

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2016»

Москва, 1–4 июня 2016

SITUATIONS-1000: КОРПУС ДЛЯ ЗАДАЧИ ИЗВЛЕЧЕНИЯ СОБЫТИЙ ИЗ ТЕКСТОВ¹

Власова Н. А. (nathalie.vlassova@gmail.com),

Ландо Н. С. (nlando1@gmail.com),

Сулейманова Е. А. (yes@helen.botik.ru),

Трофимов И. В. (itrofimov@gmail.com)

Институт программных систем им. А. К. Айламазяна РАН,
Переславль-Залесский, Россия

Ключевые слова: извлечение информации, события, русскоязычный
корпус, аннотирование

SITUATIONS-1000: A TAGGED CORPUS FOR EVENT EXTRACTION FROM TEXTS

Vlasova N. A. (nathalie.vlassova@gmail.com),

Lando N. S. (nlando1@gmail.com),

Suleymanova E. A. (yes@helen.botik.ru),

Trofimov I. V. (itrofimov@gmail.com)

Program Systems Institute of RAS, Pereslavl-Zalesky, Russia

In recent years, there has been a regrowth of interest in methods of information extraction (IE) from Russian-language texts. Evaluating the performance of such methods is impossible without tagged corpora. Developing the annotation principles and guidelines involves a lot of decision-making, and there is always an element of subjectivity in it.

The paper outlines the authors' experience with developing first-layer (surface-based) annotation for an event-tagged corpus of Russian newswire texts. The problems faced range from general issues involved in annotating

¹ Работа выполнена при поддержке РФФИ (Проект № 13-06-00483).

IE-oriented corpora (target event specification, factuality aspects) to those stemming from the text-based character of the annotation (in contrast to MUC-type annotation in the form of scenario templates)—event mention scope, participant mention boundaries, syntactic zeroes, ellipses, coordination reduction. A summary of principles and decisions underlying the annotation scheme is given, accompanied by examples to make explicit how the annotations apply to real-text cases.

Keywords: information extraction, event, Russian-language corpus, annotation scheme

1. Введение

В последнее время наблюдается возрождение интереса к методам извлечения информации из текстов на русском языке. Разметка корпусов для оценки качественных характеристик таких методов требует принятия субъективных решений на стадии выработки принципов аннотирования. Поэтому каждое решение должно быть тщательно взвешено, обосновано и, по возможности, проверено на значительном объеме языкового материала.

В задаче разметки событий ключевыми являются следующие вопросы:

1. Каким образом максимально строго определить содержательный перечень событий и их атрибутов, подлежащих разметке?
2. Что именно является объектом разметки — упоминания *ситуаций* заданного типа (т. е. случаи употребления пропозиций целевого класса) или упоминания *фактов*? Если второе, то что должно служить аннотатору основанием для принятия решения: можно ли считать, что имел место факт *целевого* события? Отражать ли в разметке различные аспекты, характеризующие событие в его отношении к действительности?
3. Если целью разметки является аннотированный текст (в противовес фрейму/шаблону, заполняемому извлеченной из текста информацией), то что будут собой представлять элементы разметки?

Неочевидность ответов на эти вопросы проиллюстрируем на примере событий назначения и отставки.

Первый вопрос — содержательное определение подлежащих разметке событий. Что такое назначение и чем оно отличается от других способов получения должности? Являются ли назначением, например, выборы, переизбрание, повышения и понижения в должности, временное исполнение обязанностей и т.п.? Обнаружить близкие по смыслу категории событий и компактно сформулировать принцип, которым мог бы пользоваться размечающий для различения целевых и нецелевых событий, — первоочередная задача. Нужно избегать как излишней детализации (чем больше объем инструкции, тем труднее ей следовать), так и переобобщений (можно определить событие потери должности таким образом, что под него попадут примеры типа *Директор ООО «Люттик» получил несовместимые с жизнью травмы*).

