

УДК 004.89:004.912

УДК 519.767.6:004.81:81'322.2

А. Н. Виноградов, И. Н. Воздвиженский, Д. А. Кормалев,
Е. П. Куршев

Моделирование временного аспекта описания ситуации в задаче извлечения информации из текстов

Аннотация. В настоящей работе рассматривается современное состояние проблемы моделирования временного аспекта описания ситуации в рамках задачи автоматического извлечения информации из текстов на естественном языке. Выделяются основные этапы анализа временного аспекта. Проанализирован опыт разметки текстов стандарта TimeML. Проанализирована коллекция русскоязычных новостных текстов с точки зрения интерпретации временных указателей.

Ключевые слова и фразы: автоматическое извлечение информации, временной аспект, темпоральные выражения, нечеткие указания на время, разметка.

Введение

Во многих практических приложениях технологии извлечения информации из текстов на естественном языке возникает потребность выявлять и интерпретировать временные аспекты описываемых явлений.

Время относится к числу концептуальных и языковых универсалий. Категория времени является одним из основополагающих элементов, формирующих в сознании носителей языка картину мира. Умение «понимать» временные аспекты явлений актуально в любых компьютерных системах, моделирующих когнитивную деятельность.

Работа выполнена в рамках НИР «Моделирование модально-временного аспекта описания ситуаций в задаче извлечения информации из текстов», номер гос. регистрации 01201455353.

- © А. Н. Виноградов, И. Н. Воздвиженский, Д. А. Кормалев, Е. П. Куршев, 2014
- © ИНСТИТУТ ПРОГРАММНЫХ СИСТЕМ ИМЕНИ А. К. АЙЛАМАЗЯНА РАН, 2014
- © ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ, 2014

В приложениях к анализу текстов на естественном языке задача моделирования времени чрезвычайно сложна и может быть отнесена к проблемам представления общих знаний о мире.

В рамках проблемы автоматического извлечения информации из текстов задача выявления темпоральных сущностей впервые была сформулирована на конференциях MUC [1, 2]. Она рассматривалась как одна из составляющих общей задачи извлечения именованных сущностей. Участники MUC-5 должны были извлекать из текста информацию о времени создания совместного предприятия. На MUC-6 в рамках общей задачи извлечения именованных сущностей требовалось извлекать «абсолютные» (absolute) указания на время [1]. На MUC-7 определение извлекаемой темпоральной информации было расширено до включения «относительных» (relative) временных оценок [2]. Никакой дифференциации при разметке не требовалось, а единственный атрибут Type заполнялся значением DATE (по существу, для всех указаний на время, кроме времени суток) или TIME (время суток).

В настоящее время независимо существуют лингвистические и формальные модели времени [3]. Однако до сих пор еще слабо исследованы способы перехода от первых к последним. Кроме того, задача интерпретации текстовых явлений в лингвистические модели времени для русского языка в полном объеме не решена.

В настоящей работе обсуждается лингвистическая сторона моделирования временного аспекта. Работа имеет следующую структуру. Сначала выделяются основные этапы моделирования временного аспекта. Затем делается попытка обзора существующих подходов к моделированию временного аспекта, в соответствии со сформулированными этапами. Далее, исследуется возможность интерпретации временных выражений в русскоязычных новостных текстах.

1. Основные этапы моделирования временного аспекта

По мнению авторов настоящей работы, общая задача лингвистического моделирования временного аспекта предполагает выполнение следующих шагов:

- (1) исследование способов представления темпоральной информации в тексте;
- (2) построение лингвистической (ориентированной на текст) модели времени;

- (3) разработка алгоритмов и программных средств для распознавания темпоральной информации в тексте и интерпретации её в лингвистической модели;
- (4) адаптация существующей логической модели времени, адекватной задаче извлечения информации (понимаемой как задача построения фактографического ресурса);
- (5) интерпретация лингвистической модели временного аспекта в формальную модель;
- (6) исследование механизмов рассуждений в модели и построение вывода для извлеченной и проинтерпретированной в лингвистической модели временной информации.

