# Structural and typological analyses of special lexical units in the sphere of information technologies in English

Tomsk Polytechnic University

Natalia Zyablova [c], Radik Ismagilov [a], Sandaara Petrova [b]

[a] School of Nuclear Science & Technology, Tomsk Polytechnic University
[b] School of Nuclear Science & Technology, Tomsk Polytechnic University
[c] School of Core Engineering Education, Tomsk Polytechnic University

**Abstract**

This article represents structural and typological analyses of special lexical units in the sphere of information technologies in English. 68 lexical units were classified into three groups due to typological analysis: abbreviations, monosyllabic terms, terminological units. According to the structural analysis, they were divided into such groups as multicomponent combinations, one-component models, two-component models. In addition, lexical units were categorized by subject areas: general, software engineering, graph theory, algorithm theory, mathematical statistics, data science, virtual reality, IT management and computer networks. Judging by results of typological analysis, terminological units are the most frequent type of lexical units. The most popular structure is two-component model. Finally, Computer Network and Data Science turned out to include more terminological units than any other subject area of information technology.

*Keywords:* Information technology, structural analysis, typological analysis;

## 1. Introduction

It is a well-known fact that information technology (IT) plays an important role in modern society. Digitalization tendency is traced in a lot of areas. According to Rosstat, in 2016 92.4% of organisations used personal computers and 89.6% could not handle without global computer networks. [9] IT sphere is constantly developing and updating with new technologies and gadgets, which are designated with new special lexical units (neologisms). Specialists all over the world exchange topical and accumulated knowledge with the help of scientific publications but it is impossible without English terminology because English is an international communicative means in a modern world.

Terminology includes terms and terminological units – special lexical units which are used by specialists of a certain sphere for designation of special objects, technics and technologies. Special vocabulary is predominantly used in publications of scientific register. Scientific register is one of functional registers of Standard English used in the sphere of science and manufacture. [10]

Structural analysis (analysis of part-of-speech compatibility) and typological analysis of special vocabulary (terms and terminological units) in sphere of information technology in modern English are represented in this article. Special lexical units of this sphere are correlated with particular theme groups which are found in scientific articles of IT sphere (Software Engineering, Graph Theory, Data Science, General, Theory of Algorithms, Virtual Reality, Computer Networks, IT-

management, High Technology and Mathematical Statistics). Special lexical units of IT sphere have also been translated into Russian and can be used as glossary.

The relevance of research is due to the need of revealing ways of designation of special objects, technics and technologies belonging to the mentioned subject area.

## 2. Methodology

68 special lexical units of IT sphere have been chosen for structural and typological analysis, including terms, terminological units and abbreviations with the help of sampling methods from scientific articles of IT sphere. Below is the list of terms, terminological units and abbreviations chosen from scientific articles [1-8, 12-14]. The terms, terminological units and abbreviations have also been translated. The abbreviations have been decrypted.

Agile (Agile Software Development) (Adj+N+N) – Гибкая методология разработки [Программная инженерия=ПИ]

Agile attributes (Adj+N) – Agile атрибуты [ПИ]

Algorithmic application (Adj+N) – Алгоритмическое приложение [ПИ]

Alpha-beta pruning (N-N+N) – Альфа-бета отсечение [Теория графов=ТГ]

Artificial intelligence (Adj+N) – Искусственный интеллект [Наука о данных=НД]

Baseline model (N+N) – Базовая модель [НД]

Big Data Analytics (аббревиатура BDA) (Adj+N+N) – Анализ больших данных [НД]

Big Data Technologies (аббревиатура BDT) (Adj+N+N) – Технологии больших данных [НД]

Breadth-first search (аббревиатура BFS) (N-Numeral+N) – Поиск в ширину [ТГ]

Brute force (Adj+N) – Полный перебор [ТГ]

Computer program (N+N) – Компьютерная программа [ПИ]

Computer Science (аббревиатура CS) (N+N) – Информатика [Информационные технологии=ИТ]

