексте, написанном 2 ноября 2012 года) и им подобные к одному и тому же дню. Не все указатели позволяют точно определить момент или интервал, на который они ссылаются: в этом случае указатель остается недоопределенным.

На конкурс TempEval-2 задание по извлечению временных указателей из текстов на ЕЯ выполняли упомянутые ранее системы и системы **HeidelTime** [27], **TERSEO**, **USDF2** [26], **KUL** [28].

Лучший результат показала система **HeidelTime**, использующая только правила вывода, заданные вручную. Каждое правило состояло из регулярного выражения, функции нормализации для этого выражения и дополнительной информации о типе указателя, необходимой для нормализации. Нормализация также выполняется на основании правил; в качестве точки отсчета принимается либо время создания документа, либо последний полностью расшифрованный временной указатель.

Следующими по эффективности стали системы **TRIPS** и **TRIOS**. Обе эти системы использовали модуль машинного обучения по методу CRF с лексическими характеристиками, результат работы которого затем фильтровался с помощью правил, заданных вручную. Классификация выделенных указателей по типу и их последующая нормализация также выполнялись на основании правил.

Третье место разделили системы **TIPSem** и **Edinburgh-LTG**.

Для извлечения временных указателей в системе **TIPSem** применяется такой же алгоритм, как и для извлечения событий. Нормализация состояла из двух этапов: определения метода нормализации для каждого указателя (с помощью классификатора, ис-

пользующего метод CRF), а затем применения нужного набора правил.

Выделение временных указателей в тексте системой Edinburgh-LTG выполняется при помощи набора заданных вручную правил и классификатора, использующего для обучения метод максимальной энтропии; так же происходит и нормализация.

Кроме того, стоит отметить систему **TERSEO** [25], использующую для извлечения указателей набор заданных вручную правил. Эта система правильно определяла тип извлеченного указателя в 98 % случаев.

#### **4.3. Извлечение временных отношений.**

Конкурс TempEval включал в себя четыре задания (C, D, E и F) на установление временных зависимостей. Например, одно из заданий формулировалось так: упорядочить во времени события и указатели в пределах одного предложения (задание C).

Восемь из представленных систем выполнили описанные задания: **TIPSem** и **TIPSem-B**, **TRIPS** и **TRIOS**, **USFD2**, **JU\_CSE\_TEMP**, **NCSU** [29] (варианты *indi* и *joint*). Все системы использовали для разметки классификаторы, построенные на основе различных методов машинного обучения и учитывающие различные характеристики.

Так, системы **TIPSem** и **TIPSem-B** для всех четырех заданий использовали **классификаторы CRF**, которые отличались характеристиками, выбранными для обучения.

Системы **TRIPS** и **TRIOS** в отличие от других систем не используют для обучения характеристики разметки, созданной вручную для стандартного обучающего корпуса, но опираются только на свойства, извлеченные на предыдущих шагах анализа, а именно: класс события, время, вид, полярность, корень слова, слово, лексический аспект и т.п.

В системах **USFD2** применяется **метод максимальной энтропии** и дополняется список стандартных характеристик событий и указателей сигнальными словами, т.е. словами, которые с большой вероятностью указывают на наличие той или иной временной зависимости (*before*, *after*, *during*, *as soon as* и т.д.).

#### **5. Выводы**

В данном обзоре были рассмотрены этапы анализа текста на ЕЯ на предмет выявления временных связей, проблемы и результаты, полученные на данный момент при решении соответствующих задач.

Для выявления временных отношений из текста следует извлечь информацию об упоминаемых событиях, указатели на конкретные моменты или промежутки времени, а затем проанализировать отношения между ними.

Этап извлечения событий требует наличия продуманной онтологии событий, договоренностей по поводу того, какие сегменты текста несут наибольший объем информации о событии, а также некоторой системы формализации полученной информации. Затруднения на этом этапе связаны, по большей части, с невозможностью извлечь необходимые аспекты значения слов, не прибегая к глубокому семантическому анализу. Это мешает отнести извлеченные события к тому или иному классу онтологии, затрудняет определение их временных свойств, а также мешает выявлению событий, выраженных частями речи, отличными от глагола. Для решения этих задач используются различные методы машинного обучения с учителем (CRF, максимальной энтропии, логико-марковских сетей), а также семантические анализаторы. Максимальная точность таких методов на настоящий момент – около 80 %.

На этапе извлечения временных указателей, помимо задачи определения границ релевантных сегментов текста, возникает проблема приведения указателей различных типов к единому виду, т.е. извлечение нормализованного значения указателей. Первая из упомянутых задач решается теми же методами, что и в случае событий, с точностью до 85%. Наиболее эффективные методы нормализации указателей опираются на тщательно выверенные наборы правил и позволяют добиться точности до 85 %.

