

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

Наталья Александровна Власова  
Исследовательский центр искусственного интеллекта  
Институт программных систем им. А.К.Айламазяна  
Российской академии наук,  
г. Переславль-Залесский  
nathalie.vlassova@gmail.com

# ИСИДА-Т - система извлечения информации ИЗ ТЕКСТОВ

- Инженерный подход к извлечению информации;
- Пошаговая обработка текста обрабатываемыми ресурсами;
- Частичный синтаксический анализ;
- Правила обработки текста на языке PSL;
- **Ресурс знаний** - словарь и связанная с ним база знаний, представляющая собой семантическую сеть концептов (концепты представляются в виде фреймов) и отношений

# Почему сложно извлекать информацию о ситуациях?

- Такая информация может располагаться неконтактно;
- Синтаксис и лексика конструкций, описывающих ситуацию, очень разнообразны;
- Участники ситуации могут задаваться сложными конструкциями;
- Омонимия
  - падежная – *Шойгу возглавил министерство обороны;*
  - ролевая – *Петров назначен директором*
- Извлечение ситуаций предполагает хорошие решения задач извлечения более низкого уровня:
  - 1) *извлечение именованных сущностей (имена, названия организаций, должностей, геополитических единиц, указаний на время);*
  - 2) *построение отношений между извлечёнными простыми объектами для правильного распределения ролей участников;*
  - 3) *разрешение кореферентности.*

# Задачи извлечения

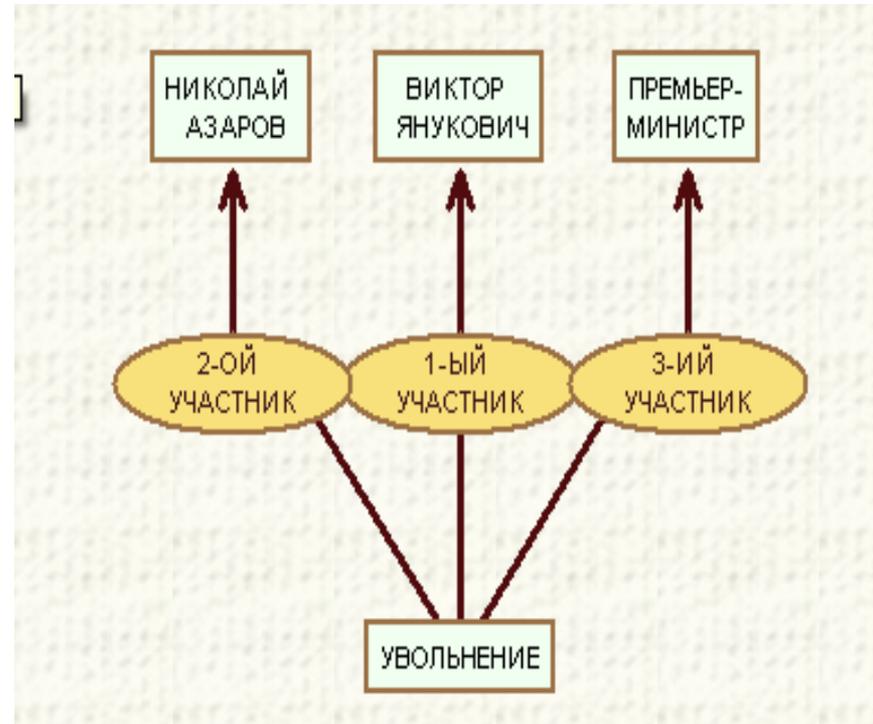
- Максимально полно и точно извлечь информацию о ситуации и её участниках;
- Выстроить систему, легко настраиваемую на разные типы ситуаций.

# Ситуации назначения-отставки

**Внетекстовая** ситуация – это ситуация с тремя (или двумя) основными участниками (пример – ситуация увольнения):

- Кто назначил
- Кого назначил
- На какую должность
- Когда назначил (факультативно)

*Вчера Виктор Янукович назначил на пост премьер-министра Николая Азарова.*



# Примеры

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" .

# Текстовая ситуация

**Текстовая ситуация** – ситуация, представленная в рамках одного предложения с помощью слова-маркера ситуации (глагола, причастия, деепричастия, отглагольного существительного) и именных групп, обозначающих участников ситуации.