Constant False Alarm Rate (аббревиатура CFAR) (Adj+Adj+N+N) – Постоянная вероятность ложных тревог [Теория алгоритмов=ТА]

Content (N) – Содержимое [ИТ]

Data algorithm (N+N) – Алгоритм данных [НД]

Data mining (N+N) – Сбор данных [НД]

Data source (N+N) – Источник данных [НД]

Database marketing (N+N) – Маркетинг на основе баз данных [НД]

Dataset (N) – Набор данных [НД]

*Depth-first search (аббревиатура DFS) (N-Numeral+N) – Поиск в глубину* [ТГ]

Digital augmentations (Adj + N) – Цифровые аугментации (улучшение виртуального присутствия за счет использования цифрового контента) [Виртуальная реальность=ВР]

Digital society (Adj+N) – Цифровое общество [ИТ]

Domain (N) - Доменное имя или домен (имя сайта) [Вычислительные сети=ВС]

E-banking (универбат) (Abbreviation-PI) – Интернет-банкинг [ВС]

E-mail (универбат) (Abbreviation-N) – Электронная почта [ВС]

Framework (с англ. - структура) (N) – Фреймворк [ВС]

Gamification (N) – Игрофикация [ИТ-менеджмент=ИТ-М]

Heuristic search (Adj+N) – Эвристический поиск [ТА]

Hidden neuron (Adj+N) – Скрытый нейрон [НД]

House of Quality (аббревиатура HOQ) (N+prep+N) – Дом качества [Управление=У]

Information technology (аббревиатура IT) (N+N) – Информационные технологии [ИТ]

Intelligent computer information processing (Adj+Adj+N+N) – Интеллектуальная компьютерная обработка информации [НД]

Internet (N) – Интернет [ВС]

Internet user (N+N) – Пользователь Интернета [ВС]

Internet website (N+N) – Интернет вебсайт [ВС]

Likelihood Ratio (аббревиатура LR) (N+N) – Отношение правдоподобия [Математическая статистика=МС]

Linear programming (Adj+N) – Линейное программирование [ПИ]

Local Internet Content (аббревиатура LIC) (Adj + N + N) – Локальный интернет-контент [ВС]

Machine learning (N+N) – Машинное обучение [НД]

Measurement model (N+N) – Модель измерения [НД]

Monte Carlo method (N+N) – Метод Монте-Карло [МС]

*Multilayer Perceptron* (аббревиатура MLP) (N+N) – Многослойный персептрон [ТГ]

Network (N) – Сообщество, сеть [ИТ]

Network Features (аббревиатура NF) (N+N) – Сетевые функции [ВС]

Neural network (N+N) – Нейронная сеть [НД]

Numerical analysis (Adj+N) – Численный анализ [ТА]

Offline (Adj) – Не в сети [ВС]

Online (Adj) – В сети [ВС]

Online Social Networks (аббревиатура OSNs) (Adj+Adj+N) – Онлайн социальные сети [ВС]

Personal computer (аббревиатура PC) (Adj+N) – Персональный компьютер [ИТ]

Quality Function Deployment (аббревиатура QFD) (N+N+N) – Структурирование (развертывание) функции качества [ИТ-М]

Radial Basis Function Neural Networks (аббревиатура RBFNN) ((Adj+N+N+Adj+N) – Нейронные сети радиально базисных функций [НД]

Robotics (N) – Робототехника [Высокие технологии]

Scrum (N) – Метод управления проектами [ИТ-М]

Self-Optimizing *Neural Networks (аббревиатура SONN)* (Adj-PI+Adj+N) – Самооптимизирующиеся нейронные сети [НД]

Signal-to-Interference Ratio (аббревиатура SIR) (N-prep-N+N) – Отношение сигнал-помеха [ТА]

Smart home technology (Adj+N+N) – Технологии умного дома [ВТ]

Smart-sensing Technology (аббревиатура SST) (Adj-PI+N) – Умная сенсорная технология [ВТ]

Social network (Adj+N) – Социальная сеть [ВС]

Software development (N+N) – Разработка программного обеспечения [ПИ]