Сложности возникают и с определением участников события. Рассматриваемые события характеризуются тремя участниками: назначаемый (увольняемый), должность и назначающий (увольняющий). Как поступать, если должность не названа явно?

- (1) ... он был назначен исполнять обязанности губернатора N-ской области ...
[должность исполняющего обязанности губернатора подразумевается]
- (2) ... он возглавил комитет по ... [должность не названа]

Какой идентифицирующей информацией должен обладать участник, чтобы событие считалось выраженным в тексте?

- (3) один из членов его команды впоследствии стал директором этой компании

Проблема содержательного определения целевого типа события переключается со вторым из поставленных ранее вопросов, а именно: что считать фактом целевого события? Этот вопрос связан не только с модальностью, полярностью, временными характеристиками и референциальным статусом, которым уделяется немалое внимание в литературе. Речь идет о разнообразии лексических средств, используемых в реальных текстах для референции к целевым фактам. Нередко граница между фактом и не-совсем-фактом оказывается размыта: к примеру, словосочетания *подписал указ о назначении*, *подал рапорт об отставке*, строго говоря, не обозначают события назначения и отставки, однако в новостных текстах по умолчанию² используются именно в таком смысле. В процессе работы над инструкцией для размечающих требуется тщательная проработка вопросов «фактуальности».

Наконец, третья группа вопросов касается собственно разметки, понимаемой как аннотирование текста:

- что принять за область поиска участников;
- как установить границы размечаемых элементов;
- как размечать различные случаи синтаксической неполноты;
- разрешать ли кореферентность при разметке и др.

В статье затрагиваются проблемы, с которыми авторы столкнулись в ходе поверхностно-синтаксической разметки событий в корпусе новостных сообщений Situations-1000³.

² Ср. *подал рапорт об отставке, который был отклонен*.

³ <http://ai-center.botik.ru/Airec/index.php/ru/collections/33-situations-1000>

2. Зарубежный опыт

Кратко, в табличной форме, представим принципы, лежащие в основе разметок MUC [1], ACE [2], TimeML [3–4].

Таблица 1. Особенности разметок событий в MUC, ACE и TimeML

Особенность разметки	MUC-7	ACE	TimeML
Разметка ограниченного перечня событий	да	да	нет
Разметка места упоминания события в тексте	нет	да	да
Разметка участников события	да	да	нет
Разметка информации об участниках в структурированном виде	да	нет	—
Область поиска информации об участниках	весь текст	предложение	—
Разметка места упоминания участника события в тексте	нет	да	—
Локализация события во времени	да	да	да
Локализация события в пространстве	да	да	нет
С учётом модальности	# ⁴	да, размечают присутствие модальности	нет
С учётом полярности	#	да, размечают	нет
С учётом общности	#	да, обобщения размечают	да, обобщения не размечают
С учётом стадии	#	нет	да, класс событий <i>aspectual</i>
Разметка событий, выраженных местоимениями и дескрипциями с отсылкой к ранее упомянутому событию-антецеденту	—	да	нет

⁴ Знак «#» означает, что данный аспект задачи не рассматривается в описании / инструкции для разметки.

3. Опыт разметки событий в корпусе

Задача фактографического извлечения информации предполагает обнаружение в тексте фактов определенного типа и выделение их участников. Описанная в статье разметка представляет собой первый шаг в создании (первый слой разметки) корпуса для оценки точности и полноты алгоритмов извлечения из текстов фактов заданного типа.