- Мэр Москвы **Сергей Собянин** 17 сентября **переназначил** **Александра Кибовского** **главой** департамента культурного наследия столицы.
- **Андрей Шаронов** **станет** **ректором** бизнес-школы "Сколково" .
- Генерал-лейтенант **Олег Макаревич** **назначен** **начальником** штаба Восточного военного округа (ВВО)
- Говорить о серьезных изменениях в игре новокузнецкой команды "Металлург" еще рано, однако **назначение** главным **тренером** **Германа Титова** ее встряхнуло.
- Распоряжением **мэра** Москвы генеральным **директором** ГУП "Мосгортранс" **назначен** **Евгений Михайлов**.

# Примеры ситуаций, которые не будут считаться текстовыми

- Бывший глава «Почты России» Александр Киселев получит после отставки с должности более 3 миллионов рублей.

**Участники ситуации увольнения не входят в синтаксическую группу слова-маркера**

- Президент Украины Виктор Янукович сменил главу Центрального управления Службы безопасности страны, назначив на этот пост человека, якобы близкого к своему сыну.

**Два участника ситуации назначения указаны в другой ситуации – смены в должности, назначение не выстраивается.**

# Принципы разметки текстовых ситуаций

- Размечаются все текстовые ситуации ;  
( оценка точности-полноты извлечения по отношению к **текстовым** ситуациям)
- Участники ситуаций только в единственном числе ;
- Слово-маркер всегда однозначное и обозначает ситуацию назначения-отставки;
- Эллиптические конструкции не размечаются;
- При наличии отрицания у слова-маркера.

Сравнение с разметкой **АСЕ**.

# Пример аннотации разметки



- **Situation** – название ситуации
  - @назначение
  - @увольнение
  - @получение\_должности
  - @уход\_с\_поста
- **first** – первый участник (кто)
- **second** – второй участник (кого);
- **third** – третий участник (куда)  
(для ситуаций @назначение,  
@увольнение)

# Типы маркеров ситуаций

На примере УВОЛЬНЯТЬ, УВОЛИТЬ

- Глагол в личной форме (ед.ч.) - *уволит, увольняет, уволил*
- Глагол в 3л. Мн.ч. - *уволит, уволят*
- Глагол в инфинитиве – *уволить, увольнять*
- Причастие действительное – *уволивший, увольняющий*
- Причастие страдательное – *уволенный, увольняемый*
- Деепричастие – *увольняя, уволив*
- Отглагольное существительное - *увольнение*
- Конструкция с относительным придаточным предложением

*Иванов, который уволил*

*Иванов, которого уволили*

*Должность, с которой уволили Иванова*

# Классификация участников ситуаций (1)

## Участники ситуации

### Основные (непосредственные)

#### 1) Атомарные

**имя человека**

**название должности**

**именная группа с гл. словами**

*ПОСТ, ДОЛЖНОСТЬ, КРЕСЛО, МЕСТО, ПОЛНОМОЧИЯ* (задаются списком)

**особые именные группы для обозначения 1-ого участника**

*(решением (указом) ПРЕЗИДЕНТА, по распоряжению ПРАВИТЕЛЬСТВА)*

#### 2) Неатомарные

**именные группы, обозначающие сразу двух участников ситуации**

*Сегодня был уволен заместитель директора Роман Николаев.*

# Классификация участников ситуаций (2)

## Вспомогательные

**фиксированные выражения при ситуациях назначения-отставки**

(задаются списком)

*по собственному желанию, по собственной инициативе и т.п.*

## Дополнительные

**именные группы, обозначающие время**

*(вчера, в прошлом году, 19 мая 2013 года)*

**именные группы, обозначающие организацию**

# Алгоритм извлечения ситуаций (1)

## Построение аннотации для слова-маркера

Советником	* Шойгу *	стала	* телеведущая *	Мария	* Китаева
		<b>AVerb [36]</b> <b>Ref</b> ->AItem "стала" [2] <b>Situation</b> @получение_должности <b>TestPhase</b> situations_2 <b>TestRule</b> appoint(850) <b>Verb</b> ->Morpho "стала" [55] »»»-TraceAddAnno situations_2::appoint			

В аннотации сохраняется информация о типе ситуации, морфологических характеристиках слова, обозначающего ситуацию

## Алгоритм извлечения ситуаций (2)

Построение аннотаций для  
именных групп – потенциальных  
участников ситуации.