В действительности в современных прикладных исследованиях вопросов извлечения темпоральной информации отсутствует такое четкое разделение на этапы. Работы ведутся по разным, трудно структурируемым направлениям. Чаще всего исследователи в работах последних лет сосредотачивают свои усилия на решении довольно сложных, но узких задач. Так, одной из самых исследуемых в последние годы тем является попытка автоматического построения последовательности событий в тексте [4–7]. В настоящей работе мы попробуем систематизировать достижения в области извлечения, интерпретации и моделирования темпоральной информации.

2. Обзор этапов моделирования временного аспекта

2.1. Языковые средства выражения темпоральной информации

Человеческие представления о времени, с одной стороны, философски-абстрактны (время как «бесконечная, непрерывная, равномерно текущая субстанция, существующая сама по себе, независимо от совершающихся в мире событий и взаимоотношений между ними»), а с другой — эмпирически-конкретны (время как неотъемлемый атрибут конкретных явлений и отношений между ними) [3]. В языке выражаются и диалектически взаимодействуют оба вида представлений.

Язык располагает целым арсеналом грамматических и лексических средств для выражения временных категорий. В лингвистике принято различать два основных типа употреблений временных значений: прямое (абсолютное) и относительное. Первое определяет

событие по отношению к моменту речи — как одновременное, предшествующее или последующее. Относительное время служит для выражения отношений одновременности/разновременности между событиями безотносительно к моменту речи. Основными языковыми средствами, служащими для передачи векторного времени, являются специальные глагольные временные формы (прошедшее, настоящее и будущее грамматическое время), возможно в сочетании с временной лексикой.

Позиционное время выражается принципиально иными языковыми средствами. Такими средствами могут быть: относительные глагольные времена (в индоевропейских языках); специальная глагольная и неглагольная лексика (*предшествовать, следовать, совпадать, раньше, позже, одновременно*); различные категории вида у предикатов; временные союзы (*до того как, после того как, в то время как* и т.п.); временные предлоги (*перед, до, после, перед, в течение, во время* и т.д.) и наречия (*затем, потом, одновременно*). Эти средства выражения позиционного времени функционируют как в пределах предложения, так и на сверхфразовом уровне. Множество позиционных временных оценок (случай установления определенных временных отношений между ситуациями безотносительно к моменту речи) весьма неоднородно. Некоторые исследователи считают необходимым различать несколько смежных категорий: относительное время, таксис, эвиденциальность [8].

Категория эвиденциальности («засвидетельствованности») интересна тем, что представляет собой модально-временной оператор [3]. К случаям, в которых проявляется категория эвиденциальности, относятся так называемые конструкции с предикативным актантом (КПА): простые предложения, семантически осложненные различными дополнительными пропозитивными единицами, выступающими в функции событийных актантов сказуемого. Это инфинитивные, причастные конструкции, разного рода номинализации. В качестве подчиняющихся могут выступать лишь определенные группы предикатов («модусные предикаты»):

- (1) предикаты ментальной (интеллектуальной, эмоциональной оценочной) деятельности: *знать, полагать, опасаться, желать, сожалеть*;
- (2) перцептивные глаголы: *чувствовать, видеть, слышать* и т.п.;
- (3) перформативы (предикаты речи): *говорить, сказать, сообщать, спрашивать, утверждать* и т.д.

В таких КПА подчиненная пропозиция передается «от лица» субъекта подчиняющей пропозиции и, следовательно, несет в себе субъективно-модальный компонент. Временная оценка при этом имеет относительный характер: положение подчиненной пропозиции на оси времени задает подчиняющая пропозиция.

В рамках задачи автоматического извлечения темпоральной информации необходимо продумать такую классификацию способов выражения временной информации, которая была бы удобна, с одной стороны, для создания лингвистической модели временного аспекта, а с другой стороны, позволяла бы оптимальным образом помечать в тексте такую информацию и соотносить ее с объектами в лингвистической модели. Для русского языка такая работа систематически не проводилась. Исследователи ограничиваются созданием простейшей лингвистической модели, с которой соотносятся обычно самые простые языковые временные указатели.

