

# Извлечение информации о ситуациях отставок-назначений в новостных текстах. Опыт разметки коллекции. Результаты тестирования.

© Н.А.Власова

Исследовательский Центр Искусственного Интеллекта  
Института Программных Систем РАН имени А.К.Айламазяна,  
г. Переславль-Залесский  
nathalie.vlassova@gmail.com

## Аннотация

В настоящей работе описан эксперимент по разметке коллекции новостных текстов с целью оценить эффективность подхода к извлечению информации о ситуациях отставки-назначения в системе ИСИДА-Т. Система ИСИДА-Т разрабатывается в ИПС РАН в течение нескольких лет и реализует инженерный подход к извлечению информации из текстов. В данной статье описывается попытка реализации подхода, описание тестовой коллекции, а также приводятся полученные результаты.

## 1 Введение

В ИЦИИ ИПС РАН в рамках проекта ИСИДА-Т [1],[6],[7] ведётся работа по извлечению из новостных текстов информации о ситуациях назначения-отставки. На RCDL-2012 был представлен доклад о концепции извлечения информации о ситуациях отставки-назначения в рамках инженерного подхода, реализуемого в системе ИСИДА-Т [11]. Для экспериментальной оценки эффективности данной разработки была размечена коллекция из 231 документа (новостные сообщения), в которых встречается 868 ситуаций отставки-назначения. Была проведена настройка системы правил, более детально проработаны контексты, описывающие целевые ситуации, и отлажено программное обеспечение. Были получены первые числовые данные, позволяющие судить об эффективности работы системы при извлечении ситуаций. При написании статьи были учтены вопросы и замечания, полученные автором на RCDL-2012.

## 2 Понятие текстовой ситуации и принципы разметки текстов. Исходные данные для анализа ситуаций

Труды 15-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2013, Ярославль, Россия, 14-17 октября 2013 г.

Прежде всего необходимо определить, что же будет считаться ситуацией, информация о которой должна быть извлечена из текста. Будем различать текстовую ситуацию и ситуацию внетекстовую. **Внетекстовая** ситуация – это ситуация, произошедшая во внеязыковой действительности. Например, факт, что 19 апреля 2013 года правительство России отправило в отставку руководителя “Почты России” Александра Киселёва, - это факт из внеязыковой действительности. В ситуации увольнения – три участника:

кто уволил (Правительство России) – **1-й участник**,  
кого уволил (Александра Киселёва) – **2-й участник**,  
с какой должности (руководителя “Почты России”) – **3-й участник**.

Вне текста наше знание о ситуации базируется на информации о её участниках и отношениях между ними. В текстах об этом событии может быть рассказано по-разному:

- 1. Решение об отставке руководителя "Почты России" правительством принято.*
- 2. Правительство России приняло решение об отставке главы "Почты России" Александра Киселёва*
- 3. Бывший глава «Почты России» Александр Киселев получит после отставки с должности более 3 миллионов рублей,*
- 4. На фоне коллапса, возникшего в работе «Почты России», правительство отправило в отставку генерального директора предприятия Александра Киселёва*
- 5. Александр Киселев, об отставке которого с поста генерального директора «Почты России»*

стало известно 19 апреля, получит после увольнения чуть более трех миллионов рублей.

6. Отставку Александра Киселева спровоцировал «посылочный коллапс» в аэропортах московского авиаузла.

7. Александр Киселев в ближайшее время может покинуть пост главы «Почты России»

8. За что увольняют директора «Почты России» Александра Киселёва?

9. Александр Киселев, уволенный с поста генерального директора "Почты России", после ухода в отставку получит золотой парашют на 3 млн рублей.

10. Уволен глава «Почты России»

(все примеры взяты с новостных порталов в сети Интернет)

Легко заметить, что в новостных текстах далеко не всегда упоминаются все три участника ситуации. Чаще всего попадают контексты с 2 участниками. И почти никогда не упоминается дата события. Кроме того, ситуация может быть описана не одним предложением, а несколькими (см. пример 2). То есть в тексте может не содержаться исчерпывающая информация о ситуации. А в отдельном предложении это вообще встречается редко. Таким образом, построить полную картину, ограничившись рамками одного предложения, практически невозможно. Более того, в тексте может говориться о возможной ситуации - ситуации, которая не произошла в действительности, и, возможно, не произойдёт (например: *на заседании речь шла о возможной отставке министра образования*).