В аннотациях сохраняется  
информация о главном и  
зависимых словах группы,  
морфологических характеристиках  
главного слова, информация о  
положении относительно слова-  
маркера.

Советником * Шойгу * стала * телеведущая * Мария * Китаева	
<b>AALink [1]</b>	<b>AALink [2]</b>
Case	Case
Т	!
<b>DomRel</b>	<b>DomRel</b>
<i>\$ситуация</i>	<i>\$ситуация</i>
<b>Master</b>	<b>Master</b>
->AVerb "стала" [46]	->AVerb "стала" [46]
<b>ORG</b>	<b>ORG</b>
-	-
<b>Position</b>	<b>Position</b>
<i>Pre</i>	<i>Post</i>
<b>Post</b>	<b>Post</b>
-	-
<b>Slave</b>	<b>Slave</b>
->AItem "Советником" [6]	->AItem "телеведущая" [9]
<b>Slave1</b>	<b>Slave1</b>
->AItem "Шойгу" [11]	->AItem "Мария Китаева" [10]
<b>Source</b>	<b>Source</b>
<i>Simple_link1</i>	<i>Simple_link1</i>
<b>TestPhase</b>	<b>TestPhase</b>
<i>situations_3</i>	<i>situations_3</i>
<b>TestRule</b>	<b>TestRule</b>
<i>noun_verb</i>	<i>verb_noun</i>
<b>Together</b>	<b>Together</b>
<i>yes</i>	<i>yes</i>
<b>TraceAddAnno</b>	<b>TraceAddAnno</b>
<i>situations_3::noun_verb</i>	<i>situations_3::verb_noun</i>

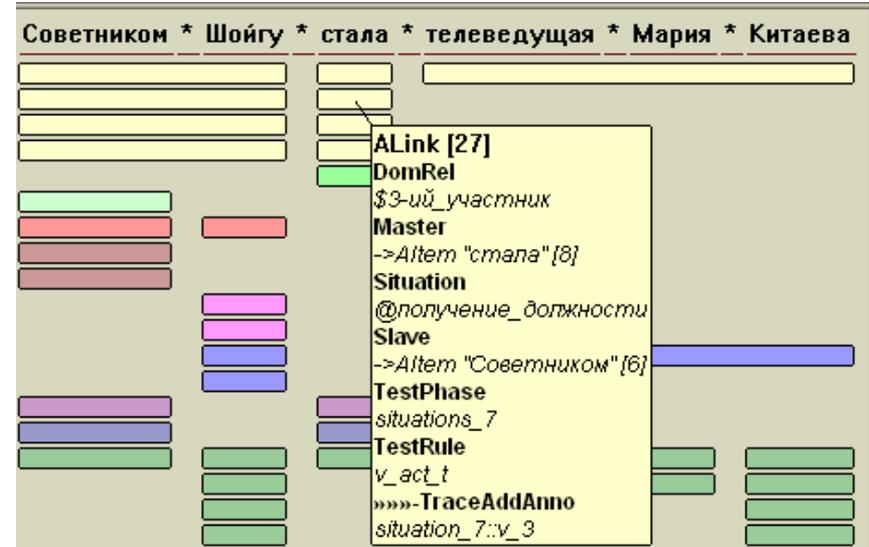
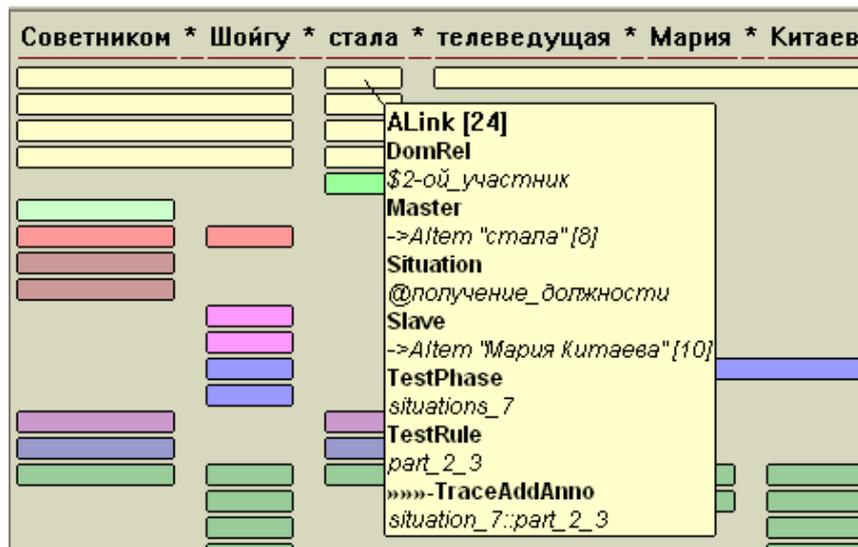
# Алгоритм извлечения ситуаций (3)