Содержательно целевыми считаются события назначения и отставки. Для задач разметки возможные лексико-синтаксические способы их выражения были сгруппированы в пять размечаемых типов событий, различающихся между собой наборами участников (типы и участники приводятся ниже). Участник с условным названием «каузатор» (лицо, назначающее на должность, принимающее отставку и т.п. — т.е. лицо, в силу своей роли в ситуации уполномоченное принимать решения) было решено считать необязательным, что дало возможность объединить в один тип размечаемых событий группы предикатов, различающихся только наличием или отсутствием валентности каузатора (таким образом были получены типы 1 и 2). С другой стороны, участник «должность» для обоих типов был объявлен обязательным (это дало возможность исключать из разметки случаи типа *Иванов устроился в Яндекс или уволился из компании*). Глагол *возглавить* и его производные пришлось отнести к отдельному размечаемому типу «возглавление», поскольку при них не может быть выражен участник «должность». Парный ему тип со значением отставки главы выделять не было смысла, т. к. аналогичных глаголов со значением ухода в отставку главы в русском языке нет (*Иванов возглавил (*обезглавил) компанию*). Из тех же соображений пришлось ввести, наряду с типом «смена лиц в должности», тип «смена главы»: к нему может быть отнесено небольшое число конструкций с невыразимой должностью (*возглавить вместо, поставить во главе вместо, встать во главе вместо*).

Размечаемые типы событий:

1. назначение / вступление в должность
 - должность
 - занимающий
 - каузатор
2. отставка / увольнение / уход с должности
 - должность
 - оставляющий
 - каузатор
3. смена лиц в должности
 - должность
 - занимающий
 - оставляющий
 - каузатор
4. возглавление
 - занимающий
 - возглавляемый объект
 - каузатор

5. смена главы
 - занимающий
 - оставляющий
 - возглавляемый объект
 - каузатор

3.1. Объекты и элементы разметки

В качестве объектов разметки выступают упоминания целевых событий.

Индикатором упоминания целевого события в тексте является маркерная конструкция (см. ниже) — при условии, что модально-коммуникативная рамка позволяет говорить, с той или иной степенью уверенности, о факте события заданного типа.

Маркерные конструкции (МК) — это глаголы и глагольные словосочетания из заранее заданного перечня, а также их синтаксические дериваты (включая номинализацию). Примеры из перечня целевых глаголов и глагольных сочетаний, сгруппированные по типам событий, приведены в таблице 2.

Упоминанием факта, подлежащим разметке, считается фрагмент текста, удовлетворяющий следующим условиям:

1. фрагмент содержит одну маркерную конструкцию;
2. во фрагменте синтаксически выражен *выделенный* участник соответствующего целевого типа события. Выделенными участниками считаются:
 - должность — для целевых типов 1, 2, 3;
 - возглавляемый объект — для целевых типов 4, 5.
3. модально-временное оформление фрагмента позволяет считать, что имел место факт события, упомянутого посредством МК. В том случае, если достоверность факта вызывает у размечающего какие-либо сомнения, он может разметить текстовый «маркер сомнения».

Так, в следующем фрагменте упоминание не размечается, хоть и содержит маркерную конструкцию *отставка* (не выполнено условие 3):

- (4) *Сейм Польши проголосовал против отставки спикера сейма Радослава Сикорского, которой добивалась оппозиция.*

Размечать МК целиком не слишком удобно, поскольку она не обязательно представляет собой непрерывный фрагмент текста. Поэтому в качестве элемента разметки упоминания было решено использовать маркерное слово (маркер целевого события, далее просто *маркер*). Если в состав МК входит словоформа лексемы со значением целевого события, то именно она помечается как маркер. При отсутствии такой словоформы маркер может быть выбран произвольно. В примерах в таблице маркеры в составе МК выделены полужирным начертанием.

