Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.41, 004.89
https://doi.org/10.32362/2500-316X-2023-11-5-7-18

RESEARCH ARTICLE

Automatic depersonalization of confidential information

Nikita G. Babak [@, 1, 2], Leonid Yu. Belorybkin [2], Shamil A. Otsokov [3], Alexey A. Terenin [2], Anastasia I. Shabrova [2]

[1] *National Research University "Moscow Power Engineering Institute," Moscow, 111250 Russia*
[2] *Sberbank of Russia, Moscow, 117312 Russia*
[3] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[@] *Corresponding author, e-mail: nikita.enrollee@gmail.com*

Abstract

Objectives. As the scope of personal data transmitted online continues to grow, national legislatures are increasingly regulating the storage and processing of digital information. This paper raises the problem of protecting personal data and other confidential information such as bank secrecy or medical confidentiality of individuals. One approach to the protection of confidential data is to depersonalize it, i.e., to transform it so that it becomes impossible to identify the specific subject to whom the data belongs. The aim of the work is to develop a method for the rapid and safe automation of the depersonalization process using machine learning technologies.

Methods. The authors propose the use of artificial intelligence models to implement a system for the automatic depersonalization of personal data without the use of human labor to preclude the possibility of recognizing confidential information even in unstructured data with sufficient accuracy. Rule-based algorithms for improving the precision of the depersonalization system are described.

Results. In order to solve this problem, a model of named entity recognition is trained on confidential data provided by the authors. In conjunction with rule-based algorithms, an F_1 score greater than 0.9 is achieved. For solving specific depersonalization problems, a choice between several implemented anonymization algorithm variants can be made.

Conclusions. The developed system solves the problem of automatic anonymization of confidential data. This opens an opportunity to ensure the secure processing and transmission of confidential information in many areas, such as banking, government administration, and advertising campaigns. The automation of the depersonalization process makes it possible to transfer confidential information in cases where it is necessary, but not currently possible due to legal restrictions. The distinctive feature of the developed solution is that both structured data and unstructured data are depersonalized, including the preservation of context.

Keywords: automated system, anonymization, information protection, cybersecurity, sensitive information, machine learning, neural networks, depersonalization, personal data, named entity recognition

• **Submitted:** 10.02.2023 • **Revised:** 14.06.2023 • **Accepted:** 13.07.2023

For citation: Babak N.G., Belorybkin L.Yu., Otsokov Sh.A., Terenin A.A., Shabrova A.I. Automatic depersonalization of confidential information. *Russ. Technol. J*. 2023;11(5):7–18. https://doi.org/10.32362/2500-316X-2023-11-5-7-18

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Автоматическое обезличивание конфиденциальной информации

Н.Г. Бабак [@, 1, 2], Л.Ю. Белорыбкин [2], Ш.А. Оцоков [3], А.А. Теренин [2], А.И. Шаброва [2]

[1] *Национальный исследовательский университет «МЭИ», Москва, 111250 Россия*
[2] *Публичное акционерное общество «Сбербанк России», Москва, 117312 Россия*
[3] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[@] *Автор для переписки, e-mail: nikita.enrollee@gmail.com*

Резюме

Цели. В то время как объем персональных данных, передаваемых по сети, продолжает расти, законодательные органы все более жестко регулируют процессы хранения и обработки цифровой информации. В работе рассматривается проблема защиты персональных данных и другой конфиденциальной информации (КИ), например, банковской или врачебной тайны, физических лиц. Одним из способов защиты конфиденциальных данных является их обезличивание – преобразование, в результате которого становится невозможно установить принадлежность этих данных конкретному субъекту. Цель работы – построение автоматической системы, позволяющей быстро и безопасно обезличивать данные с помощью технологий машинного обучения.

Методы. Предлагается использовать модели искусственного интеллекта для реализации системы автоматического обезличивания КИ, т.к. это дает возможность распознавать КИ даже в неструктурированных данных с достаточно высокой точностью без привлечения человеческого труда. Для повышения точности всей системы обезличивания предлагается использовать алгоритмы на основе правил.

Результаты. На конфиденциальных данных, размеченных авторами для решения данной задачи, обучена модель распознавания именованных сущностей, которая в связке с алгоритмами на основе правил в результате имеет значение F_1-меры больше, чем 0.9. Реализовано несколько вариаций алгоритмов обезличивания, что позволяет выбирать между ними для каждой конкретной задачи.

Выводы. Разработанная система решает задачу автоматического обезличивания КИ. Это открывает возможность для безопасной обработки и передачи КИ во многих областях, например, в банковской деятельности, государственном управлении, рекламных кампаниях. Также автоматизация процесса обезличивания делает возможной передачу КИ в тех случаях, когда это необходимо, но невозможно в силу правовых ограничений. Отличительная особенность разработанного решения заключается в том, что обезличиваются как структурированные данные, так и неструктурированные, в т.ч. с сохранением контекста.