При разметке ситуаций назначения-отставки в текстах коллекции для экспериментальной работы необходимо чётко представлять, что будет считаться ситуацией. Под **текстовой ситуацией** мы будем понимать ситуацию, описанную в одном предложении (возможно, не произошедшую в действительности, но просто упомянутую в тексте) и выраженную с помощью целевого слова-маркера ситуации (слова, называющего ситуацию, – **уволнил, назначил, отставка, назначив** и т.п.) и именных групп, описывающих участников ситуации. При этом мы исходим из предположения, что участники расположены контактно справа и слева от маркера ситуации (между участником и словом-ситуацией допускаются наречия, указания на время, частицы).

Наша задача – максимально полно и точно извлечь информацию из текстовых ситуаций, то есть определить слово-ситуацию, собрать всех участников и правильно распределить роли. В дальнейшем, располагая информацией о текстовых ситуациях в разных предложениях текста и зная их последовательность, мы сможем приблизиться к полной информации о внетекстовой ситуации.

В силу особенностей работы всех модулей системы ИСИДА-Т не размечаются и не включаются в общую группу ситуации:

- если хотя бы один из участников выражен именной группой с главным словом во множественном числе,

- если хотя бы один из участников выражается несколькими синтаксически однородными именными группами (*уволнил Иванова и Петрова*),

- в случае эллипсиса (*Иванова назначили директором, а Петрова – его заместителем*),

- при наличии отрицания у целевого слова.

Конструкции с модальными глаголами с учётом определения текстовой ситуации учитываются и размечаются.

Итогом извлечения информации о текстовой ситуации должен стать набор отношений, связывающих ситуацию, обозначенную в тексте словом-маркером, и её участников, выраженных в тексте именными группами. Рассмотрим пример:

*Президент Украины Виктор Янукович отправил в отставку премьер-министра Николая Азарова.*

В результате работы модуля извлечения информации о ситуации должен получиться следующий набор отношений (см. рис. 1)

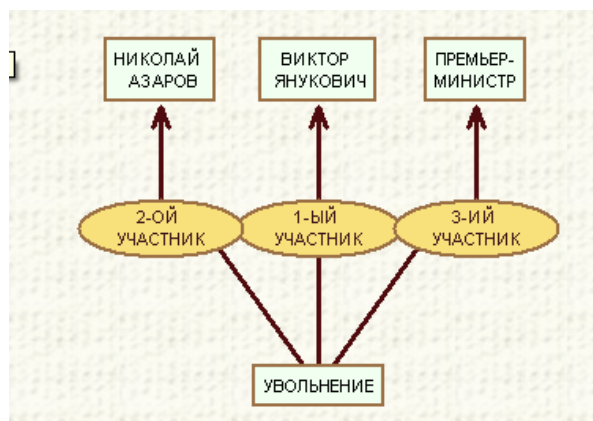


Рис.1. Представление информации, полученное в результате работы модуля анализа ситуаций

Какой уже извлечённой информацией мы можем располагать перед началом работы модуля извлечения ситуаций? Про каждое слово в предложении известна его морфологические характеристики, с помощью ресурса знаний системы ИСИДА-Т [6],[8] и специального модуля извлечения имён построены специальные аннотации для имён людей (включая разбиение на имя, отчество и фамилию), названий должностей, организаций, геополитических единиц. Отдельными аннотациями помечаются временные указатели (включая составные – например, *вчера вечером, в следующем году* и т.п.). Возможности синтаксического анализатора системы ИСИДА-Т ограничиваются анализом именных групп с зависимыми прилагательными или существительными в родительном падеже, а также с приложениями, включая предложные группы, в которые могут входить более простые именные группы. Для каждого слово известны его графематические характеристики (написано с большой или маленькой буквы, латинский или русский шрифт – вся информация о написании слова в тексте.)

Основная идея описываемого подхода – извлечь всё, что возможно извлечь из текста средствами локального микросинтаксиса и информацией об извлечённых сущностях.