2.2. Алгоритмы распознавания темпоральных выражений

Методы извлечения темпоральных сущностей делятся на инженерные («knowledge-based») и методы машинного обучения.

Инженерные методы основываются на правилах и словарях. Словари обычно содержат наборы ключевых слов, характерных для временных выражений, списки глаголов и т.п. Правила содержат темпоральные шаблоны (форматы дат, времен и т.п.).

Для задачи извлечения инженерные подходы относительно просты и эффективны, хотя любая перенастройка требует значительных трудозатрат [9]. В существующих решениях задачи извлечения информации о времени пока преобладает инженерный подход, в то время как в общей задаче извлечения информации налицо явное преимущество обучаемых методов.

Методы машинного обучения требуют создания большой коллекции размеченных текстов, на которой будет происходить обучение. Эти методы обычно используют в качестве признаков всю имеющуюся лингвистическую информацию (результаты синтаксических, морфологических разборов), знания, накопленные в системе, графематическую информацию (наличие чисел определенного формата, слов, написанных с большой буквы и т.п.), а также словари и списки специальных слов. Здесь возможно выявление сложных зависимостей, неподвластных инженерному подходу, однако и число признаков может быть довольно велико и приводить к переобучению.

Поэтому возникает необходимость в специальных процедурах отбора релевантных признаков или сокращения размерности пространства признаков.

Нам представляется, что методы машинного обучения хорошо работают для сегментации темпоральных выражений. То есть для выявления в тексте границ такого выражения. В недавней работе [10] были получены достаточно высокие результаты такой сегментации (точность 96,3%, полнота 89,9%, F-мера 93,05%). Авторы использовали метод условных случайных полей, естественно возникающий для разметки последовательностей с взаимными зависимостями членов.

В то же время, интерпретация выявленного временного выражения на настоящий момент производится методами машинного обучения только для выражений специального вида. Эта задача пока нуждается в дальнейших исследованиях. Методы машинного обучения могут быть привлечены для автоматической классификации временных выражений.

2.3. Формальная модель времени

Полное решение задачи автоматического извлечения временной информации из текста предполагает расположений всех событий, извлеченных из данного текста, на абсолютной временной шкале. Эта шкала фактически является промежуточным звеном на пути от лингвистической к формальной модели временного аспекта.

В настоящий момент разработан целый ряд формальных моделей времени и правил вывода в них [11–13]. Модели различаются по способу представления времени:

- (1) интервальное или точечное время,
- (2) наличие шкал (метрические модели времени),
- (3) наличие периодичности,
- (4) конкретное и вероятностное время.

Кроме того, модели различаются по назначению (например, ограниченные модели, ориентированные на вычислительно эффективный вывод) и по области применения.

Например, в работе [12] хорошо разработана формальная модель времени, содержащая понятия интервального и точечного времени, учитывающая наличие периодичности событий. Однако эта модель предназначена для узкой практической задачи извлечения информации из расписаний.

В русскоязычных публикациях имеется описание одного реализованного на практике решения в области лингвистического моделирования времени в мультязычных коллекциях документов [11]. Подход ограничен извлечением «конструкций, соответствующих временным сущностям различных типов», которые впоследствии рассматриваются как актанты ситуации. Однако об онтологической модели времени и интерпретации извлеченных фрагментов речи не идет.

2.4. Разметка аннотированных корпусов текстов как попытка интерпретации лингвистической модели в формальную

Для оценки качества извлечения информации о времени, сравнения различных методов извлечения такой информации, обучения машинных статистических методов необходимо иметь стандарт разметки текстовых корпусов. Такой стандарт фактически задает способ интерпретации лингвистической модели временного аспекта в формальную модель представления знаний о времени.

