

## **Semantically Similar Sentences Identification and Generating Methods**

Petro Zdebskyi, Roman Kempnyk<sup>[0000-0002-8433-976X]</sup>, Vladyslav Chyhura<sup>[0000-0003-0588-1276]</sup>

Lviv Polytechnic National University, Lviv, Ukraine

The task can be formalized as a Recognising Textual Entailment (RTE) task. RTE is seen as recognizing two snippets of text, or the value of one can be deduced from the other. This task is not domain-bound and is proposed to recognize the variability of the semantic expression that is commonly required in many tasks.

The phenomenon of natural language is the variety of semantic expressions, in which the same meaning can be expressed or logically inferred from different texts. This phenomenon can be considered as a problem of language ambiguity, which are formed in communication many to many between language expressions and meanings. Text relationship inference between two texts: T (the text) and H (the hypothesis) represents a fundamental phenomenon of natural language. It is denoted as  $T \rightarrow H$  and means that the value of H can be logically inferred from T. This relationship is right because the value of one expression (e.g. "buy") can usually be deduced from the other (e.g. "own"), in turn, the logical conclusion in the other direction is less obvious. The recognition of textual inference is one of the hardest tasks natural language processing and progress in this task is the key to the solution of other tasks, such as finding answers to questions (Question Answering), information extraction (Information Extraction), information retrieval (Information Retrieval), summarization text (Text Summarization), and others.

The problem of generating semantically similar sentences can be reduced to the problem of generation of the text and validate or generate the text semantically similar to the sample.

The innovative component of this thesis is to study transitional in the RTE task, because the existing models do not explicitly implement this relationship. That is, if one sentence logically follows the second, and the third from the second, you should also run the logical conclusion of the third sentence of the first. That is, if  $A \rightarrow B$  and  $B \rightarrow C$  must be performed and  $A \rightarrow C$ . Also in the scope of tasks of the thesis includes the study of the limitations of existing data sets to solve the problem RTE and restrictions this task as such from the point of view of philosophy.

### **References**

1. Method and apparatus for performing full bidirectional translation between a source language and a linked alternative language. - Access mode: <https://patents.google.com/patent/US6275789B1/en>

2. Method and apparatus for language translation. - Access Mode: <https://patents.google.com/patent/CA2202696A1/en>
3. Shanidze, O., Petrasova, S.: Extraction of semantic relations from Wikipedia text corpus. In: Computational linguistics and intelligent systems, COLINS, 2, 74-75. (2019)
4. Chuiko, V., Khairova, N.: Semantic similarity identification for short text fragments. In: Computational linguistics and intelligent systems, COLINS, 2, 57-59. (2019)
5. Malyeyeva, O., Nosova, N., Fedorovych, O., Kosenko, V.: The Semantic Network Creation for an Innovative Project Scope as a Part of Project Knowledge Ontology. In: Computational linguistics and intelligent systems, COLINS, 301-311. (2019)
6. Shubin, I., Skovorodnikova, V., Kozyriev, A., Pitiukova, M.: Mining Methods for Adaptation Metrics in E-Learning. In: Computational linguistics and intelligent systems, COLINS, 288-300. (2019)
7. Kovaliuk, T., Kobets, N.: Semantic Analysis and Natural Language Text Search for Internet Portal. In: Computational linguistics and intelligent systems, COLINS, 277-287. (2019)
8. Kovaliuk, T., Kobets, N.: The Object Model of the Subject Domain with the Use of Semantic Networks. In: Computational linguistics and intelligent systems, COLINS, 228-242. (2019)
9. Bakumenko, N., Strilets, V., Ugryumov, M.: Application of the C-Means Fuzzy Clustering Method for the Patient's State Recognition Problems in the Medical Monitoring System. In: Computational linguistics and intelligent systems, COLINS, 218-227. (2019)
10. Dashkevich, A.: Semantic Segmentation of a Point Clouds of an Urban Scene. In: Computational linguistics and intelligent systems, COLINS, 208-217. (2019)
11. Basyuk, T., Vasyliuk, A., Lytvyn, V.: Mathematical Model of Semantic Search and Search Optimization. In: Computational linguistics and intelligent systems, COLINS, 96-105. (2019)
12. Kuprianov, Ye.: Semantic state superpositions and their treatment in virtual lexicographic laboratory for spanish language dictionary. In: 1st International Conference Computational Linguistics and Intelligent Systems, COLINS, P. 37-46. (2017)
13. Vysotska V. Linguistic analysis and modelling semantics of textual content for digest formation / Victoria Vysotska, Lyubomyr Chyrun // MEST Journal (Management Education Science & Society Technologie). – Vol.3 No.1. – PP. 127-148 [Online]. – ISSN 2334-7171, ISSN 2334-7058 (Online), DOI 10.12709/issn.2334-7058. This issue: DOI 10.12709/mest.02.02.02.0. – [http://mest.meste.org/MEST\\_1\\_2015/5\\_15.pdf](http://mest.meste.org/MEST_1_2015/5_15.pdf).
14. Vysotska V. Generative regular grammars application to modeling the semantics of sentences in natural language / Victoria Vysotska // Комп'ютерні системи проектування. Теорія і практика, Вісник Національного університету "Львівська політехніка" № 808,- Львів 2014 – Стор.43-56.
15. Берко А.Ю. Семантична інтеграція неповних та неточних даних / А.Ю. Берко, В.А. Висоцька, В.В. // Збірник наукових праць «Системи обробки інформації. Безпека та захист інформації в інформаційних системах», Випуск 7 (79). – Харків 2009. – Стор. 93-98.
16. Висоцька В.А. Особливості генерування семантики речення природною мовою за допомогою породжувальних необмежених та контекстно-залежних граматик / В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету "Львівська політехніка". – № 783. – Львів, 2014. – Стор. 271-292.