Этап установления временных отношений на данный момент является наименее успешным, отчасти потому, что зависит от результатов, полученных на предыдущих этапах. Первая задача, возникающая на данном этапе, состоит в том, чтобы выбрать

подходящую теорию формализации времени. Теория времени должна позволять удобно формализовать временные аспекты событий и временных указателей, а также описывать такой набор отношений, которым впоследствии можно было бы с легкостью оперировать.

Решение задачи построения временной структуры текста позволило бы обработать огромные объемы информации, которые на данный момент хранятся в виде текстов на ЕЯ – отчеты, истории болезни, инструкции, рапорты и т.д. Однако результаты в этой области пока что являются недостаточно точными для практического применения, что обуславливает актуальность исследований в области семантической обработки текстов на ЕЯ.

#### Список использованной литературы

1. Verhagen, M. The TARSQI Toolkit [Text] / M. Verhagen, J. Pustejovsky // Proceedings of the Eight International Conference on Language Resources and Evaluation, May 23-25, 2012, Istanbul/ ed. N. Calzolari, K. Choukri, T. Declerck, M.U. Dogan, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis. — Istanbul: ELRA, 2012. – P. 2043–2048.
2. Passonneau, R. J. A Computational Model of the Semantics of Tense and Aspect [Text] / R. J. Passonneau // Computational Linguistics. – 1988. – Vol. 14. – No 2. – P. 44–60.
3. Еремеев, А. П. Темпоральные модели и их применение в интеллектуальных системах [Текст] / А. П. Еремеев // Научная сессия МИФИ-2011: сб. науч. тр., 24-28 января 2011 г., Москва. – М.: МИФИ. – 2011. – P. 45–46.
4. Заболеева-Зотова, А. В. Гибридный подход к обработке временной информации в тексте на русском языке [Текст] / А. В. Заболеева-Зотова, Д. К. Фамхынг, С. С. Захаров // КИИ-2008: сборник трудов международной научно-технической конференции, 29 сентября – 3 октября 2008 г., Дубна: – М.: URSS. – 2008. – P. 228–235.
5. Reichenbach, H. The Tenses of Verbs [Text] / H. Reichenbach // Elements of Symbolic Logic. – New York : Macmillan, 1947. – P. 287–298.
6. Kolomiyets, O. Machine Learning Approaches for Temporal Information Extraction: A comparative study [Text] / O. Kolomiyets, M.-F. Moens // GI Jahrestagung: September 28 - October 3, 2009, Lubeck / ed. S. Fischer, E. Maehle, R. Reischuk. – GI, 2009. – P. 3150–3161.
7. Prior, A. N. Papers on Time and Tense [Text] / A. N. Prior. – Oxford University Press, 1968. – 342 p.
8. Binnick, R. I. Time and the Verb: A Guide to Tense and Aspect [Text] / R. I. Binnick. – Oxford University Press, USA. – 1991. – 584 p.
9. Smith, C. S. Modes of Discourse: The Local Structure of Texts [Text] / C. S. Smith. — Cambridge University Press, 2003. – 334 p.
10. Dowty, D. R. Word Meaning and Montague Grammar [Text] / D. R. Dowty. — Dordrecht: Reidel, 1979. – 448 p.
11. Bach, E. The Algebra of Events [Text] / E. Bach // Linguistics and Philosophy. – 1986. – Vol. 9. – P. 5–16.
12. Pustejovsky, J. The Syntax of Event Structure [Text] / J. Pustejovsky // Cognition. – 1991. – Vol. 41. – P. 47–81.
13. Verhagen, M. SemEval-2010 Task 13: TempEval-2 [Text] / M. Verhagen, R. Sauri, T. Caselli, J. Pustejovsky // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 57–62.
14. Grover, C. Edinburgh-LTG: TempEval-2 system description [Text] / C. Grover, R. Tobin, B. Alex, K. Byrne // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL. – 2010. – P. 333–336.
15. Lafferty, J. D. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [Text] / J. D. Lafferty, A. McCallum, F. C. N. Pereira // Proceedings of the 18th International Conference on Machine Learning, June 28 - July 1, 2001, Williamstown / ed. C. Brodley, A. Danilyuk. – Morgan Kaufmann. – 2001. – P. 282–289.
16. Curran, J. R. Language Independent NER Using a Maximum Entropy Tagger [Text] / J. R. Curran, S. Clark // Proceedings of the Seventh Conference on Natural Language

Learning HLT-NAACL, May 31, 2003, Edmonton, Canada. – ACL, 2003. – P. 164–167.

17. Fisher, M. Handbook of Temporal Reasoning in Artificial Intelligence (Foundations of Artificial Intelligence) [Text] / M. Fisher, D. Gabbay, L. Vila. - New York: Elsevier Science Inc., 2005. – 750 p.