Для разметки аннотированных корпусов разработан метастандарт TimeML [14]. Он позволяет размечать в документах события и их временную привязку. Схема TimeML предусматривает разметку глаголов, прилагательных и существительных, соответствующих событиям и состояниям; определение грамматического времени, вида, «полярности» (отрицательная или положительная); разметку модальных операторов и даже число упоминаний одного и того же события. Выражения времени не только распознаются, но и нормализуются. Размечаются различные указатели темпоральных отношений (темпоральные предлоги и союзы), а также случаи модально-временного подчинения событий. Кроме того, отношения в парах событий и отношения между событиями и выражениями времени помечаются специальными отношениями временного порядка (в терминах временной логики Аллена).

Система разметки TimeML устроена аналогично xml-разметке и содержит следующие основные теги:

- (1) Event — элементы текста, которые означают события. Как правило, это глаголы, а также участники временного отношения.
- (2) TimeX3 — временные выражения: время, даты, интервалы.
- (3) Signal — функциональные слова, отвечающие за наличие временного отношения. Например, *когда, в, после, до*.
- (4) TLink — устанавливает связь событий (возможно нескольких) и временных выражений. Связь путем упорядочивания во времени.

Аннотации событий содержат следующие атрибуты: тип события, грамматическое время, грамматические характеристики, полярность (положительная или отрицательная), модальность, кратность (сколько раз упомянуто в тексте).

Остановимся подробнее на аннотациях типа TLink. Они соответствуют отношениям между событиями и/или временными интервалами (момент времени рассматривается как интервал, у которого совпадают начало и конец). Существует развитая система типов таких отношений, она восходит к *interval calculus* [15], которая содержит 14 типов.

В стандарте TimeML оставлено 12 типов отношений: BEFORE, BEGINS, ENDS, IBEFORE, INCLUDES, SIMULTANEOUS и их обращения. Отношениям соответствуют условия (*metric constrains*), т.е. неравенства [16].

В качестве ответа автоматической системы допускается дизъюнкция отношений, но аннотатор при разметке текста вручную ставит только одно однозначно определенное отношение.

В 2004 году агентством DARPA и программой ACE была запущена инициатива TERN — Time Expression Recognition and Normalization. Её цель — разметка времени в корпусе текстов на английском и китайском языках на основе требований аннотационного стандарта TIMEX2. Время, указанное в атрибутах аннотаций разметки TimeML, нормализуется в соответствии с этим стандартом. Существуют следующие размеченные корпуса:

- (1) TimeBank 1.2 содержит 186 текстов, 64 тысячи слов [17].
- (2) Opinion Corpus содержит 73 текста, 39 тысяч слов [18].

Для русского языка таких общедоступных коллекций пока не создано. Вопрос о приложении стандарта разметки TimeML к русскоязычным текстам пока не имеет однозначного ответа.

2.5. Выводы в темпоральной модели

Работа со временем в практических приложениях, использующих текстовые данные, не может ограничиваться выделением в тексте явных темпоральных выражений и констатацией факта наличия того или иного временного отношения. Интеллектуальный анализ текста предполагает и умение выполнять рассуждения в некоторой темпоральной модели.

Для различных моделей разработаны индивидуальные системы отношений и правил вывода. Вопрос выбора модели, адекватной для интерпретации сведений, содержащихся в тексте, исследован слабо.

Некоторое подобие вывода существует в модели TimeML. По TLinkам строится граф (вершины которого суть события или временные интервалы, а ребра соответствуют отношениям). Рассматривается «замыкание» этого графа. Т.е. достраиваются ребра, которые можно вывести из имеющихся с помощью аксиом [19–21]. Некоторые TLinkи, полученные замыканием, будут дизъюнктивными, такие TLinkи уничтожаются.

Большое количество работ посвящено установлению отношений порядка на множестве событий, извлеченных из текста, например, [4–7]. Однако, так или иначе авторы этих работ резко сужают общую постановку задачи. Например, в работе [4] рассматривается корпус документов, содержащих информации об одном и том же событии (при этом в них могут упоминаться и другие события). В работе [7] рассматриваются расшифровки устных новостных радиосообщений за короткий период времени. В то же время нам представляется, что предварительная задача извлечения событий и соотнесения им временных указателей еще далека от идеального решения.