Советником \* Шойгу \* стала \* телеведущая \* Мария \* Китаева

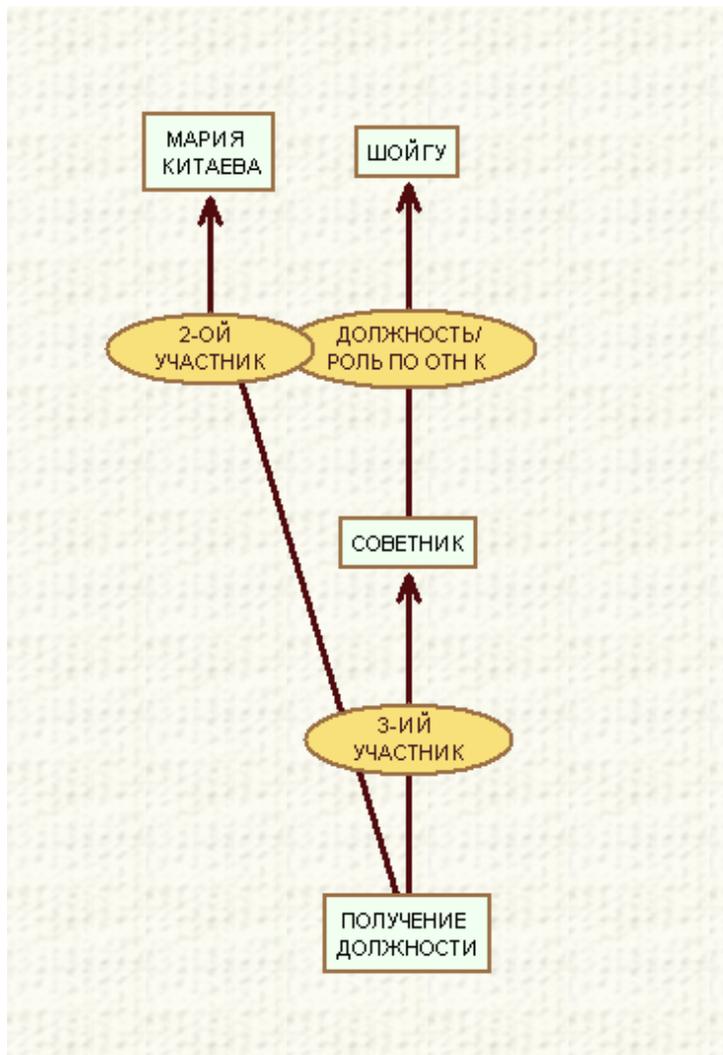
```
AVerb [46]
I_name_part
yes
Number_part
2
Part_I
->AItem "телеведущая" [9]
Part_I_name
->AItem "Мария Китаева" [10]
Part_T
->AItem "Советником" [6]
Part_T1
-
Ref
->AItem "стала" [8]
Sit
+
Situation
@получение_должности
T_part
yes
TestPhase
situations_2
TestRule
appoint(850)
Verb
->Morpho "стала" [66]
part_source
Simple_link1
»»»-TraceAddAnno
situations_2::appoint
```

В аннотацию слова-маркера записывается информация о потенциальных участниках.

# Алгоритм извлечения ситуаций (4)



# Алгоритм извлечения ситуаций (5)



Полученный результат извлечения.

# Результаты тестирования

Размечен 231 текст

(868 текстовых ситуаций назначения-отставки).

Точность 0,76

Полнота 0,72

F-мера 0,74

# Перспективы исследований

- Применение метода для других типов ситуаций – оценка быстроты перехода и качества обработки;
- Обогащение обрабатываемых контекстов (какие могут быть методики?)
- Классификация и разбор сложных случаев с омонимией;
- Разработка универсальных принципов разметки для ситуаций и создание размеченной коллекции;
- Классификация нетекстовых ситуаций и постепенное включение их в рассмотрение и обработку.