Есть “ядерные” вещи, которые можно классифицировать и систематизировать, а есть словарно-текстовые, которые надо будет задать списком. Конечно, есть контексты с текстовыми ситуациями, которые средствами системы ИСИДА-Т не могут быть обработаны. Например, такие:

1. *им стал долгое время проработавший руководителем клиники, в Москве, заведующий кафедры ортопедической стоматологии факультета последипломного образования Московского государственного медико-стоматологического университета им. А.И.Евдокимова, доктор медицинских наук, профессор, человек, который знает дагестанскую и современную мировую медицину — Танка Ибрагимов* – анализ именной группы, которая называет участника ситуации получения должности, выходит за рамки возможностей синтаксического анализатора системы ИСИДА-Т.

2. *Президент Украины Виктор Янукович сменил главу Центрального управления Службы безопасности страны, назначив на этот*

*пост человека, якобы близкого к своему сыну* – в данном примере участник отделён от слова-ситуации НАЗНАЧИВ синтаксическими группами, которые не являются участниками ситуации.

В разметке участвовали следующие ситуации:

@увольнение, @назначение,  
@получение\_должности, @уход\_с\_поста.

Аннотация разметки, атрибуты которой сравниваются с результатами, полученными в ходе обработки текста, содержит следующие атрибуты:

1. **Situation** – название ситуации
2. **first** – первый участник (при разметке сюда записывается главное слово именной группы, соответствующей по значению данному участнику ситуации);
3. **second** – второй участник;
4. **third** – третий участник.

Если в текстовой ситуации нет информации о каком-либо из участников, соответствующий атрибут остаётся незаполненным.

См. на рисунке 2 пример аннотации разметки:

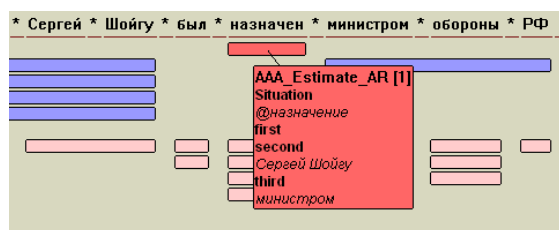


Рис.2. Пример аннотации разметки.

Одной из основных задач при разметке и тестировании коллекции текстов было оценить, какой результат мы сможем получить при существующих возможностях системы для текстовых ситуаций. Принципы разметки ситуаций для тестирования примерно соответствуют принципам разметки событий ACE [13] в том, что помечаются ключевые слова и они могут быть выражены не только глаголом в личной форме, но и отглагольным существительным, причастием, деепричастием, помечаются целиком участники. Отличия же заключаются в том, что не размечаются ситуации, обозначенные местоимением (например, *она (отставка) произошла сегодня утром*).

### 3 Классификация слов-маркеров ситуации

Ниже приводятся слова, которые маркируют ситуации, участвующие в разметке и тестировании:

Ситуация	Слова-маркеры
Увольнение	увольнение, уволить, увольнять  снятие, снять, снимать  освобождение, освободить, освобождать  отстранение, отстранить, отстранять  отправить, отправлять (в отставку)  прекратить, прекращать (полномочия)
Назначение	назначение, назначить, назначать  Переназначение, переназначить, переназначать  ставить  поставить  утверждение, утвердить, утверждать
Получение_должности	стать, становиться  сохранить  занять  получение, получить, получать  назначаться  приход, придти, приходиться (на должность)  Быть (формы будущего времени)

Уход_в_отставку	уход, уйти, уходить  покинуть, покидать  оставить, оставлять  лишиться, лишаться  подать, подавать (в отставку)  сложить, складывать (полномочия)
-----------------	---

В следующей таблице приведена классификация контекстов со словом-маркером ситуации в зависимости от формы этого слова:

	Форма целевого слова	Образец (на примере глагола увольнять)
1	Глагол в личной форме (в единственном числе)	Уволил, увольняет и т.п.
2	Возвратный глагол	Уволился
3	Глагол в 3 лице мн.числа	Уволили, увольняют
4	Глагол в инфинитиве	Уволить (например, в приказе или с модальным глаголом)
5	Причастие действительное	Уволивший, увольняющий (оч. редко)
6	Причастие страдательное	Уволенный, увольняемый
7	Отглагольное существительное	увольнение
8	Деепричастие	Уволив, увольняя
9	Относительное предложение	Иванов, который уволил