Таблица 2. Примеры маркерных конструкций⁵⁶⁷

назначение / вступление в должность	<p>назначить (а также его синтаксические дериваты назначение, (был) назначен, назначенный, назначив, и т. п.) назначить на должность/пост/место вступить в должность занять должность/пост/место получить должность/пост/место перевести на должность/пост/место перейти на должность/пост/место повысить до (должности) подписать указ о назначении (на должность) получить назначение (на должность) прийти на должность и т. п.</p>
отставка / увольнение / уход с должности	<p>отставка² уйти в отставку (с должности/поста) лишиться должности объявить об отставке (с должности/поста) подать рапорт об отставке (с должности/поста)) освободить от (занимаемой) должности освободить кресло оставить должность/пост/место отправить в отставку (с должности) подать в отставку (с должности) подписать указ об увольнении (с должности/поста) покинуть должность/пост и т. п.</p>
смена лиц в должности	<p>сменить в должности занять место (кого/чье) назначить на должность... вместо назначить на место (кого/чье) получить должность... вместо поставить на место смениться (о должности) уступить должность стать вместо</p>
возглавление ³	<p>возглавить встать во главе поставить во главе</p>
смена главы	<p>возглавить вместо встать во главе вместо</p>

⁵ Полный перечень МК содержится в сопроводительной документации к корпусу.

⁶ Слово *отставка*, которое может самостоятельно выступать в качестве МК (*после его отставки*), включено в перечень глаголов как номинализация фиктивного глагола *отставить (ср. *уволить* > *увольнение*).

⁷ МК *назначить главой*, *стать главой* мы относим к типу «назначение», где должность — ИГ с вершиной *глава*.

Разметка упоминания включает в себя следующие шаги:

1. обнаружение в тексте МК, принятие решения о разметке (подлежит ли фрагмент разметке в качестве упоминания целевого события); определение типа размечаемого события (событие должно быть отнесено к одному из пяти целевых типов), выделение маркера;
2. определение **области разметки события** — установление границ упоминания, в которых будет выполняться разметка ИГ-участников (см. ниже);
3. разметка **ИГ-участников** — именных групп, представляющих актанты целевой пропозиции и входящих в область разметки события. Разметка предполагает определение границ именной группы и приписывание ей роли соответствующего участника.

При разметке тип события приписывается упоминанию (маркеру) аннотатором. Маркер, таким образом, служит носителем типа события и фиксирует место упоминания в тексте. Содержательными элементами разметки упоминания являются только ИГ-участники.

3.2. Область разметки события

Чтобы обеспечить поверхностно-синтаксический характер разметки, было решено ограничить область разметки события пределами клаузы, содержащей МК. Эта установка вполне оправдана для случаев, когда синтаксическая вершина МК — глагол в личной форме или краткое причастие. В других случаях последовательно придерживаться ее означает оставлять за пределами разметки этого уровня часть очевидно релевантной информации. Решение, принятое на данном этапе, состоит из нескольких положений, которые мы проиллюстрируем примерами.

Для наглядности в приведенных примерах маркер выделен полужирным шрифтом, синтаксическая вершина МК — подчеркиванием (таким образом, маркер-вершина выделен одновременно полужирным шрифтом и подчеркиванием). Область разметки события, соответствующего выделенному маркеру, ограничена фигурными скобками.

1. Если вершина МК — глагол в личной форме или краткое причастие, то областью разметки события считается содержащая МК финитная клауза. Вложенные клаузы в область разметки данного события не включаются, за исключением тех случаев, когда они вложены в ИГ-участника.
- (5) {Тогда Доренко **покинул** пост главного редактора радиостанции «Русская служба новостей»} [...]
- (6) {Высокопоставленный северокорейский политик, являвшийся де-факто «номером два» в иерархии, **уволен** с должности руководителя Главного политического управления армии}.

