

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський національний технологічний університет
Університет Інформатики і прикладних знань, м.Лодзь, Польща
Національний технічний університет України «Київський
політехнічний інститут»
Навчально-науковий інститут комп'ютерних систем і технологій
«Індустрія 4.0» ім. П.М. Платонова

XXIII Всеукраїнська науково-технічна конференція
молодих вчених, аспірантів та студентів

«СТАН, ДОСЯГНЕННЯ ТА ПЕРСПЕКТИВИ **ІНФОРМАЦІЙНИХ СИСТЕМ І ТЕХНОЛОГІЙ»**

Матеріали конференції



Одеса

20-21 квітня 2023 р.

Стан, досягнення та перспективи інформаційних систем і технологій / Матеріали XXIII Всеукраїнської науково-технічної конференції молодих вчених, аспірантів та студентів. Одеса, 20-21 квітня 2023 р. - Одеса, Видавництво ОНТУ, 2023 р. – 449 с.

Збірник включає матеріали доповідей учасників конференції, які об'єднані за тематичними напрямками конференції.

Збірник буде корисним як для фахівців і працівників фірм, зайнятих в області ІТ, так і для викладачів, магістрів і студентів вищих навчальних закладів, які навчаються за напрямками і спеціальностями програмного забезпечення, обчислювальної техніки і автоматизованих систем, прикладної математики та обробки інформації, буде корисним професіоналам з комп'ютерного моделювання та розробки комп'ютерних ігор.

Результати досліджень у збірнику представляють собою своєрідний зріз сучасного стану справ в перерахованих галузях знань, який може допомогти як фахівцям, так і студентам університетів скласти загальну картину розвитку інформаційних технологій та пов'язаних з ними питань.

Наукові праці згруповані за напрямками роботи конференції та наведені в алфавітному порядку прізвищ авторів.

Матеріали (тези доповідей) друкуються в авторській редакції. Відповідальність за якість та зміст публікацій несе автор.

Матеріали подано українською та англійською мовами.

Редактор збірника Котлик С.В.

Яковенко М.І., Корнієнко Ю.К. (Одеський національний технологічний університет)	
Розділ 5: Комп'ютерні телекомунікаційні мережі та технології	300
1. Алгоритм попередньої обробки зображень для алгоритму QOI. Доценко Д., Крайник Я. М. (Чорноморський національний університет імені Петра Могили)	300
2. Аналіз сучасних архітектур GPU. Завальнюк Є.К., Романюк О.Н., Снігур А.В., Шевчук Р. П. (Вінницький національний технічний університет, Західноукраїнський національний університет)	302
3. Дослідження інструментальних засобів розробки програмного забезпечення для електронної комерції. Клівчук Д.К. (Волинський національний університет імені Лесі Українки)	304
4. Основні принципи роботи сучасних навігаційних систем. Наголюк Д. О. (Донецький національний університет імені Василя Стуса)	305
5. Сучасний стан і перспективи розвитку глобальних мереж інфокомунікацій. Нєнов О. Л. (Одеський національний технологічний університет)	307
6. Розробка захищеної корпоративної локальної мережі. Рижков М.С., Сахарова С.В., Нєнов О.Л. (Одеський національний технологічний університет)	309
7. Вимірювання параметрів оптичних компонентів мережі. Сахарова С.В., Рибалов Б.О. (Одеський національний технологічний університет)	311
8. Аналіз сучасних HTML-редакторів. Терешко Д. С., Романюк О. Н., Романюк О. В. (Вінницький національний технічний університет)	313
9. Оптимізація роботи алгоритму розподілу навантаження між серверами в мережі шляхом поєднання Rest і Soup. Тоха В.В. (Вінницький національний технічний університет)	314
10. Автоматизація процесу перебудови характеристик частотно-залежних компонент при обробці сигналів датчиків у робототехнічних системах. Чумаченко Н.К., Бадерко І.В., Ситніков В.С. (Національний університет "Одеська політехніка")	317
11. Розробка мережевого фільтра на базі міні комп'ютера Raspberry Pi. Шевчук М.С., Іванова Л.В., Сахарова С.В. (Одеський національний технологічний університет, Одеський технічний фаховий коледж ОНТУ)	319
Розділ 6: Штучний інтелект і автоматизація робототехнічних систем	322
1. Terms clustering hybrid service with word2vec, k-means, and majorclust algorithms for knowledge processing systems with cloud-based architecture. Malakhov K.S. (Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine)	322
2. Safety and ethics in the use of automated systems. Rysbek Akerke. (University "Turan", Kazakhstan)	324
3. Exploring extramae: a scalable self-supervised approach to synthetic time series generation. Аблець А. В. (Криворізький національний університет)	325
4. Синтетичні набори даних в штучному інтелекті. Антонова А.Р., Юрченко І.С. (Одеський національний технологічний університет)	326
5. Використання штучного інтелекту у 3D-моделюванні. Бойцова М.П., Бойцова О.С. (Одеський національний технологічний університет)	328
6. Розробка сайту психологічної допомоги на базі штучного інтелекту . Босенко Л.С., Болтач С.В. (Одеський національний технологічний університет)	330
7. Програма для відстеження пози та рухів людини на основі аналізу відео потоку з використанням MediaPipe. Вишневський В., Рябенський В., Вишневський В. (Національний Університет Кораблебудування ім. адмірала Макарова)	332
8. Використання штучного інтелекту в освіті: переваги, виклики та можливості. Горбачов О.С. (Донбаська державна машинобудівна академія)	334
9. Огляд метода знаходження оптимальної розкладки клавіатури за допомогою генеративного алгоритму штучного інтелекту (гаші). Горільський Е.О., Шаповалова Н. Н. (Криворізький національний університет)	335

Розділ 6.