17. Висоцька В.А. Концептуальна модель процесу формування семантики речення природною мовою / В.А. Висоцька // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”, № 805.- Львів 2014 – Стор. 258-278.
18. Шестакевич Т.В. Моделювання семантики речення природною мовою за допомогою породжувальних граматик / Т.В. Шестакевич, В.А. Висоцька, Л.В. Чирун, Л.Б. Чирун // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”, № 814.- Львів 2015 – Стор. 335-352.
19. Литвин В.В. Метод визначення семантичної метрики на основі тезаурусу предметної області / В.В. Литвин, В.А. Висоцька, Р.М. Оливко // V Всеукраїнська науково-практична конференція «Інтелектуальні системи та прикладна лінгвістика». – Харків, 2016. – 14 квітня 2016 р. – НУ «ХПІ». – Стор. 10-12.
20. Висоцька В.А. Генерування речень українською за допомогою породжувальних граматик / В.А. Висоцька, Т.В. Шестакевич // Міжнародна наукова конференція “Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMIT’2012)”. – Євпаторія. – 27-31 травня 2012. – Стор.48-50.
21. Шестакевич Т.В. Застосування породжувальних граматик для генерування речень українською мовою / Т.В. Шестакевич, В.А. Висоцька // Східно-Європейський журнал передових технологій. – Харків, 2012. – № 3/2 (57). – С. 51-53.
22. Висоцька В.А. Застосування породжувальних граматик для моделювання синтаксису речення / В.А. Висоцька, Т.В. Шестакевич, НУ університету “Львівська політехніка”. – № 743. – Львів 2012. – Стор. 175-190.
23. Висоцька В. Порівняння складності автоматичного опрацювання англійських та українських текстів з врахуванням семантики та синтаксису природних мов / Висоцька В. А., Наум О. // Інформаційні системи та мережі. Вісник Національного університету “Львівська політехніка”, № 872.- Львів, 2017. – Стор. 149-162.
24. Kovalchuk, V., Lytvyn, V., Vysotska, V., Hrendus, M., Naum, O.: The Information System for Identification of Content Set Based on Analysis of Similar Texts. In: Computational Linguistics and Intelligent Systems, COLINS, 2, 122-127. (2018)
25. Lytvyn, V.: The similarity metric of scientific papers summaries on the basis of adaptive ontologies. In: Proceedings of 7th International Conference on Perspective Technologies and Methods in MEMS Design, MEMSTECH, 162. (2011)
26. Veres, O., Kis, Y., Kugivchak, V., Rishniak, I.: Reverse-search System of Similar or Identical Images. In: Computational Linguistics and Intelligent Systems, COLINS, 2, 129-138. (2018)
27. Lytvyn, V., Vysotska, V., Peleshchak, I., Basyuk, T., Kovalchuk, V., Kubinska, S., Chyrun, L., Rusyn, B., Pohreliuk, L., Salo, T.: Identifying Textual Content Based on Thematic Analysis of Similar Texts in Big Data. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 84-91. (2019)
28. Vysotska, V., Lytvyn, V., Kovalchuk, V., Kubinska, S., Dilai, M., Rusyn, B., Pohreliuk, L., Chyrun, L., Chyrun, S., Brodyak, O.: Method of Similar Textual Content Selection Based on Thematic Information Retrieval. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 1-6. (2019)
29. Vysotska, V., Burov, Y., Lytvyn, V., Demchuk, A.: Defining Author's Style for Plagiarism Detection in Academic Environment. In: International Conference on Data Stream Mining and Processing, DSMP, 128-133. (2018)
30. Shakhovska, N., Shvorob, I.:The method for detecting plagiarism in a collection of documents. In: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 142-145. (2015)