### 4 Классификация участников ситуации

Прежде всего, нужно определить, какие именные группы будут считаться участниками ситуации. Помимо именных групп, которые задают непосредственно участников ситуаций, - это

именные группы, которые могут находиться между целевым словом и именной группой основного участника. В свою очередь такие группы делятся на **вспомогательные** (они нужны для того, чтобы целевое слово указывало на ситуацию, например, глагол **уйти** обозначает ситуацию @уход\_в\_отставку только при наличии зависимой именной группы с главным словом ОТСТАВКА) и **дополнительные** (например, указание на время, место, организацию – такая информация не относится к основным участникам, но при этом такие именные группы могут отделять группу основного участника от целевого слова). Кроме того, основные участники делятся на **атомарные** и **неатомарные** группы. Атомарные именные группы указывают на одного основного участника ситуации, а неатомарные – на двух. Например:

1. *Директор предприятия Виктор Сергеев уволил своего заместителя Романа Николаева.*
2. *Сегодня был уволен заместитель директора Роман Николаев.*

В рассматриваемых примерах 2-й (кого уволили) и 3-й (с какой должности) участники выражены в одной именной группе.

Итак, ниже приводится классификация участников ситуаций отставки-назначения:

- I. Основные. Атомарные
  1. Именные группы с главным словом ПОСТ, ДОЛЖНОСТЬ, КРЕСЛО, МЕСТО + название должности или + местоимение ЕГО, ЕЁ, СВОЙ, ЭТОТ;
  2. Именная группа – имя человека;
  3. Именная группа – название должности
  4. Именные группы, устроенные особым образом и выражающие первого участника (решением президента, по распоряжению правительства и т.п.)
- II. Основные. Неатомарные
  1. Имя + должность (могут называть 2-го и 3-го участников), при этом должность может называть прежнее место работы
- III. Вспомогательные. Именные группы – фиксированные выражения при ситуациях назначения-отставки (задаются списком) – *по собственному желанию, по собственной инициативе* и т.п.
- IV. Дополнительные
  1. Именные группы – указание на время

## 2. Именные предложные группы с указанием на организацию

Кроме общей классификации, отдельными пометами снабжаются такие слова в именных группах, которые помогают правильно распределить участников ситуации. Это, например, прилагательные - **бывший, новый, экс-, действующий**.

В следующих двух таблицах представлены слова-маркеры ситуаций и участники, наличие которых обязательно для формирования контекста ситуации (обязательные участники). Отбор таких слов проводился эмпирически на основе анализа текстов об отставках и назначениях на новостных порталах в сети Интернет. Конечно, такой список не может описывать всё языковое многообразие конструкций, описывающих целевые ситуации, но, безусловно, включает в себя самые частотные и типичные языковые выражения.

Ситуации с тремя участниками - @назначение, @увольнение

Целевое слово	Обязательный участник
Уволить, увольнять, увольнение	-
Снять, снимать, снятие	должность
Отправить, отправлять	отставка
прекратить	полномочия
Освободить, освобождать, освобождение	- должность
Отстранить, отстранять, отстранение	должность
Назначить, назначать, назначение, переназначить	Должность, организация или геополитическая единица
Поставить, ставить	Должность, организация или геополитическая единица
утвердить	должность

Ситуации с двумя участниками (первый никогда не выражен при данных целевых словах) –  
 @получение\_должности, @уход\_в\_отставку.

Целевое слово	Обязательный участник
Стать, становиться	должность
Сохранить, сохранять, сохранение	должность
Получить, получать, получение	должность
Быть	должность
назначаться	должность
Уходить, уйти, уход	Должность или отставка
Покинуть, покидать	Должность
Оставить, оставлять	Должность
Подать, подавать	Отставка
Сложить, сложение	полномочия
Лишиться, лишение	Должность или полномочия
уволиться	-

## 5 Описание алгоритма извлечения информации о ситуациях

Именные группы участников ситуаций собираются последовательно справа и слева от целевого слова. Работа правил организована следующим образом: правила собраны в фазы, которые выполняются последовательно одна за другой. Всего есть 9 фаз правил. Информация о работе фаз приведена в таблице ниже:

№	Описание работы фазы	Аннотации, которые строятся правилами
1	Словам, описывающим ситуацию, ставятся в соответствие специальные аннотации, в атрибуты	<b>AVerb</b> – аннотация, маркирующая слово-ситуацию

	которых в следующих фазах будет записываться информация об участниках. Эти аннотации также играют роль маркера для фаз, в которых происходит поиск потенциальных участников ситуации	
2	Построение аннотаций для потенциальных участников, выраженных нестандартными конструкциями	<b>UKAZ</b> – аннотация для именных групп с главным словом “указ”, “решение”, “распоряжение” и т.п.  <b>POST</b> – аннотация для именных групп с главным словом “пост”, “должность”, “место”, “кресло”.
3	Построение аннотаций для ключевых слов – 2-й этап	<b>AVerb</b> – аннотация, маркирующая ситуацию
4	Построение аннотаций для участников ситуаций, расположенных непосредственно справа и слева от целевого слова	<b>AALink</b> – аннотации для именных групп – потенциальных участников ситуации, которые выделяются правилами фазы непосредственно справа и слева от аннотации <b>AVerb</b>
5	Построение аннотаций для участников ситуаций, расположенных справа или слева от участников, определенных в фазе 4	<b>AALink</b> – аннотации для именных групп – потенциальных участников ситуации, которые выделяются справа и слева от аннотаций <b>AALink</b> , построенных правилами фазы 4 (эти аннотации включают в себя аннотации <b>AALink</b> , построенные ранее)
6	Построение аннотаций для участников ситуаций,	<b>AALink</b> – аннотации для именных групп – потенциальных

	расположенных справа или слева от участников, определённых в фазе 5	участников ситуации, которые выделяются справа и слева от аннотаций <b>AALink</b> , построенных правилами фазы 5 (эти аннотации включают в себя аннотации <b>AALink</b> , построенные ранее)
7	Добавление в аннотации <b>AVerb</b> атрибутов, соответствующих найденным потенциальным участникам (обязательные участники)	Аннотация <b>AVerb</b> получает новые атрибуты
8	Добавление в аннотацию <b>AVerb</b> атрибутов, соответствующих найденным потенциальным участникам (только для тех целевых слов, у которых на предыдущем этапе был опознан обязательный участник или обязательный участник не нужен)	Аннотация <b>AVerb</b> получает новые атрибуты
9	Определение отношений между целевым словом и найденным участником ситуации (первый участник, второй участник, третий участник)	<b>ALink</b> – аннотации, которые задают отношение между словом, маркирующим ситуацию, и участником, – приписывание номера участнику ситуации

Рассмотрим работу алгоритма на примере

*Советником Шойгу стала телеведущая Мария Китаева.*

Эталон разметки выглядит так:

Советником Шойгу стала телеведущая Мария Китаева

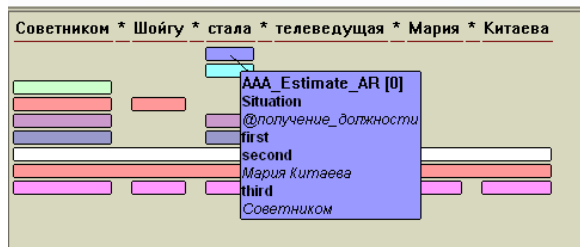


Рис. 3 Эталон разметки для разбираемой ситуации.

Результат работы фазы 1:

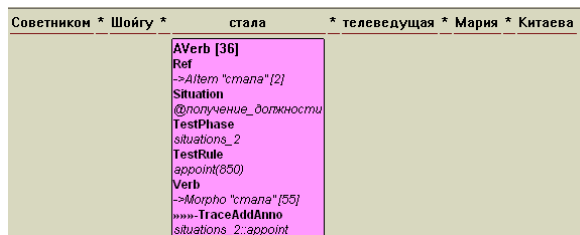


Рис.4. Построение аннотации **AVerb**, соответствующей слову-маркеру ситуации.

Построена аннотация **AVerb**, в атрибутах которой есть ссылка на ситуацию, которую маркирует эта аннотация ( в данном примере - @получение должности), есть атрибут, по ссылке из которого можно попасть в аннотацию, где хранится информация обо всех морфологических характеристиках слова.