18. McCarthy, J. Some Philosophical Problems From the Standpoint of Artificial Intelligence [Text] / J. McCarthy, P. Hayes // Machine Intelligence 4. – Edinburgh University Press. – 1969. – P. 463–502.

19. Kowalski, R. A Logic-based Calculus of Events [Text] / R. Kowalski, M. Sergot // New Generation Computing. – 1986. – Vol. 4. – No. 1. – P. 67–95.

20. Allen, J. F. Towards a General Theory of Action and Time [Text] / J. F. Allen // Artificial Intelligence. – 1984. – Vol. 23. – No. 2. – P. 123–154.

21. Pustejovsky, J. The Specification Language TimeML [Text] / J. Pustejovsky, B. Ingria, R. Sauri[et al.] // The Language of Time: A Reader / I. Mani, J. Pustejovsky, R. Gaizauskas. – Oxford University Press. – 2004. – P. 545–559.

22. UzZaman, N. TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text [Text] / N. UzZaman, J. F. Allen // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 276–283.

23. Llorens, H. TIPSem (English and Spanish): Evaluating CRFs and semantic roles in TempEval-2 [Text] / H. Llorens, E. Saquete, B. Navarro // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 284–291.

24. Kolya, A. K. JU\_CSE\_TEMP: A First Step towards Evaluating Events, Time Expressions and Temporal Relations [Text] / A. K. Kolya, A. Ekbal, S. Bandyopadhyay // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 345–350.

25. Saquete, E. ID 392: TERSEO + T2T3 Transducer: a systems for recognizing and normalizing TIMEX3 [Text] / E. Saquete // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 317–320.

26. Derczynski, L. USFD2: Annotating Temporal Expressions and TLINKs for TempEval-2 [Text] / L. Derczynski, R. Gaizauskas // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – Uppsala, Sweden: Association for Computational Linguistics, 2010. – P. 337–340.

27. Strotgen, J. HeidelTime: High quality rule-based extraction and normalization of temporal expressions [Text] / J. Strotgen, M. Gertz // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 321–324.

28. Kolomiyets, O. KUL: Recognition and normalization of temporal expressions [Text] / O. Kolomiyets, M.-F. Moens // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 325–328.

29. Ha, E. Y. NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology [Text] / E. Y. Ha, A. Baikadi, C. Licata, J. C. Lester // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 341–344.

Получено 19.11.2012

## References

1. Verhagen, M., Pustejovsky J. The TAR-SQI Toolkit [Text] / ed. N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis. // Proceedings of the Eight International Conference on Language Resources and Evaluation, May 23-25, 2012, Istanbul. – Istanbul : ELRA, 2012. – P. 2043–2048 [in English].

2. Passonneau, R. J. A Computational Model of the Semantics of Tense and Aspect [Text] / R. J. Passonneau // Computational Linguistics. – 1988. – Vol. 14. – No 2. – P. 44–60 [in English].

3. Yermeev, A. P. Temporalnye modeli i ih primeneniye v intellektualnyh sistemah [Text] / A. P. Yermeev // MIFI-2011: conference proceedings, January 24-28, 2011, Moscow. – Moscow : MIFI. – 2011. – P. 45–46 [in Russian].