- (7) *Беседем с членом РСМ и АСО России Арсеном Гаджиевым, {который назначен на должность министра по делам молодежи Республики Дагестан}.*
2. Если вершина МК — полное причастие, образующее обособленный оборот, то область разметки события ограничивается этим обособленным оборотом.
- (8) *Также к структуре Дирекции присоединились Илья Дронов (экс-глава российского LiveJournal) и Андрей Ильичев, {занявший должность заместителя Директора по продуктам}.*
3. Если вершина МК — деепричастие, то область разметки события ограничивается деепричастным оборотом.
- (9) *Т. Мокридж возглавил News International в июле 2011г., {сменив на этой должности Ребекку Брукс, которая была одним из ключевых фигурантов дела News of the World}.*
4. Если вершина МК — полное небособленное причастие, то область разметки события включает само причастие, его зависимые (при их наличии) и определяемую причастием именную группу.
- (10) *Связаться с {покинувшей свой пост ведущей}, а также получить какие-либо комментарии у руководства RT пока не удалось.*
5. Если вершина МК — инфинитив, подчиненный глаголу (глагольному словосочетанию) в личной или неличной форме, то в область разметки события включается не только инфинитив с зависимыми, но и клауза, подчиняющая инфинитив.
- (11) *{Губернатора Волынской области вынудили уйти в отставку}.*
6. Если вершина МК — отглагольное существительное, то область разметки события включает только вершину МК с зависимыми.
- (12) *Пока что им удалось добиться {отставки главы Волынской области Бориса Климчука}.*
- (13) *[...] после {назначения новым гендиректором РСН Арама Габрелянова}.*

3.3. Разметка ИГ-участников

В качестве участников события размечаются именные группы (далее ИГ), являющиеся поверхностным выражением соответствующих актантов целевой пропозиции и находящиеся в пределах области разметки упоминания, а также удовлетворяющие дополнительным ограничениям (см. ниже).

Синтаксические нули в качестве участников события не размечаются.

Участники «занимающий» и «оставляющий»

Разметке в качестве участника любого из этих типов подлежит ИГ (включая местоимение и сочиненную группу) с референцией к лицу или нескольким лицам.

Участник «каузатор»

В качестве каузатора может быть размечена ИГ, удовлетворяющая ограничениям для занимающего и оставляющего, а также ИГ с референцией к организации или геополитической единице.

Участник «должность»

В качестве участника *должность* размечается ИГ (в т. ч. местоимение и сочиненная группа) с референцией к должности или нескольким должностям, включая, если есть, слово-вершину *должность, пост, место* и т. п. В порядке исключения как *должность* в контексте назначения размечается инфинитивная группа «*исполнять обязанности + название должности* в род. падеже».

Если название должности входит в (или совпадает с) ИГ занимающего или оставляющего эту должность, то в составе такой ИГ размечаются и участник-должность, и участник-лицо:

(14) [[*Главед информгентства «Новый регион»*]_{должность} *Иванов*]_{оставляющий}
уволился из-за давления властей.

Участник «возглавляемый объект»

В качестве возглавляемого объекта может быть размечена ИГ с референцией к организации или геополитической единице.

3.3.1. Границы размечаемой ИГ-участника

ИГ, обозначающая актанта ситуации, размечается целиком, включая обособленные определения и приложения. Последовательной альтернативы этому решению пока найти не удалось, ни по формальным, ни по содержательным критериям. Примеры:

(15) [*Алексей Соколов из администрации президента*]_{занимающий} назначен
[*замглавы Минкомсвязи*]_{должность}.

- (16) [Высокопоставленный северокорейский политик, являвшийся де-факто «номером два» в иерархии]_{ОСТАВЛЯЮЩИЙ} уволен с [должности руководителя Главного политического управления армии]_{ДОЛЖНОСТЬ}.
- (17) В понедельник [президент]_{КАУЗАТОР} принял отставку [еще одного губернатора — главы Мурманской области Марины Ковтун]_{ОСТАВЛЯЮЩИЙ}.
- (18) [Пост президента]_{ДОЛЖНОСТЬ} занял [Федерико Франко, бывший вице-президент страны]_{ЗАНИМАЮЩИЙ}.

3.3.2. Особенности разметки участников, вытекающие из способа определения границ упоминания

Принятый нами подход к определению области разметки упоминания накладывает определенные ограничения на отбор ИГ, размечаемых как ИГ-участники.