Следующий рисунок – результат работы фазы 4 ( правила фаз 2 и 3 в данном контексте не работают)

Советником Шойгу стала телеведущая Мария Китаева

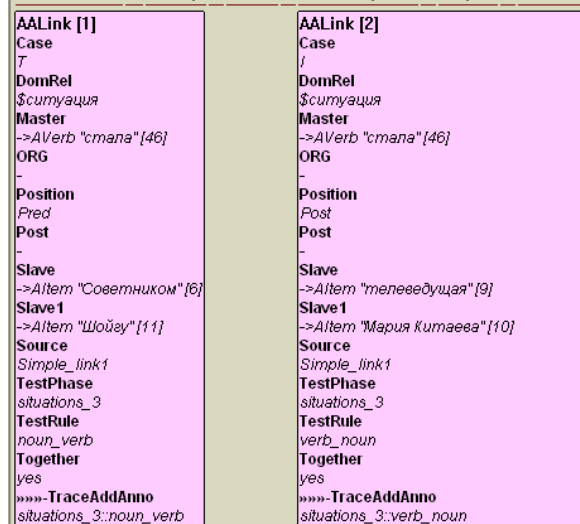


Рис.5. Построение аннотаций **AALink**, соответствующих именным группам участников ситуации.

В результате работы правил 4-й фазы построились две аннотации класса AALink – соответственно слева и справа от слова-маркера ситуации. Эти аннотации соотносятся с именными группами “Советником Шойгу” и “телеведущая Мария Китаева”.

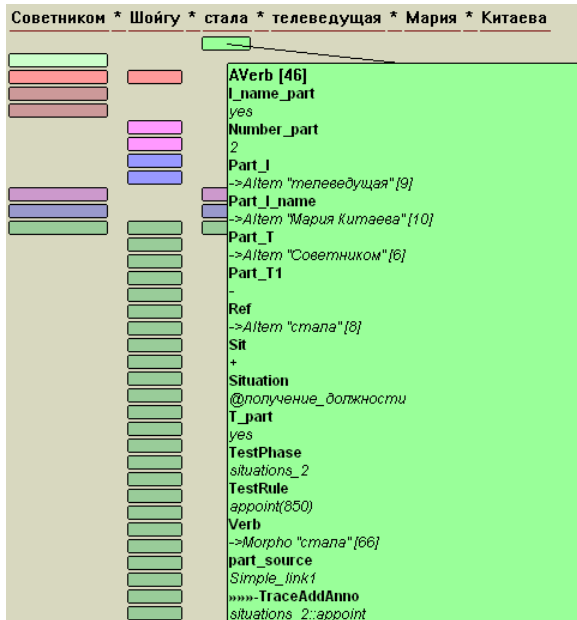


Рис.6. Добавление атрибутов в аннотацию AVerb.

Рисунок 6 иллюстрирует результат работы правил фаз 7 и 8, когда в аннотацию слова-маркера ситуации добавляются атрибуты-ссылки на найденные именные группы потенциальных участников. При этом названия атрибутов отражают падеж и некоторые другие важные для дальнейшей работы характеристики. В разбираемом примере - это Part\_I (“телеведущая”), Part\_I\_name (“Мария Китаева”), Part\_T (“Советником”).

На следующих двух рисунках показаны аннотации ALink, которые построены правилами последней, 9-й, фазы и соответствуют отношениям между словом-маркером ситуации и словом из именной группы, указывающим непосредственно на участника ситуации. Атрибут DomRel показывает номер участника ситуации, атрибут Situation – название ситуации, атрибуты Master и Slave – это ссылки на слово-маркер ситуации и на слово, обозначающее участника.

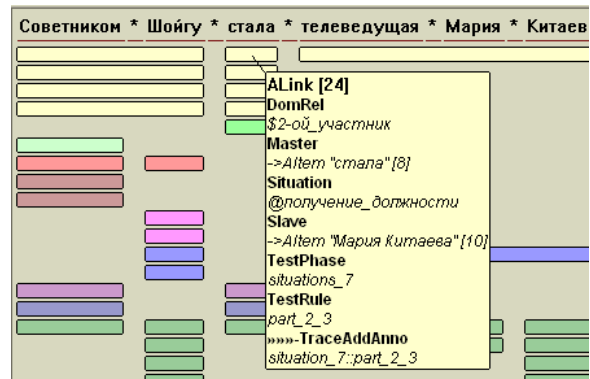


Рис.7. Построение аннотации ALink, моделирующей отношение между словом-маркером ситуации и её 2-м участником.

