

Sentiment Analysis in Russian

Future evaluations

Natalia Loukachevitch

Outline

- Previous sentiment analysis evaluation in Russian and their results
 - General sentiment analysis
 - Users' reviews, news quotes
 - Entity and aspect-oriented sentiment analysis
 - Users' reviews, Twitter messages
- Future possible tasks and collections

General Sentiment Analysis 2011-2012

- Classification of users' reviews
 - In three domains (movies, books, digital cameras)
 - Three scales (two class, three class, five class)
- Results
 - All best results (in 9 tasks) were achieved with SVM classifier

General Sentiment Analysis 2012-2013

- New task: classification of news quotes
 - Three classes: positive, negative, neutral
 - *Example: Посредством этих структур десяткам тысяч избирателей предлагают деньги в обмен на паспортные данные и подписи за какого-либо кандидата*
 - Broad scope of topics
- Results
 - Best approaches were language-engineering, based on rules and lexicons
- Repetition of users' reviews classification task

Entity and Aspect-oriented Sentiment Analysis (2014-2015)

- Aspect-oriented sentiment analysis of Russian reviews
- Entity-oriented tweet analysis
 - Reputation monitoring
 - Banks and telecommunication companies
 - Results
 - Training and test collections were collected in different time intervals (divided by Ukraine events)
 - Approaches were fully based on the training collections
 - not enough for qualitative classification of the test collection because of data sparsity and time gap.

Entity and Aspect-oriented Sentiment Analysis (2015-2016)

- Entity-oriented tweet analysis
- The best results were obtained by the neural network
- Participants combined machine-learning and lexical resources
 - word clusters (word2vec)
 - manual lexicon RuSentiLex
<http://www.labinform.ru/pub/rusentilex/index.htm>
 - PMI-based lexicon based on tweets collection
study.mokoron.com
 - the crowdsourced lexicon Linis-crowd
<http://linis-crowd.org/>

Conclusion on Previous Evaluations


- «Simple» texts are under analysis
 - Reviews are usually devoted to a single entity
 - Tweets are short, small number of entities are mentioned
 - Not only in Russian Evaluation, also in English (SemEvaL)
- Other variants?
 - Analytical texts (for example, politics)

Analytical articles

- A lot of named entities (persons, states, organizations)
 - Positive or negative relations between named entities
 - Positive or negative events with the participation of entities
 - The author attitude towards some entities or absence of attitude
 - Combination of all issues in the same text fragment

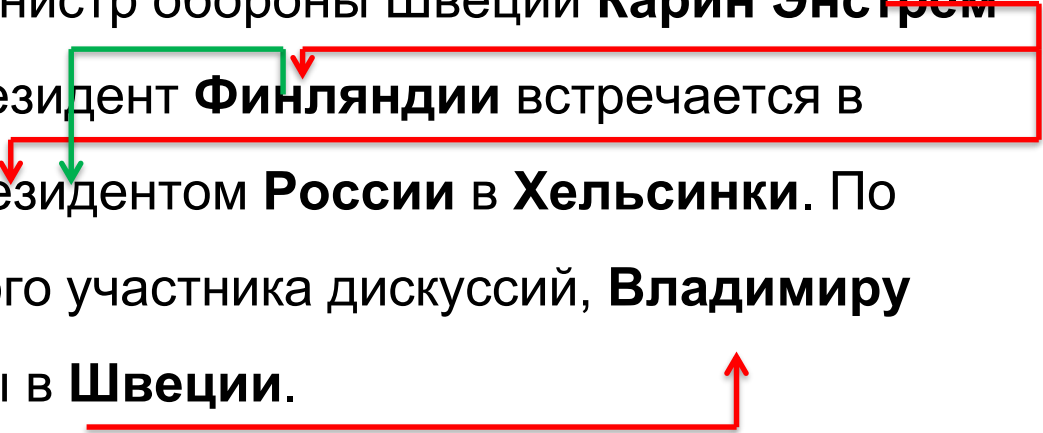
Example

- As is apparent in **Washington**, there is no place for objectivity on the subject of **Russia**, irrespective of facts and events.
- (<http://www.counterpunch.org/2017/05/26/ukraine-and-the-nato-military-alliance/>)
- Washington is **negative** to Russia
- Author is **negative** to Washington
- Author attitude to Russia?



Отношения **Финляндии** и **Швеции** можно считать хорошими. Ведь входили же страны в состав одного королевства до 1809 года. Страны объединяет также и то, что они не входят в **НАТО**, но являются партнерами альянса. Кроме того, **Финляндия** и **Швеция** укрепляют двустороннее сотрудничество в области обороны.

