Multi-Sieve Approach to Character Identification and Classification

VIP Senior, 2017, Spring

Student: **Geeticka Chauhan**[1], Mentor: Mark Finlayson[1], Instructor: Masoud Sadjadi[1], Francisco Ortega[1]
[1]Florida International University

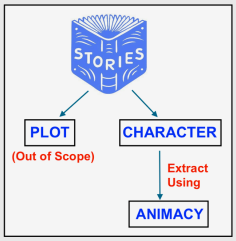
ABSTRACT

Our aim is to perform automatic classification of characters in folktales by using the concept of animacy. We define animacy as the property of being alive and communicative, for example: people and animals. In addition to these entities, folktales often involve characters that are not seen as animate in the real world, and this makes our problem a challenging one. Unlike other work, we attempt to classify the animacy of coreference chains by the use of a multi-sieve approach by combining a rule-based method and classifier. Our results are promising so far, and reaching very close in F1 measure of the state of the art techniques.

GOAL

Understand Stories and Culture

Tool: Natural Language Processing- NLP



IBM Watson

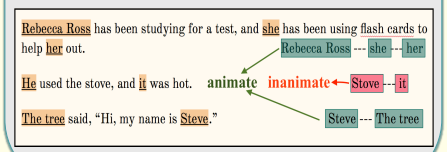
THIS IS FAKE

TERMINOLOGY

Animacy: Quality of being alive and communicative. Measure for Coreference Chains.

Referring Expression: Yellow Highlight

Coreference Chain: Chain of green boxes



ANNOTATION

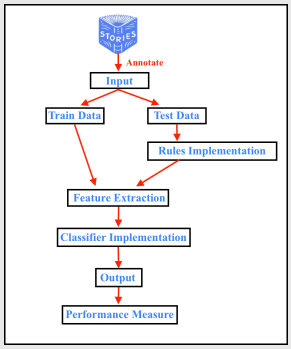
15 Russian Folktales with NLP Pipeline

+ Hand annotation of coreference chains by animacy

MODEL

Input: Annotated Russian Foktales

Output: Animacy of Coreference Chains



TECHNOLOGIES USED

- MIT Story Workbench
- Word2Vec Java Library
- MIT Java WordNet Interface
- Stanford CoreNLP

RULES

Design Rules for animacy of:

- **Part of Speech**: noun and pronoun
- **PERSON**: using Named Entity Recognition
- **Sense**: using WordNet

MACHINE LEARNING

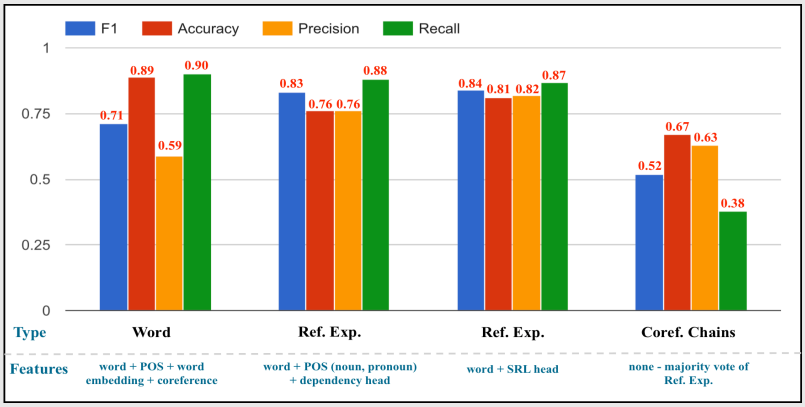
Support Vector Machine (SVM)

Split on Stories: 10 training, 5 test

Features:

- **Word**: 3 words before and after
- **Part of Speech**
- **Word Embedding**: Word2vec
- **Coreference**: belong to a chain?
- **Dependency Head**: Grammatical subject
- **SRL Head**: Semantic Subject

RESULT



CONTRIBUTIONS

- Clean Data
- Experimental Setup for animacy identification
- SVM Classifier
- Good Results of 84% F1

FUTURE WORK

Ongoing

Rules + Integrate more stories

Future

Role + Name Extraction

ACKNOWLEDGEMENT

I would like to thank my team mate, Labiba Jahan as well as my mentor, Dr. Finlayson for having me on board the Cognac Lab's project. I would also like to thank Dr. Ortega and Mohsen Taheri for their support.