4. Zaboloeva-Zotova, A. V. Gibridnyy podhod k obrabotke vremennoy informatsii v tekste na russkom yazyke [Text] / A. V. Zaboloeva-Zotova, D. K. Famhyng, S. S. Zakharov // KII-2008: conference proceedings, September, 29 – October, 3, 2008, Dubna. – Moscow : URSS, 2008. – P. 228–235 [in Russian].
5. Reichenbach, H. The Tenses of Verbs [Text] / H. Reichenbach // Elements of Symbolic Logic. – New York : Macmillan, 1947. – P. 287–298 [in English].
6. Kolomiyets, O. Machine Learning Approaches for Temporal Information Extraction: A comparative study [Text] / O. Kolomiyets, M.-F. Moens // GI Jahrestagung: September 28 - October 3, 2009, Lubeck / ed. S. Fischer, E. Maehle, R. Reischuk. – GI, 2009. – P. 3150–3161 [in English].
7. Prior, A. N. Papers on Time and Tense [Text] / A. Prior. – 1968. – Oxford University Press, – 342 p. [in English].
8. Binnick, R. I. Time and the Verb: A Guide to Tense and Aspect [Text] / R. I. Binnick. – 1991 – Oxford University Press, USA, – 584 p. [in English].
9. Smith, C. S. Modes of Discourse: The Local Structure of Texts [Text] / C. S. Smith. – 2003. – Cambridge University Press, – 334 p. [in English].
10. Dowty, D. R. Word Meaning and Montague Grammar [Text] / D. R. Dowty. – Dordrecht: Reidel, – 1979. – 448 p. [in English].
11. Bach, E. The Algebra of Events [Text] / E. Bach // Linguistics and Philosophy. – 1986. – Vol. 9. – P. 5–16. [in English].
12. Pustejovsky, J. The Syntax of Event Structure [Text] / J. Pustejovsky // Cognition. – 1991. – Vol. 41. – P. 47 – 81 [in English].
13. Verhagen, M. SemEval-2010 Task 13: TempEval-2 [Text] / M. Verhagen, R. Sauri, T. Caselli, J. Pustejovsky // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 57–62 [in English].
14. Grover, C., Edinburgh-LTG: TempEval-2 system description [Text] / C. Grover, R. Tobin, B. Alex, K. Byrne // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 333–336 [in English].
15. Lafferty, J. D. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data [Text] / J. D. Lafferty, A. McCallum, F. C. N. Pereira // Proceedings of the 18th International Conference on Machine Learning, June 28 - July 1, 2001, Williamstown / ed. C. Brodley, A. Danilyuk. – Morgan Kaufmann : – 2001. – P. 282–289 [in English].
16. Curran, J. R. Language Independent NER Using a Maximum Entropy Tagger [Text] / J. R. Curran, S. Clark // Proceedings of the Seventh Conference on Natural Language Learning HLT-NAACL, May 31, 2003, Edmonton : Canada. – ACL, 2003. – P. 164–167 [in English].
17. Fisher, M. Handbook of Temporal Reasoning in Artificial Intelligence (Foundations of Artificial Intelligence) [Text] / M. Fisher, D. Gabbay, L. Vila. – New York : Elsevier Science Inc., 2005. – 750 p. [in English].
18. McCarthy, J. Some Philosophical Problems From the Standpoint of Artificial Intelligence [Text] / J. McCarthy, P. Hayes // Machine Intelligence 4. – Edinburgh University Press, 1969. – P. 463–502 [in English].
19. Kowalski, R. A Logic-based Calculus of Events [Text] / R. Kowalski, M. Sergot // New Generation Computing. – 1986. – Vol. 4. – No. 1. – P. 67–95 [in English].
20. Allen, J. F. Towards a General Theory of Action and Time [Text] / J. F. Allen // Artificial Intelligence. – 1984. – Vol. 23. – No. 2. – P. 123–154 [in English].
21. Pustejovsky, J. The Specification Language TimeML [Text] / J. Pustejovsky, B. Ingria, R. Sauri // The Language of Time: A Reader / I. Mani, J. Pustejovsky, R. Gaizauskas. – 2004 – Oxford University Press, – P. 545–559 [in English].
22. UzZaman, N. TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text [Text] / N. UzZaman, J. F. Allen // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 276–283.
23. Llorens, H. TIPSem (English and Spanish): Evaluating CRFs and semantic roles in TempEval-2 [Text] / H. Llorens, E. Saquete, B. Navarro // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 284–291 [in English].

24. Kolya, A. K. JU\_CSE\_TEMP: A First Step towards Evaluating Events, Time Expressions and Temporal Relations [Text] / A. K. Kolya, A. Ekbal, S. Bandyopadhyay // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 345–350.

25. Saquete, E. ID 392: TERSEO + T2T3 Transducer: a systems for recognizing and normalizing TIMEX3 [Text] / E. Saquete // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 317–320 [in English].

26. Derczynski, L. USFD2: Annotating Temporal Expressions and TLINKs for TempEval-2 [Text] / L. Derczynski, R. Gaizauskas // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – Uppsala, Sweden: Association for Computational Linguistics, 2010. – P. 337–340 [in English].

27. Strotgen, J. HeidelTime: High quality rule-based extraction and normalization of temporal expressions [Text] / J. Strotgen, M. Gertz // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 321–324 [in English].

28. Kolomiyets, O. KUL: Recognition and normalization of temporal expressions [Text] / O. Kolomiyets, M.-F. Moens // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. — ACL, 2010. – P. 325–328 [in English].

29. Ha, E. Y. NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology [Text] / E. Y. Ha, A. Baikadi, C. Licata, J.C. Lester // Proceedings of the 5th International Workshop on Semantic Evaluation, July 15-16, 2010, Uppsala. – ACL, 2010. – P. 341–344 [in English].



Крапивный Юрий Николаевич, канд. физ.-мат. наук, доц. каф. математического обеспечения компьютерных систем Одесского нац. ун-та им. Мечникова  
e-mail: y.kravivny@gmail.com



Касаткина Галина Владимировна, аспирант каф. математического обеспечения компьютерных систем Одесского нац. ун-та им. Мечникова  
e-mail: drmdra@gmail.com