Structural model (Adj+N) - Структурная модель [НД]

Support Vector Machine (аббревиатура SVM) (N+N+N) – Метод опорных векторов [НД]

Technology acceptance (N+N) – Принятие технологий [ВС]

Tree (N) – Дерево [ТГ]

Uniform Resource Locator (аббревиатура URL) (Adj+N+N) – Унифицированный указатель ресурса [ВС]

Web design (N+N) – Веб-дизайн [ВС]

Web Domain Dataset (аббревиатура WDD) (N+N+N) – Набор данных веб-домена [ВС]
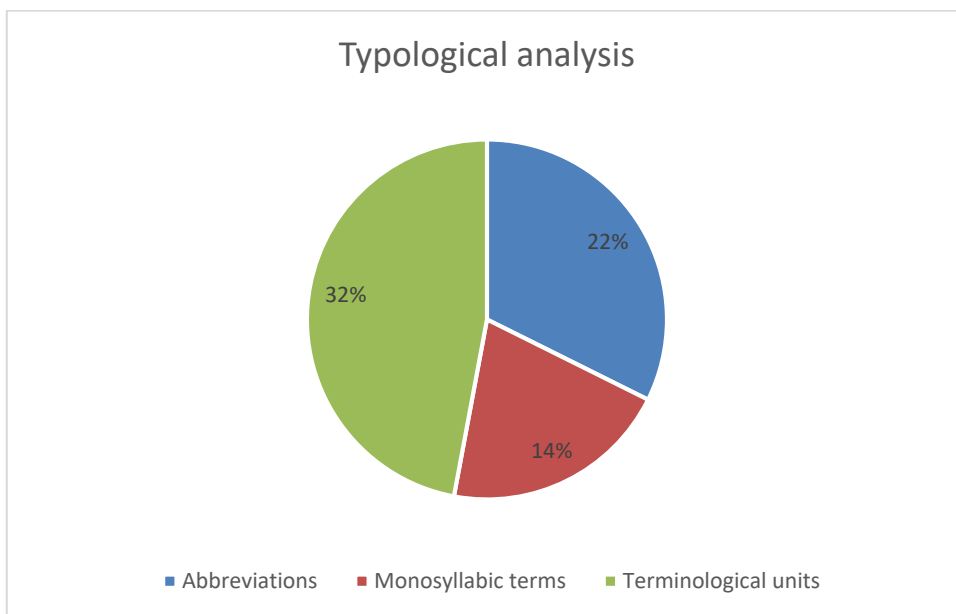
Webpage (N) – Веб-страница [ВС]

Fig. 1 Typological analysis of English terms, terminological units and abbreviations in the field of IT

In the process of typological analysis, it has been revealed that most part of the English terminology in IT (32%) are terminological units. The second largest group - abbreviations, occupying 22% of the total sample size, this indicates a tendency to reduce the most frequently-used special vocabulary. The lowest share (14%) are monosyllabic terms. (Fig. 1)
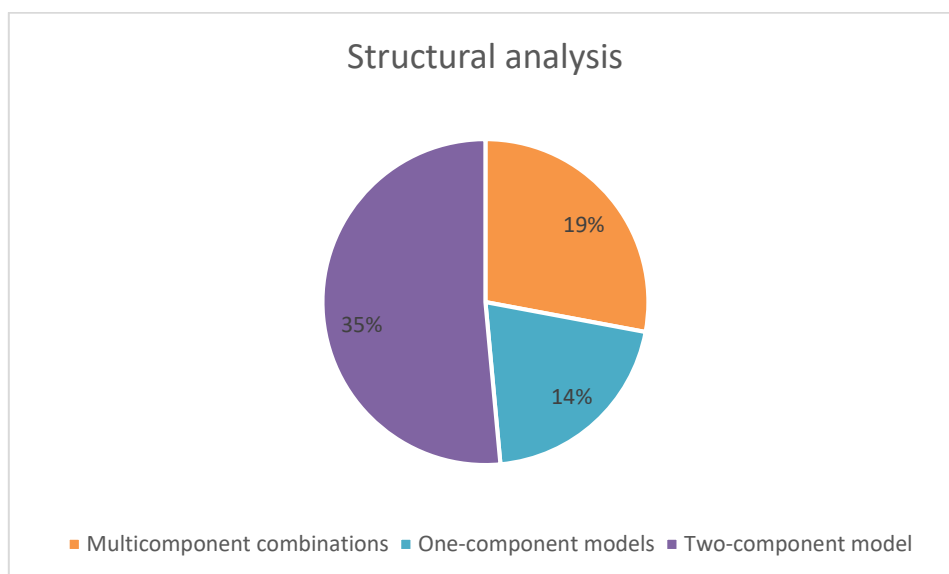