3. Сложности интерпретации указаний на время в русскоязычных текстах

Стандарт TimeML изначально разработан для текстов на английском языке. Аналогичных стандартов для русского языка не существует. По всей видимости, применение стандарта TimeML для русского языка столкнется с некоторыми трудностями, так как лингвистические способы выражения временной информации в русском и английском языках сильно различаются. Кроме того, различается и восприятие временной информации, выраженной в тексте, разными носителями языка.

В качестве попытки исследования многообразия средств выражения временной информации в русском языке и выяснения проблем ее интерпретации нами был проведен опрос нескольких десятков информантов. Им было предложено проинтерпретировать в свободной словесной форме временные маркеры в небольшом корпусе новостных текстов. Предварительного отбора текстов в целях сужения задачи не проводилось. В частности, эти тексты содержали нечеткие временные указатели.

Цель исследования — понять, насколько различаются интерпретации одних и тех же временных указателей разными носителями языка.

Был создан набор текстов, в которых выделены представляющие интерес для исследователей указания на время. Был разработан шаблон интерпретации, т.е. таблица с количественными оценками для каждого указания на время. Масштаб шкалы времени для каждого указания на время следовало выбирать индивидуально (часы, дни, месяцы, годы, столетия).

Далее приводятся примеры и ниже различные возможные интерпретации временных выражений.

3.1. Примеры временных выражений, задающих интервалы с фиксированным концом

Живя с *2001 года* в доме престарелых, Жозе Агинело душ Сантуш часто рассказывал работникам заведения о событиях, которые имели место *в начале XX века* и даже *ранее*.

в начале XX века = 1900–1920 ∨ 1901–1925 ∨ 1900–1906 ∨ 1900–1910
ранее = 1890–1900 ∨ 1885–1900 ∨ 1896–1900 ∨ 1875–1899

Они вызвали экспертов, которые задали мужчине вопросы и установили: он не обманывает, утверждая, что родился *в конце 19 века*.

в конце 19 века = 1890–1899 ∨ 1880–1900 ∨ 1885–1899 ∨ 1875–1900

Напомним, что работы по строительству второго пути железной дороги Выборг-Каменногорск *в начале октября 2012 года* были приостановлены.

в начале октября 2012 года = 1–10 октября 2012 года ∨ 1–7 октября 2012 года

3.2. Примеры временных выражений, задающих интервалы с фиксированной точкой внутри

Самолет АН-26 *в ночь на пятницу* забрал из Ростовской области четверых раненых военнослужащих Украины.

в ночь на пятницу = 23:00–4:00 ∨ 0:00–6:00 ∨ 22:00–3:00 ∨ 0:00–1:00

Сегодня *около 01.30 мск* украинский борт забрал из Ростова-на-Дону оставшихся четверых раненых военнослужащих Украины.

около 01.30 мск = 1:20–1:40 ∨ 1:15–1:45 ∨ 1:25–1:35 ∨ 1:27–1:32

3.3. Примеры временных выражений, задающих интервалы

Заметим, что конец интервала может отличаться от точки отсчета.

Премьер-министр Польши Дональд Туск заявил, что угроза военного вторжения российских Вооруженных сил на территорию Украины *за последние несколько дней* возросла.

за последние несколько дней = 1–3 дня ∨ 3–5 дней ∨ неделя ∨ 1–5 дней ∨ 15 дней

Мы имеем основания полагать, что *в течение последних часов* мы получили соответствующую информацию, что в настоящий момент риск прямой интервенции России значительно выше, чем *несколько дней назад*.

в течение последних часов = 1–5 часов ∨ 3–5 часов ∨ 20 часов

несколько дней назад = 1–5 дней ∨ 2–4 дня ∨ две недели

В последнее время в СМИ неоднократно появлялась информация об увеличении численности российских войск на границе с Украиной.

В последнее время = месяц ∨ неделя 5–10 дней ∨ две недели

Информанты могут по-разному интерпретировать масштаб интервала (день, месяц, год).

Американские ученые назвали *недавно* открытый вид рыб в честь президента США Барака Обамы.