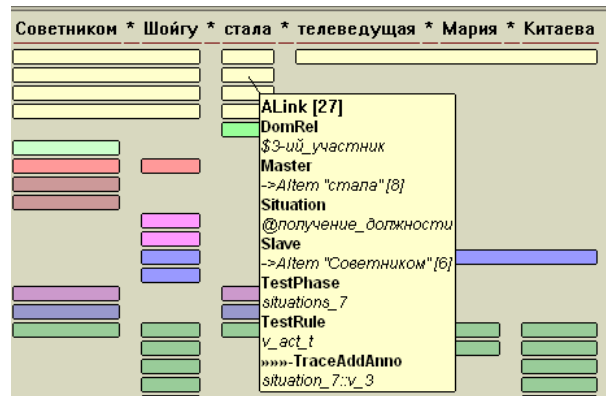


Рис.8. Построение аннотации ALink, моделирующей отношение между словом-маркером ситуации и её 3-м участником.

Наконец, на рисунке 9 приведён результат работы алгоритма извлечения информации о ситуации:



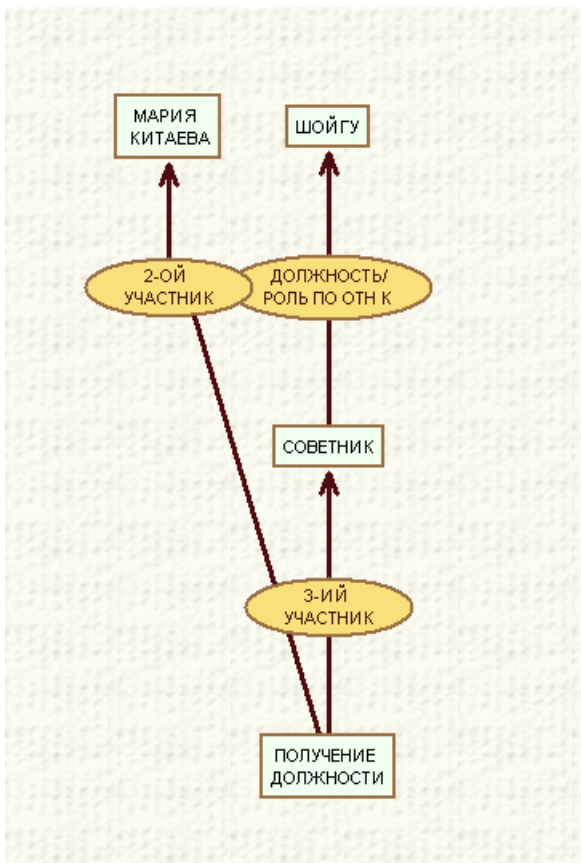


Рис.9. Графическое представление результатов извлечения.

## 6 Результаты тестирования работы алгоритма на размеченной коллекции.

Для проверки эффективности работы системы была размечена тестовая коллекция из 231 текста (тексты взяты на новостных порталах в сети Интернет). В текстах встретились и были размечены описанным выше способом 868 ситуаций отставки-назначения. Результаты тестирования приведены в таблице ниже:

Точность	0,76
Полнота	0,72
F-мера	0,74

При оценке результатов необходимо иметь в виду, что данные по всем трём оценкам приводятся по ситуациям, у которых при анализе атрибуты полностью совпали с атрибутами эталона.

По всей видимости, данные результаты отражают технологический предел возможностей системы ИСИДА-Т на данном этапе разработки. На результат анализа ситуаций оказывают влияние эффективность работы алгоритма выявления имён (F-мера – 0,95), а также особенности синтаксических конструкций, задающих ситуации.

## Заключение

В результате работы с размеченной коллекцией были получены числовые данные, позволяющие оценить эффективность разработанного подхода к извлечению информации о ситуациях отставки-назначения. На данном этапе этот результат позволяет надеяться, что при развитии разработок в направлении анализа всего текста и разрешения кореферентности будут улучшены показатели извлечения информации о ситуациях и мы сможем в большей степени приблизить наши знания, полученные из текста с помощью автоматического извлечения, к знаниям, которые получает читающий человек.