Fig. 2 The pie chart of structural analysis of special lexical units

In the course of the study, it has been found that the two-component model is the most common, including combinations of N + N, Adj + N and universal banking (E-banking) [11] with the noun

(N). 19% is multicomponent combinations (combinations with a noun, for example, Agile attributes (Adj + N)), they also include abbreviations consisting of more than two abbreviated words (RBFNN). The smallest group consisted of one-component models - 14%. (Fig. 2)

Thus, the results of typological and structural analyzes allow us to conclude that the two-component model is the predominant model for designation of special lexical units.

The field of information technology is represented by the following areas: general, software engineering, graph theory, algorithm theory, mathematical statistics, data science, virtual reality, IT management and computer networks.
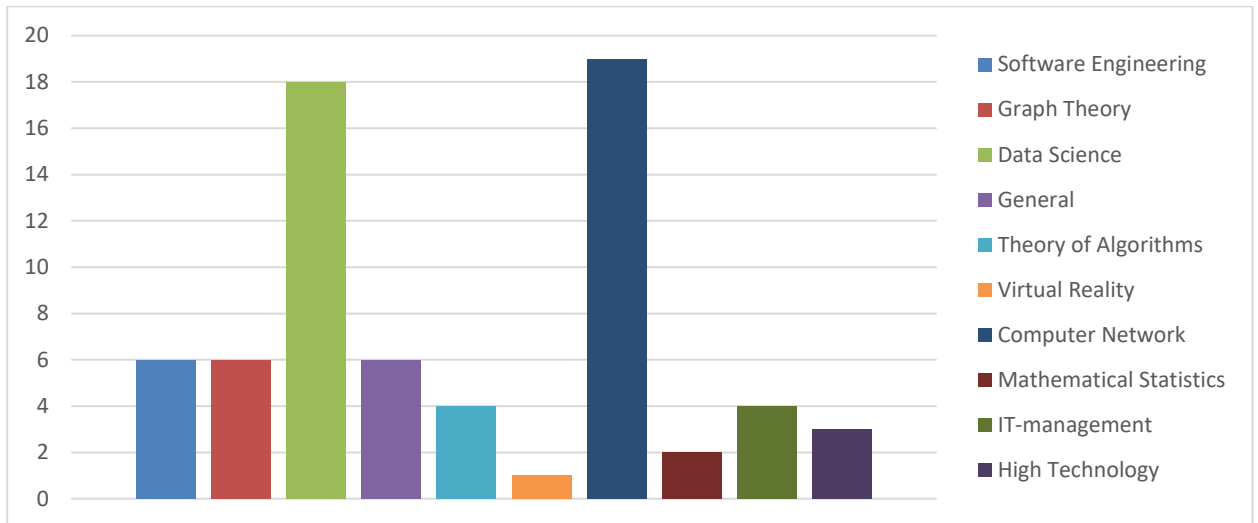


Fig. 3 The histogram of analysis of the ownership of special units in the subject areas of the IT sphere

During the study the largest number of special lexical units (19 out of 68) related to the field of computer networks including neologisms has been discovered. (Fig. 3). 18 special lexical units from the sample relate to data science, the most developing branch of the IT sphere. The remaining selected subject areas include a small number of terminological units relative to the total sample.

