

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

Vlasova N.A., Lando N.S., Suleymanova E.A., Trofimov I.V.  
Program Systems Institute of RAS, Pereslavl-Zalessky, Russia

# Information extraction task

- **search** for relevant information in electronic textual documents
- **structuring** retrieved information for further analysis

# Subtasks ( $\pm$ )

- Tokenization
- Morphological analysis & disambiguation
- Named entity recognition
- Syntactic analysis
- Coreference resolution
- Detecting target information
- Structuring and normalizing information
- ...

# Situations-1000 purpose

- Evaluation of target event detecting algorithms:
  - independently of coreference subtask
  - independently of structuring and normalization subtask

# Target events

## appointments and dismissals

- What qualifies as a target event? (criteria)
  - reelection, downgrading to a position of, temporary duty assignment...
- Solution: list of triggers
  - подписать указ о назначении на должность (sign a decree on appointing ... as)
  - получить назначение на должность (get/obtain the position of)
  - прийти на должность (come to take the position of)
  - вернуться на должность (get back the position of)
  - поставить на должность (put ... to the position)
  - освободить кресло (vacate the position of)
  - оставить пост (leave the post of)
  - ...

# Event types

Event triggers were grouped into 5 types of target events – each with a unique set of syntactic roles.

“++” stands for the role that must be filled explicitly for a fragment to be marked up

“+” stands for an optionally filled role

	Appoint / Get a position	Dismiss / Leave a position	Replace / Take over a position	Appoint the head / Head	Management succession
Position	++	++	++		
Person-in	+		+	+	+
Person-out		+	+		+
Causator	+	+	+	+	+
Object of management				++	++

# Annotation

- ACE (Automatic Content Extraction) – a prototype
- Surface-based approach
  - event scope – the clause formed by the trigger root node
  - participants – noun phrases filling the argument positions of the target proposition + semantic constraints

# Surface-based approach: features(1)

Mark-up within the clause containing the trigger:

- *Иванов, который лишился своей должности...*
- *Иванов, **лишившийся** своей должности ....*



# Surface-based approach: features(2)

No coreference resolution for anaphoric expressions:

– *Министр обороны **лишился** своей должности.*

# Surface-based approach: issues and shortcomings

## Ellipsis

- *Министром связи назначен Руди Эриво, министром внутренних дел — Реджинальд Дельва [...]*
- *В 1991 г. назначен директором Кузнечного завода ОАО «Курганмашзавод».*
- *В октябре 2008 года полковник Йав получил повышение по службе и был назначен на должность начальника штаба второго военного округа.*
- *Исполняющим обязанности главы ГТРК «Крым» назначен первый замгендиректора Игорь Кульчик, Севастопольской РГТРК — замгендиректора по технике Александр Слободенюк.*

# Corpus profile

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

- Sources: news agency sites
- Size: 1000 documents
- Events annotated: 2249
- Tools: online browser

Thank you!