Штучний інтелект та автоматизація робототехнічних систем

UDC 004.822

TERMS CLUSTERING HYBRID SERVICE WITH WORD2VEC, K-MEANS, AND MAJORCLUST ALGORITHMS FOR KNOWLEDGE PROCESSING SYSTEMS WITH CLOUD-BASED ARCHITECTURE

MALAKHOV K.S. (malakhovks@nas.gov.ua)

Microprocessor technology lab, Glushkov Institute of Cybernetics
of the National Academy of Sciences of Ukraine

The article discusses the algorithm and software implementation of the research module for distributional semantic analysis of natural language text documents in Ukrainian (extraction and clustering of existing concepts or other defined entities) in a system for transdisciplinary research support. The proposed approach to the natural language text processing uses an ordered set of universal procedures that allows to identify the basic units of language (words, phrases, or other defined entities), to classify them, to cluster them, and to define concepts in a text document.

Automated analysis of natural language texts continues to be an outstanding research challenge. By concepts we are referring to noun terms, as well as noun groups that have definitions in specialized dictionaries of domain knowledge or in the natural language text itself.

The architecture of the network tool component (in the form of a web service with an API) is presented as a pipeline for processing electronic text documents containing natural language text and consists of the following stages and components:

1. Extract text from .pdf, .docx, .doc documents.
2. Basic text normalization - the so-called linguistic preprocessing
3. Spell checking and automatic typos correction.
4. Lexical analysis – tokenization – it is the process of converting a sequence of characters into a sequence of tokens.
5. Grammatical analysis – marking up the text with parts of speech, in general, this is the process of assigning a syntactic category to each of the tokens.
6. Morphological analysis – this is the process of lemmatization. Syntactic analysis – parsing and chunking, in general, is the process of analyzing sentences, which first identifies their constituent parts and then associates them with higher-order units that have discrete grammatical meanings. The selected nouns and noun groups are checked against specialized dictionaries of the domain knowledge and against the availability of their definitions in the text sentences.
7. Formation of specialized XML and HTML text structures. To implement this stage, appropriate functions and procedures have been developed that allow obtaining the results of the electronic text document processing pipeline as output data.
8. Software and information utilities that define the functionality of the electronic text document processing pipeline ensure the solution of the task of automated processing of natural language text documents in Ukrainian based on linguistic and semantic analysis of the provided source data.

The next step is the choice of a clustering algorithm. While there are a few popular ones like K-means and DBSCAN, MajorClust is somewhat less known, but has an important advantage: it does not require stopping criteria like the number of clusters, the number of elements in a cluster, or the maximum distance between elements within a cluster, to be pre-set in advance: in many NLP

tasks such as word clustering these are very hard to guess, even though lots of experimentation. The algorithm is described in [1] in application to document clustering (see p.4 for pseudocode and explanation).

The algorithm begins by assigning every document to its own cluster. At each iteration, each document gets assigned to the cluster, to which it has the biggest similarity. A document's similarity to a cluster is calculated as the sum of the document's cosine similarities to each of the cluster's members. Thus, a document can be re-clustered to other clusters at later iterations. Clustering stops when no document changes its cluster.

However, I found that two modifications to this algorithm are necessary in my case. First, if many elements in the matrix have non-zero similarities to many other elements (e.g., word vectors have lots of non-zero values in the same columns, as is the case with word2vec vectors), then MajorClust may end up putting all elements into one single cluster. Therefore, it was necessary to introduce a threshold below which similarities get cancelled out, i.e., set to 0. After some experimentation I found that good results can be achieved by removing up to 99% of the smallest similarity pairs.

Second, after cancelling out low similarities, there will be many elements with no similarity to any other element. I.e., each of them should either constitute a cluster of its own or be treated as noise. MajorClust will merge two elements together even if the similarity between them is 0, thus producing one large catch-all cluster with lots of unrelated words. So, the other modification is to prevent assigning an element to a cluster if its similarity equals 0. My implementation requires that all similarity values are equal or greater than 0, so first the features in the vectors need to be scaled to a range, e.g. (0, 1).

The development and training of predictive models of distributional semantics for the Ukrainian language using the word2vec algorithm and others are described in [2]–[5]. We have developed an experimental implementation of the MajorClust algorithm with word2vec models for word clustering. The selected concepts form a network model for representing the knowledge of the relevant domain.