## 3. Conclusion

Information technology (IT) is one of the most developing and requested spheres in our modern world which causes appearance of new special words (terms and terminological units) in different subject areas of IT. Abundance of neologisms including abbreviations which became popular due to the tendency to cut words down makes it difficult to comprehend and translate foreign scientific articles in the sphere of IT. This causes necessity to find out different ways of coining terms and terminological units.

According to the results of typological analysis the dominant group of special lexical units in the sphere of IT is terminological units. Structural analysis showed the prevalence of the composition of two components in terminological units. However, the quantity of two-component terminological units slightly exceeds the number of one-component and multi-component terms and terminological units accordingly. The number of subject sub-areas of the sphere of IT revealed in the research is 10.

The results of the typological and structural analyses presented in the article can be useful for further research of special lexical units in the area of IT.

**References**

1. Artificial intelligence techniques for small boats detection in radar clutter. Real data validation analytics. *Engineering Applications of Artificial Intelligence*, January 2018, pp. 296-308. [Available at https://www.sciencedirect.com/science/article/pii/S0952197617302610] [Accessed on 12.04.2020]

2. Benjamin, K. Sovacool. (2020). Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies. *Renewable and Sustainable Energy Reviews*. [Available at https://www.sciencedirect.com/science/article/abs/pii/S1364032119308688] [Accessed on 12.04.2020]

3. Emmanouil Tranos. (2020). Individual internet usage and the availability of online content of local interest: A multilevel approach. *Computers, Environment and Urban Systems*. [Available at https://www.sciencedirect.com/science/article/pii/S0198971519300808] [Accessed on 12.04.2020]

4. Heli Hallikainen. (2020). Fostering B2B sales with customer big data analytics. *Industrial Marketing Management*. pp. 90-98. [Available at https://www.sciencedirect.com/science/article/pii/S0019850118304656] [Accessed on12.04.2020]

5. Herbert A. Simon. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence*. pp. 95-127. [Available at https://www.sciencedirect.com/science/article/pii/000437029500039H] [Accessed on 12.04.2020]

6. Luís Filipe Rodrigues. (2017). How does the web game design influence the behavior of e-banking users? *Computers in Human Behavior*. pp. 163-174. [Available at https://www.sciencedirect.com/science/article/pii/S0747563217302765] [Accessed on 12.04.2020]

7. Methodology for the Design of Agile Product Development Networks. *Procedia CIRP* [Available at https://www.sciencedirect.com/science/article/pii/S2212827119307942] [Accessed on 12.04.2020]

8. Po Yang. (2018). The Internet of Things (IoT): Informatics methods for IoT-enabled health care things. *Journal of Biomedical Informatics*. pp 154-156. [Available at https://www.sciencedirect.com/science/article/pii/S1532046418302028] [Accessed on 12.04.2020]

9. Rossiya v cifrah. 2018: Krat.stat.sb./Rosstat. M. 522 s.

10. Russkij yazyk i kul'tura rechi. Tol'yatti: TGU, 2005. 99 s.

11. Tatarinov, V.A. (1996). Teoriya terminovedeniya. M.: Mosk. Licej. T.1: Teoriya termina: Istoriya i sovremennoe sostoyanie. 311 s.

12. Xiaoli Xu. (2012). Intelligent fault prediction system based on internet of things. *Computers & Mathematics with Applications*. pp. 833-839. [Available at https://www.sciencedirect.com/science/article/pii/S0898122111011059] [Accessed on 12.04.2020]

13. Yong Jin Park. (2020). Contextualizing privacy on health-related use of information technology. *Computers in Human Behavior* [Available at https://www.sciencedirect.com/science/article/pii/S0747563219304248#ec-research-data] [Accessed on 12.04.2020]

14. Zyablova, N.N. (2016). Leksiko-semanticheskoe pole «Vozobnovlyaemye istochniki energii»: leksikologicheskij i normativnyj status rekurrentnyh edinic v sovremennom anglijskom yazyke: avtoref. dis. … kand. fil. nauk: 10.02.04. Tomskij politekhnicheskij universitet.