1. Если МК входит в относительное придаточное и одну из целевых валентностей МК заполняет местоимение (*который, кто*), то именно оно, а не его антецедент в главном предложении (вершина придаточного), размечается в качестве участника.

(19) Аксенов ранее объявил о создании в республике собственных военно-морских сил, {*которые*]_{ВОЗГЛАВЛЯЕМЫЙ ОБЪЕКТ} возглавил бывший главнокомандующий ВМС Украины Денис Березовский, присягнувший на верность крымскому народу}.

2. Если вершина МК — полное причастие, образующее обособленный оборот, или деепричастие с зависимыми, то антецедент (или постцедент) нулевого подлежащего при причастии или деепричастии в качестве участника данного события не размечается (поскольку находится в главной клаузе).

(20) Борис Дубровский, {будучи назначенным на пост исполняющего обязанности региона президентом страны}, получил серьёзный вотум доверия.

3. Если вершина МК — инфинитив, подчиненный глаголу (глагольному словосочетанию) в личной или неличной форме, то ИГ, соответствующая глубинно-синтаксическому подлежащему инфинитива, может быть размечена в качестве участника (но только если находится в пределах области разметки упоминания, как в примере 21; в примере 22 подлежащее инфинитива в качестве участника отставки не размечается).

(21) Сегодня же стало известно, {что оставить свою должность согласился [глава волинского облсовета Владимир Войтович]_{ОСТАВЛЯЮЩИЙ}}.

(22) *Председатель Заксобрания Севастополя Алексей Чалый заявил сегодня на сессии регионального парламента, {что намерен уйти в отставку}.*

3.4. Особые случаи: сочинение, эллипсис, множественные актанты

3.4.1. Эллипсис МК при сочинении

Упоминания с эллипсисом МК при сочинении клауз подлежат разметке. Область разметки упоминания с эллиптической МК определяется по аналогии с неэллиптическим упоминанием. В качестве маркера размечается тире (при его отсутствии допускается разметка текстового пробела в предполагаемом месте пропуса МК).

(23) {[Министром связи]_{должность} **назначен** [Руди Эриво]_{занимающий}},
{[министром внутренних дел]_{должность} = [Реджинальд Дельва]_{занимающий}},
{[министром транспорта]_{должность} = [Жак Руссо]_{занимающий}}.

3.4.2. Опущение/усечение ИГ-участника

Упоминание с эллипсисом ИГ, соответствующей выделенному участнику, не размечается.

Упоминания с эллипсисом других ИГ-участников размечаются, но опущенные ИГ не реконструируются и не размечаются.

(24) *В 1991 г. **назначен** [директором Кузнечного завода
ОАО «Курганмашзавод»]_{должность}.*

Антецедент/постцедент ИГ, подвергшейся в упоминании (в конъюнкте, содержащем МК) сочинительному сокращению, в качестве ИГ-участника данного упоминания не размечается:

(25) *В октябре 2008 года полковник Иван получил повышение по службе {и **был назначен** на [должность начальника штаба второго военного округа]_{должность}}.*

В том случае, если одна из ИГ-участников подверглась частичному сокращению (эллипсису), упоминание размечается, только если оставшаяся часть ИГ содержит её вершину. Пример упоминания, которое не размечается:

(26) {[Исполняющим обязанности главы ГТРК «Крым»] **назначен** [первый замгендиректора Игорь Кульчик]_{занимающий}}, Севастопольской РГТРК — замгендиректора по технике Александр Слободенюк.

Ср. пример, подлежащий разметке:

(27) {[Директором завода]_{должность} **назначен** [Петров]_{занимающий}},
{[заместителем]_{должность} = [Сидоров]_{занимающий}}.