недавно = несколько месяцев — год ∨ 10 лет ∨ 10–40 дней ∨ 0–3 месяца

Отношения между Индией и Пакистаном *в последние годы* остаются напряженными, хотя *недавно* обе стороны приняли ряд шагов по смягчению разногласий.

в последние годы = 0–10 лет ∨ 3–4 года ∨ 0–5 лет

недавно = 0.5–1 год ∨ 3 месяца ∨ 2–3 года

Таким образом, можно сделать вывод, что временные указатели интерпретируются носителями языка очень неоднозначно. Это говорит о том, что для построения лингвистической модели временного аспекта в целях практического извлечения информации из текстов необходимо резко сужать множество категорий объектов и отношений между ними, которые лягут в основу модели временного аспекта для автоматического извлечения информации.

Заключение

Таким образом, в работе сформулированы основные этапы моделирования временного аспекта, освещены основные достижения в этой области. Кроме того, в качестве попытки построения лингвистической модели времени для системы автоматического извлечения информации из русскоязычных текстов было проведено исследование по интерпретации произвольно отобранных указаний на время. Результаты этого исследования показали необходимость построения лингвистической модели времени и алгоритмов ее интерпретации в формальную модель для конкретной задачи автоматического извлечения информации определенного вида. Необходимо разработать принципы разметки русскоязычных документов. Авторы предлагают создать открытую коллекцию текстов с размеченными временными указателями.

Благодарности. Авторы благодарят Н. А. Власову, Е. А. Сулейманову, И. В. Трофимова за полезные обсуждения и А. В. Подобреева за помощь в подготовке статьи к публикации.

Список литературы

- [1] Sundheim В., Chinchor N., “Named Entity Task Definition, Version 2.0, 31 May 95”, *Proceedings of the 6th Message Understanding Conference*, MUC-6, Morgan Kaufman, 1995, pp. 319–332 ↑ 216.
- [2] Chinchor N., “MUC-7 Information Extraction Task Definition, Version 5.1, 23 July 1998”, *Proc. of the 7th Message Understanding Conference*, MUC-7, 1998 ↑ 216.
- [3] Полянский С. М., «Таксис — относительное время — эвиденциальность (к проблеме критериев разграничения)», Сибирский лингвистический семинар, **2**, Новосибирск, 2001 ↑ 216, 217, 218.
- [4] Filatova E., Novy E., “Assigning Time-Stamps to Event-Clauses”, *Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing (Toulouse, France, 2001)*, pp. 88–95 ↑ 217, 223.
- [5] Mani I., Schiffman B., Zhang J., “Inferring Temporal Ordering of Events in News”, *Proceedings of HLT-NAACL 2003 (Edmonton, May–June 2003)*, pp. 55–57 ↑ 217, 223.
- [6] Muller P., Tannier X., “Annotating and measuring temporal relations in texts”, *Proceedings of the 20th International Conference on Computational Linguistics (Coling 04) (Geneva, Switzerland, 2004)*, pp. 50–56 ↑ 217, 223.
- [7] Mani I., Wilson G., “Robust Temporal Processing of News”, *ACL/2000 (2000)*, pp. 69–76 ↑ 217, 223.