Работа выполнена при поддержке РФФИ, грант № 13-06-00483а.

## Литература

- [1] Александровский Д.А., Кормалев Д.А., Кормалева М.С., Куршев Е.П., Сулейманова Е.А., Трофимов И.В. Развитие средств аналитической обработки текста в системе ИСИДА-Т // Тр. Десятой нац. конф. по искусственному интеллекту с междунар. участием КИИ-2006, Обнинск, 25-28 сентября 2006 г.: В 3 т. — М.: Физматлит, 2006. — Т. 2. — С. 555—563.
- [2] Гершензон Л. М., Ножов И. М., Панкратов Д. В. Система извлечения и поиска структурированной информации из больших текстовых массивов СМИ. Архитектурные и лингвистические особенности. Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог'2005» (Звенигород, 1–6 июня, 2005 г.) / Под ред. И. М. Кобозевой, А. С. Нариньяни, В. П. Селегея. — М.: Наука, 2005.
- [3] Ермаков А.Е. Автоматическое извлечение фактов из текстов досье: опыт установления анафорических связей. Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции "Диалог 2007". – Москва, Наука, 2007
- [4] Ермаков А.Е. Извлечение знаний из текста и их обработка: состояние и перспективы. Информационные технологии 2009, № 7
- [5] Киселев С.Л., Ермаков А.Е., Плешко В.В. Поиск фактов в тексте естественного языка на основе сетевых описаний. Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2004. — М.: Наука, 2004.
- [6] Кормалев Д.А., Куршев Е.П., Сулейманова Е.А., Трофимов И.В. Архитектура

- инструментальных средств систем извлечения информации из текстов. Труды международной конференции "Программные системы: теория и приложения", Переславль-Залесский, М.: Физматлит, 2004, т.2, с.49—70
- [7] Кормалев Д.А., Куршев Е.П., Сулейманова Е.А., Трофимов И.В. Извлечение данных из текста. Анализ ситуаций ньюсмейкинга. Труды Восьмой национальной конференции по искусственному интеллекту с международным участием КИИ-2002. Москва, Физматлит, 2002, с. 199-206
- [8] Кормалев Д.А., Куршев Е.П., Сулейманова Е.А., Трофимов И.В.. Технология извлечения информации, из текстов, основанная на знаниях. Программные продукты и системы, 2009, №2
- [9] Куршев Е.П., Сулейманова Е.А. Ресурсы предметных знаний в системах интеллектуального анализа текста // Тр. междунар. конф. «Программные системы: теория и приложения», ИПС РАН, Переславль-Залесский, октябрь 2006 г.: В 2 т. — М.: Физматлит, 2006. — Т.1. — С. 379—390.
- [10] <http://www.mlg.ru/>
- [11] Власова Н.А. Подход к автоматическому извлечению информации о назначениях и отставках лиц (на материале новостных сообщений) // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. XIV Всероссийская научная конференция RCDL-2012. Труды конференции. — Переславль-Залесский : Университет города Переславля, 2012. — С. 374—378.
- [12] Котельников Д.С., Лукашевич Н.В. Итерационное извлечение шаблонов описания событий по новостным кластерам Электронные библиотеки: перспективные методы и технологии, электронные коллекции. XIV Всероссийская научная конференция RCDL-2012. Труды конференции. — Переславль-Залесский : Университет города Переславля, 2012. — С. 362—373.
- [13] <http://projects.ldc.upenn.edu/ace/docs/> - электронный документ – принципы разметки ACE

**Extracting information on appointments and dismissals from news texts. An experience in developing an annotated corpus. Testing results.**

Natalia Vlasova

The paper describes an experiment on annotating a collection of Russian-language news texts for an information extraction task. The objective was to evaluate the efficiency of an approach to solving the appointment-dismissal task, which is implemented in the ISIDA-T software. ISIDA-T has been developed during the past decade in the Program Systems Institute, RAS. It is based on a knowledge-engineering approach to information extraction. The paper describes the implementation of information extraction method, the annotation principles, the test collection, and presents some evaluation results.