3.4.3. Упоминания с множественными актантами

Поскольку реконструировать в таких упоминаниях несколько упоминаний с единичными участниками, оставаясь в рамках поверхностной разметки, далеко не всегда возможно, было принято решение размечать упоминания с множественными актантами только в том случае, если одноименные участники в них образуют единую ИГ, например:

(28) [Сопредседателями комитета]_{должность} **стали** [Джим Мессина, руководитель избирательной кампании Барака Обамы на выборах 2012 года, и бывший губернатор штата Мичиган Дженнифер Грэнхолм, которая ранее призывала Клинтон выдвинуть свою кандидатуру в 2016 году]_{занимающий}

(29) [Своих должностей]_{должность} **лишились** [заместители главы Минобороны Елена Козлова, курировавшая при Сердюкове контрольно-финансовую инспекцию и медицину, а также Дмитрий Чушкин, отвечавший в ведомстве за информационно-техническую политику]_{оставляющий}

Для сравнения — пример, который сейчас не размечается, поскольку упоминания должностей (а это выделенный участник) не образуют во фрагменте единой ИГ:

(30) Государственный комитет телевидения и радиовещания Украины издал приказы об **увольнении** гендиректора государственной телерадиокомпании «Крым» Степана Гулеватого и гендиректора Севастопольской региональной ГТРК Сергея Норманского.

4. Заключение

Размеченный нами корпус Situations-1000 по принципам разметки близок к ACE, но имеет некоторые отличия (Табл. 3).

Таблица 3. Сравнение разметок ACE и Situations-1000

Особенность разметки	ACE	Situations-1000
Разметка ограниченного перечня событий	да	да
Разметка места упоминания события в тексте	да	да
Разметка участников события	да	да
Разметка информации об участниках в структурированном виде	нет	нет
Область поиска информации об участниках	предложение	клауза, содержащая маркерную конструкцию

Особенность разметки	ACE	Situations-1000
Разметка места упоминания участника события в тексте	да	да
Локализация события во времени	да	нет
Локализация события в пространстве	да	нет
С учётом модальности	да, размечают присутствие модальности	частично (маркер сомнения)
С учётом полярности	да, размечают	да, отрицания не размечаются
С учётом общности	да, обобщения размечают	да, обобщения не размечают
С учётом стадии	нет	нет
Разметка событий, выраженных местоимениями и дескрипциями с отсылкой к ранее упомянутому событию-антецеденту	да	нет

Корпуса аннотационного типа (с привязкой разметки к тексту), несмотря на сложность их создания, обладают некоторыми достоинствами, которых оказываются лишены разметки типа MUC. Разметка непосредственно «по тексту» полезна не только для отладки и оценки отдельных алгоритмов и этапов решения задачи, но и с прикладной точки зрения. Зная, что алгоритмы могут ошибаться, и желая проверить полученный системой результат, пользователь прикладной системы имеет возможность обратиться не просто к тексту, а к предложению (нескольким предложениям), содержащим извлеченную информацию.

Situations-1000 свободно доступен для использования в научно-исследовательских целях по URL: <http://ai-center.botik.ru/Airec/index.php/ru/collections/33-situations-1000>

Литература

1. *Chinchor N., Marsh E.*, (1998), MUC-7 Information Extraction Task Definition, Proceedings of the Seventh Message Understanding Conference, Fairfax, available at: http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/ie_task.html.
2. *Linguistic Data Consortium*, (2005), ACE (Automatic Content Extraction) English Annotation Guidelines for Events. Version 5.4.3.
3. *Saurí R., Littman J., Gaizauskas R., Setzer A., Pustejovsky J.* (2006), TimeML Annotation Guidelines. Version 1.2.1, available at: http://www.timeml.org/publications/timeMLdocs/annguide_1.2.1.pdf.
4. *Saurí R., Goldberg L., Verhagen M., Pustejovsky J.*, (2009), Annotating Events in English. TimeML Annotation Guidelines, Version TempEval-2010, available at: <http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/EventGuidelines-050409.pdf>.