- [8] Якобсон Р. О., «Шифтеры, глагольные категории и русский глагол», *Принципы типологического анализа языков различного строя*, пер. с англ., М., 1972 ↑ 218.
- [9] Wong K.-F., Xia Y., Li W., Yuan C., “An Overview of Temporal Information Extraction”, *International Journal of Computer Processing of Oriental Languages*, **18**:2 (2005), pp. 137–152 ↑ 219.
- [10] Kudinov M. S., Romanenko A. A., Piontkovskaja I. I., “Conditional Random Field in Segmentation and Noun Phrase Inclination Tasks for Russian”, Papers from the Annual International Conference “Dialogue”, Computational Linguistics and Intellectual Technologies, **13**, 2014 ↑ 220.
- [11] Ефименко И. В., «Время в мультязычных коллекциях документов: лингвистическая модель и ее реализация в среде GATE», *Тр. Девятой национальной конференции по искусственному интеллекту с междунар. участием*, КИИ-2004, Физматлит, М., 2004, с. 508–516 ↑ 220, 221.
- [12] Weiser S., Laublet P., Minel J.-L., “Automatic Identification of Temporal Information in Tourism Web Pages”, The International Conference on Language Resources and Evaluation (LREC) (2008) ↑ 220.
- [13] Fisher M. D., *Handbook of temporal reasoning in AI*, Elsevier Academic Press, 2010 ↑ 220.
- [14] Pustejovsky J., Ingria B., Sauri R., Castano J., Littman J., Gaizauskas R., Setzer A., Katz G., Mani I., “The Specification Language TimeML”, *The Language of Time: A Reader Mani*, eds. Pustejovsky J., Gaizauskas R., Oxford University Press, 2005 ↑ 221.
- [15] Allen J., “Towards a General Theory of Action and Time”, *Artificial Intelligence*, **23**:2 (1984), pp. 123–154 ↑ 222.
- [16] Kautz H. A., Ladkin P. B., “Integrating Metric and Qualitative Temporal Reasoning”, AAAI’91 (1991) ↑ 222.
- [17] TimeBank 1.2, <http://timeml.org/site/timebank/timebank.html> ↑ 222.
- [18] Opinion Corpus, http://timeml.org/site/timebank/aquainttimeml/aquaint_timeml.1.0.tar.gz ↑ 222.
- [19] Verhagen M., Mani I., Sauri R., Knippen R., Littman J., Pustejovsky J., *Automating Temporal Annotation with TARSQI*, Demo Session, ACL, 2005 ↑ 223.
- [20] Setzer A., Gaizauskas R., “Annotating Events and Temporal Information in Newswire Texts”, Proceedings of LREC-2000 (2000), pp. 1287–1294 ↑ 223.
- [21] Vilain M., Kautz H., Beek P. V., “Constraint propagation algorithms for temporal reasoning: A revised report”, *Readings in Qualitative Reasoning about Physical Systems*, eds. Weld D. S., de Kleer J., Morgan-Kaufman, 1989, pp. 373–381 ↑ 223.

- [22] *Situations-1000. Размеченная коллекция новостных текстов на русском языке, содержащих информацию о назначениях и отставках мч*, <http://ai-center.botik.ru/Airec/index.php/ru/collections/33situations-1000>, Исследовательский центр искусственного интеллекта, ИПС им. А. К. Айламазяна РАН, 2014 ↑.

Рекомендовал к публикации

д.ф.-м.н. С.В. Знаменский

Об авторах:



Андрей Николаевич Виноградов

Заместитель руководителя Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН, к.ф.-м.н.

e-mail:

andrew@andrew.botik.ru



Илья Николаевич Воздвиженский

Младший научный сотрудник Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН, к.т.н.

e-mail:

vozdvin@yandex.ru



Дмитрий Анатольевич Кормалев

Младший научный сотрудник Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН

e-mail:

dkormalev@gmail.com



Евгений Петрович Куршев

Руководитель Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН, к.т.н.

e-mail:

epk@epk.botik.ru

Образец ссылки на эту публикацию:

А. Н. Виноградов, И. Н. Воздвиженский, Д. А. Кормалев, Е. П. Куршев. *Моделирование временного аспекта описания ситуации в задаче извлечения информации из текстов // Программные системы: теория и приложения: электрон. научн. журн.* 2014. Т. 5, № 4(22), с. 215–229.

URL http://psta.psiras.ru/read/psta2014_4_215-229.pdf

Andrey Vinoradov, Ilya Vozdvizhenskiy, Dmitrii Kormalev, Evgenii Kurshev.
The time aspect modelling of situation description for information extraction task.

ABSTRACT. In this paper we give a brief overview of the state of the art in information extraction of time aspects from natural language texts. We introduce the main steps of such analysis. We analyze TimeML experience in event and time annotation. We study human interpretation of temporal expressions in the Russian-language news texts. (*In Russian*).

Key Words and Phrases: automatic information extraction, modal aspect, temporal information, temporal expressions, confusing temporal expressions, markup.