The article discusses the algorithm and software implementation of the research module for distributional semantic analysis of natural language text documents in Ukrainian as well as the extraction and clustering of existing concepts using the word2vec and MajorClust algorithms. Currently, the latest version of the service is available here and is free for use for research and educational purposes: <https://konspekt.ai-service.pp.ua/ua>. An experimental implementation of the MajorClust algorithm with word2vec models for word clustering is available here: <https://gist.github.com/malakhovks/a8b2e895a755419bc49b21e888267e2f>.

Important note: ken project and documentation are in active development. For any technical clarifications and questions contact us via email: malakhovks@nas.gov.ua or via Issues [6]. The recent Russian's rocket shelling on critical infrastructure in Ukraine and Kyiv led to the shutdown of the <https://konspekt.ai-service.pp.ua> server.

References

- [1] B. Stein and S. Meyer, "Document Categorization with MAJORCLUST," 2002. Accessed: Apr. 13, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/Document-Categorization-with-MAJORCLUST-Stein-Meyer/2380d838f03d0564631475904dc61e4c077a2997>
- [2] O. V. Palagin, V. Y. Velychko, K. S. Malakhov, and O. S. Shchurov, "Distributional semantic modeling: A revised technique to train term/word vector space models applying the ontology-related approach," in *CEUR Workshop Proceedings*, Kyiv, Ukraine: CEUR-WS, Sep. 2020, pp. 342–353. [Online]. Available: http://ceur-ws.org/Vol-2866/ceur_342-352palagin34.pdf
- [3] O. V. Palagin, V. Yu. Velychko, K. S. Malakhov, and O. S. Shchurov, "Research and development workstation environment: The new class of current research information systems," in *CEUR Workshop Proceedings*, Kyiv, Ukraine: CEUR-WS, May 2018, pp. 255–269. [Online]. Available: <http://ceur-ws.org/Vol-2139/255-269.pdf>

- [4] V. Velychko *et al.*, *New Information Technologies, Simulation and Automation*. Iowa State University Digital Press, 2022. doi: 10.31274/isudp.2022.121.
- [5] O. Palagin, M. Petrenko, and K. Malakhov, "Technique for designing a domain ontology," *Technique for designing a domain ontology*, no. 10, 2011, Accessed: Dec. 04, 2022. [Online]. Available: <http://dspace.nbuv.gov.ua/xmlui/handle/123456789/46447>
- [6] K. Malakhov, "KEen." Mar. 15, 2023. Accessed: Apr. 13, 2023. [Online]. Available: <https://github.com/malakhovks/ken>

SAFETY AND ETHICS IN THE USE OF AUTOMATED SYSTEMS

RYSBEK AKERKE (akerkerysbek98@gmail.com)

Turan University, Kazakhstan

The modern world is becoming more and more automated, and this leads to the emergence of new security and ethics issues. Automated systems such as robots, drones, automated vehicles and others are used in many fields, including manufacturing, transportation, medicine, military and others. However, their use may cause certain risks and raise ethical issues that need to be taken into account and addressed.

One of the main safety issues is the possibility of automated systems to cause harm to people and the environment. For example, robots operating in production environments can cause injury to workers if they are not provided with sufficient safety. Automatic vehicles can cause accidents if their software does not take into account all possible situations on the road. Drones can violate privacy if used unauthorized. In addition, the use of automated systems may raise ethical questions. For example, the use of automatic vehicles can lead to the loss of drivers' jobs, which can affect the economy and society as a whole. Robots replacing people in some jobs can cause dissatisfaction among people who lose their jobs. Military robots may raise questions about what level of control and responsibility the people controlling these systems should have. Addressing these issues requires joint efforts from manufacturers, government agencies, the public and other stakeholders. Manufacturers must guarantee the safety and reliability of their systems.

Nowadays, when we are more and more dependent on technology, the issues of safety and ethics in the use of automated systems are becoming more and more relevant. First of all, we would like to address the issue of security. Automated systems can be indispensable in operation, but they can also pose a security threat. For example, hackers can use weaknesses in the system to break into databases and obtain confidential information. Therefore, it is necessary to take all necessary measures to ensure the safety of automated systems. This can be regular software updates, the installation of strong passwords, as well as the use of a two-factor authentication system. In addition, it is necessary to take into account ethical aspects when using automated systems. For example, we need to make sure that the systems do not violate the rights and freedoms of people. This may be a violation of privacy rights if the system collects and stores personal data without the consent of users. It is also necessary to take into account the possibility of using our systems to manipulate public opinion. For example, it may be the use of algorithms to spread fake news or manipulate election results.

Another important aspect to consider when using automated systems is transparency. Companies should be ready to communicate openly and transparently with users and the public about how automated systems are used and what data they collect and process. This will help establish trust on the part of users and protect companies from possible accusations of violating privacy rights or ethical principles. In addition, it is necessary to take into account that automation can lead to the dismissal of people and job cuts. Therefore, companies should be prepared for social responsibility and the provision of alternative options for employees who may be affected by the