
CLPL6: Semantic Roles and Conclusion

—

Tafseer Ahmed

—

Plan

- Word Embedding Code
- SemEval Tasks
- Semantic Roles
- Morphological Analyzer - Potential Project
- Part of Speech Tags - Urdu
- Conclusion

SemEval (SEMantic EVALuation)

- an ongoing series of evaluations of computational semantic analysis systems;
- organized by SIGLEX, a Special Interest Group on the Lexicon of ACL (the Association for Computational Linguistics)
- Datasets for SemEval tasks

SemEval 2020

Lexical semantics

- Task 1: Unsupervised Lexical Semantic Change Detection [[mailing list](#)] [[email organizers](#)]
- Task 2: Predicting Multilingual and Cross-Lingual (Graded) Lexical Entailment [[mailing list](#)] [[email organizers](#)]
- Task 3: Graded Word Similarity in Context (GWSC) [[discussion forum](#)] [[mailing list](#)] [[email organizers](#)]

Common Sense Knowledge and Reasoning, Knowledge Extraction

- Task 4: Commonsense Validation and Explanation [[mailing list](#)] [[email organizers](#)]
- Task 5: Modelling Causal Reasoning in Language: Detecting Counterfactuals [[mailing list](#)] [[email organizers](#)]
- Task 6: DeftEval: Extracting Definitions from Free Text in Textbooks [[mailing list](#)] [[email organizers](#)]

Humour, Emphasis, and Sentiment

- Task 7: Assessing Humor in Edited News Headlines [[mailing list](#)] [[email organizers](#)]
- Task 8: Memotion Analysis [[mailing list](#)] [[email organizers](#)]
- Task 9: Sentiment Analysis for Code-Mixed Social Media Text [[mailing list](#)] [[email organizers](#)]
- Task 10: Emphasis Selection for Written Text in Visual Media [[mailing list](#)] [[email organizers](#)]

Societal Applications of NLP

- Task 11: Detection of Propaganda Techniques in News Articles [[mailing list](#)] [[email organizers](#)]
- Task 12: OffensEval 2: Multilingual Offensive Language Identification in Social Media [[mailing list](#)] [[email organizers](#)]

SemEval2016

Track I. Textual Similarity and Question Answering Track

- [Task 1: Semantic Textual Similarity: A Unified Framework for Semantic Processing and Evaluation](#)
- [Task 2: Interpretable Semantic Textual Similarity](#)
- [Task 3: Community Question Answering](#)

Track II. Sentiment Analysis Track

- [Task 4: Sentiment Analysis in Twitter](#)
- [Task 5: Aspect-Based Sentiment Analysis](#)
- [Task 6: Detecting Stance in Tweets](#)
- [Task 7: Determining Sentiment Intensity of English and Arabic Phrases](#)

III. Semantic Parsing Track

- [Task 8: Meaning Representation Parsing](#)
- [Task 9: Chinese Semantic Dependency Parsing](#)

IV. Semantic Analysis Track

- [Task 10: Detecting Minimal Semantic Units and their Meanings](#)
- [Task 11: Complex Word Identification](#)
- [Task 12: Clinical TempEval](#)

V. Semantic Taxonomy Track

- [Task 13: TExEval-2 -- Taxonomy Extraction](#)
- [Task 14: Semantic Taxonomy Enrichment](#)

Semantic Roles

- The **glass** broke.
- **John** broke the **glass**.
- The **glass** was broken by **John**.

Semantic Roles

- The **glass**_[subject] broke.
- **John**_[subject] broke the **glass**_[object].
- The **glass**_[subject] was broken by **John**_[object-of-prep].

Semantic Roles

- The **glass**_[patient] broke.
- **John**_[agent] broke the **glass**_[patient].
- The **glass**_[patient] was broken by **John**_[agent].

List of Semantic Roles

- Agent
- Experiencer
- Causer
- Patient
- Theme
- Stimulus
- Instrument
- Causee

PropBank

Propbank Role		Example
Arg0	Agent, Experiencer	John broke the glass. John felt cold.
Arg1	Patient, Theme	John broke the glass. John bought the glass.
Arg2	Beneficiary, Recipient	I gave John a book. I bought a book for John.
...
Argm	Modifiers	John read the book in the evening.

Urdu PropBank

Predicate: *DagmagA*

DagmagA: created by Tafseer - Hindi dagmagA

Roleset id: **DagmagA.01** , *To stagger/tremble/shake*

Roles:

Arg1: *The entity that staggers/trembles/shakes*

Example:

<http://www.express.pk/story/44282/>

ہیلی کاپٹر بُری طرح ڈگمگایا

Arg1: ہیلی کاپٹر

Argm-mnr: بُری طرح

Rel: ڈگمگایا

Morphological Analyzer - a potential project

- As the first step, create morphological analyzer covering all the words and its inflectional form in the Swadesh List of the language.

https://en.wiktionary.org/wiki/Appendix:Indo-Aryan_Swadesh_lists

https://en.wiktionary.org/wiki/Appendix:Indo-Iranian_Swadesh_lists

- List all forms of the noun, and then find different set of morphological rules.

Morphological Analyzer - Noun

- Some possible reasons for different forms
 - Singular vs Plural vs ...
 - following/preceding post/prepositions
 - In the absence of case markers,
 - subject of perfect
 - Locatives
 - objects
 -

Part of Speech - Urdu

http://www.cle.org.pk/Downloads/ling_resources/wordlists/Closed%20Class%20Word%20List%20v1.1.pdf

Open class words	Closed class words	Other
<u>ADJ</u>	<u>ADP</u>	<u>PUNCT</u>
<u>ADV</u>	<u>AUX</u>	<u>SYM</u>
<u>INTJ</u>	<u>CCONJ</u>	<u>X</u>
<u>NOUN</u>	<u>DET</u>	
<u>PROPN</u>	<u>NUM</u>	
<u>VERB</u>	<u>PART</u>	
	<u>PRON</u>	

Revisiting Goals

Linguistic Analysis - Goals (of this series)

- Learning and creating **Linguistic Representations** that
 - are widely used in computer applications
 - deals with peculiar features of Pakistani languages
- Training the **annotators** for creating linguistically correct training dataset

Linguistic Analysis - Goals (of this series)

- Creating linguistic analyzers by using **small training datasets**
 - Unsupervised Learning
 - Word Embeddings
 - Transfer Learning
 -
- Using the **libraries and tools** of computational linguistics

Not the Goals/Focus of this series

- Text Mining
- Corpus Linguistics
- NLP algorithms - in depth
- Machine (including Deep) Learning - in depth
 - However, the topics related to linguistic analysis will be discussed

Thank You

شکریہ

Contact: tafseer@gmail.com