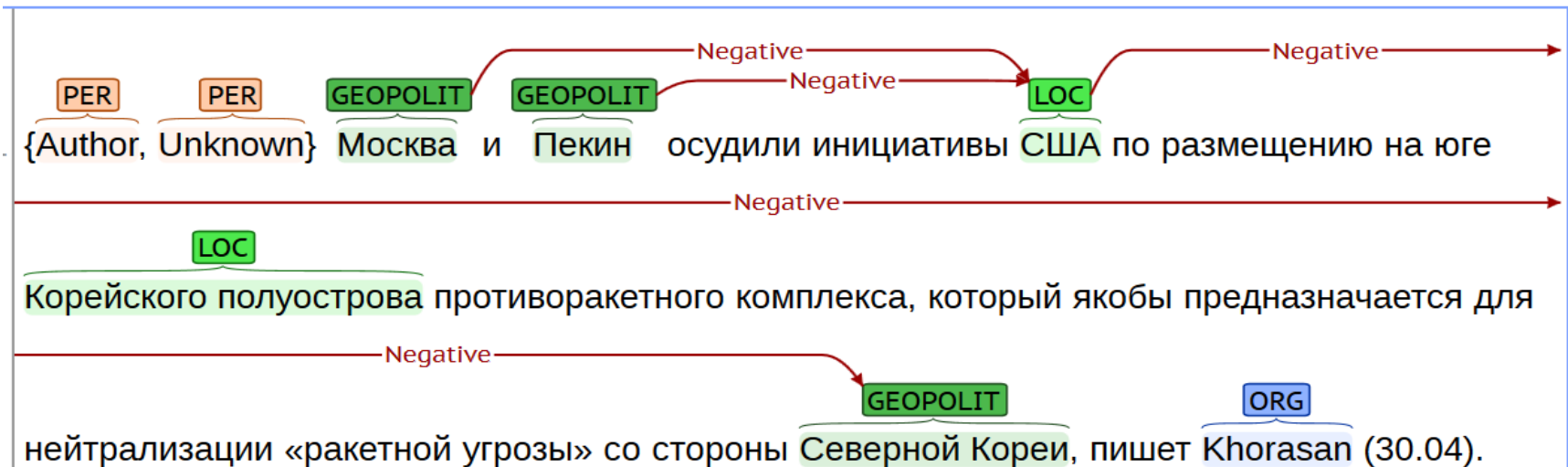
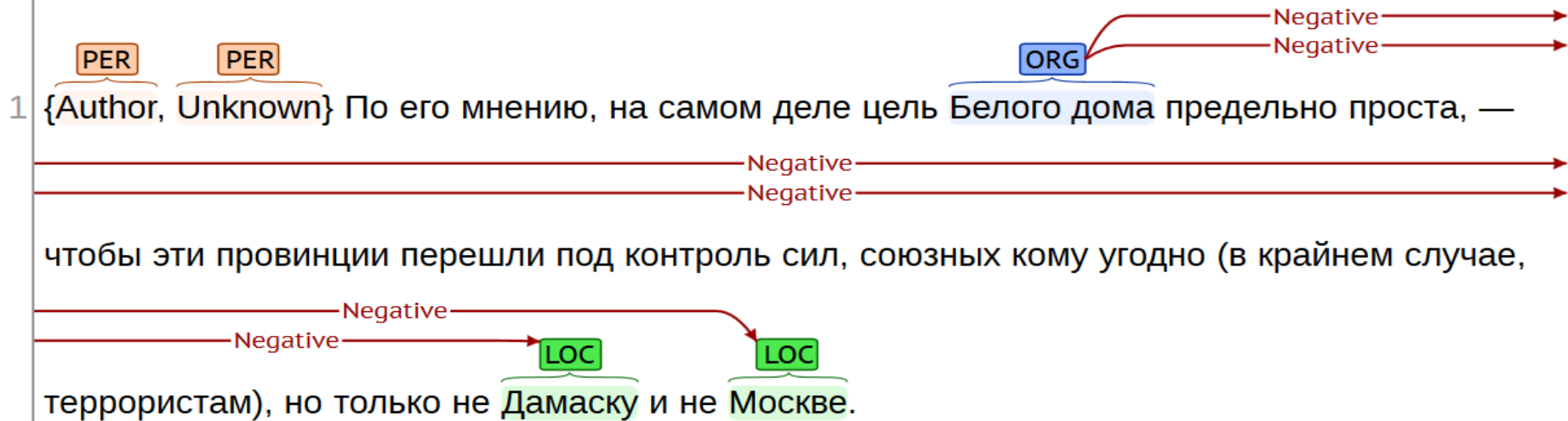
Несмотря на все это, в ходе обсуждений в **Култаранте** возникли разногласия. Бывший министр обороны Швеции **Карин Энстрём** возмутилась тем, что президент **Финляндии** встречается в следующем месяце с президентом **России** в **Хельсинки**. По мнению второго шведского участника дискуссий, **Владимиру Путину** не были бы рады в **Швеции**.



New collection

- Russian translations of political articles in portal inosmi.ru
- Articles contains a lot of opinions and attitudes between entities and author's attitude towards some entities
 - Labeling of single sentences
 - Labeling whole texts
 - Currently preliminary labeling of 71 whole texts
 - And their sentences

Sentence Labeling



Whole Text Labeling

Обама, Асад, neg, current (Obama, Asad, neg, current)

США, ИГИЛ, neg, current (USA, ISIL, neg, current)

Иран, Асад, pos, current (Iran, Asad, pos, current)

США, Ирак, neg, past (USA, IRAK, neg, past)

США, Афганистан, neg, current (USA, Afganistan, neg, current)

Япония, США, pos, current (Japan, USA, pos, current)

Южная Корея, США, pos, current (South Korea, USA, pos, current)

Австралия, США, pos, current (Australia, USA, pos, current)

неизв, Обама, pos, past (Unknown, Obama, pos, past)

неизв, Обама, neg, current (Unknown, Obama, neg, current)

автор, Обама, pos, current (Author, Obama, pos, current)

Possible tasks

- For mentioned named entities to determine polarity of relations and author attitude
 - In a single sentence
 - In the whole text
- Interesting issues
 - Interaction between different attitudes,
 - Relations vs. Events vs. Author sentiment
 - negation, direct speech, anaphora,
 - logical inference
 - Equivalence of entities relative to attitudes (Moscow-Russia-Putin)