Ключевые слова: автоматизированная система, анонимизация, защита информации, кибербезопасность, конфиденциальная информация, машинное обучение, нейросети, обезличивание, персональные данные, распознавание именованных сущностей

• **Поступила:** 10.02.2023 • **Доработана:** 14.06.2023 • **Принята к опубликованию:** 13.07.2023

Для цитирования: Бабак Н.Г., Белорыбкин Л.Ю., Оцоков Ш.А., Теренин А.А., Шаброва А.И. Автоматическое обезличивание конфиденциальной информации. *Russ. Technol. J.* 2023;11(5):7–18. https://doi.org/10.32362/2500-316X-2023-11-5-7-18

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

In today's digitally intermediated world, the scope of stored and processed data is constantly growing, requiring increased reliability in terms of data protection. The issue of protecting personal data transmitted via computer networks and stored in information systems becomes particularly relevant. The list and procedure for processing personal data is outlined in Federal Law No. 152 "On Personal Data." Here, personal data is defined as any information pertaining to a directly or indirectly identified or identifiable individual[1].

Modern computer systems allow organizations to collect and process large amounts of data necessary for their effective functioning and development. The access to various kinds of data facilitated by the rapid development of information technology in turn increases the risk of information leakage [1]. The high risk of illegal access to confidential information (CI) makes the task of ensuring its protection particularly relevant.

One of the measures aimed at minimizing the risks of harm to an individual in the event of leakage of personal data from automated systems (AS) is depersonalization as required by law. The depersonalization of personal data comprises an action that makes it impossible to determine, without additional information, the specific subject to whom the personal data belongs. By means of such anonymization, legal data-processing requirements can be reduced, leading to lower costs for organizations when developing such systems. Thus, the depersonalization of personal data not only protects people from cyber threats, but also has a positive economic effect. This problem was considered in some works [2–5], but peculiarities of data processing in Russian, whose morphology has additional complexities, were not taken into account. Moreover, in these works, the detailing of recognizable CI entities was not adequately carried out, which reduces the quality of impersonal data.

[1] Federal Law No. 152-FZ dated July 27, 2006 "On Personal Data" (in Russ.). https://docs.cntd.ru/document/901990046. Accessed February 09, 2023.

1. TERMS OF REFERENCE

When carrying out data depersonalization, it is necessary to understand what data elements should be hidden. Therefore, we can say that the preliminary stage of depersonalization of CI (in particular personal data) is its separation from all other information. For this purpose, manual extraction of a certain type of information is not only more costly, but also subject to the risk of human error.

Based on above, there is a task of automatic recognition and subsequent CI depersonalization in the data processed and transmitted in the AS. Data can be transmitted in the form of exchanged files, various information flows, etc. In this regard, it is necessary to provide the ability to extract information from files of different extensions and byte representation.

2. CI RECOGNITION

There are several basic automated ways to recognize information, such as vocabulary search, regular expressions, and machine learning algorithms. While the recognition of any kind of information in structured data is quite often solved using rule-based systems, things are not so straightforward with unstructured data. Moreover, there is a large variety of data that directly or indirectly identify a person, such as name, first name and patronymic, passport series and number, phone number. For each type of data, large vocabularies will have to be compiled and constantly updated, along with the encoding of complex rules.

These problems can be solved by using machine learning algorithms to recognize personal data in structured and unstructured information. In particular, the task of personal data recognition is reduced to the task of Named Entity Recognition (NER) [6]. There are several basic ways to solve this problem:

- using statistical methods, for example, according to the number of certain characters;
- using rules based on vocabularies and regular expressions;
- using neural networks.

Statistical methods currently used to perform this task lack sufficient recognition quality, especially when dealing with unstructured data. Rule-based systems, although relatively fast, require more frequent updates and are prone to errors in more complex data, such as organization names, surnames, and first names. In addition, statistical and rule-based approaches do not take context into account. Neural networks can address these shortcomings. For the tasks of natural language processing and, in particular, NER, the most advanced are neural networks with transformer-type architecture [7]. Transformers transform natural language into a numerical vector representation called embedding, which in turn can be processed by machine. Such embeddings, unlike classical vectors, take into account the semantic proximity of token words.

When working with structured data, it is not always necessary to use neural networks to recognize some types of CI—simple rules and statistical methods suffice. Preliminary analysis and separation of data into structured and unstructured allows choosing a suitable recognition algorithm. For recognition of numerical data, regular expressions with check digits are more suitable, especially in structured data. It is also worth noting that some numeric personal data are well recognized by neural networks working with a sufficiently large set of unstructured data. In any case, in order to use machine learning algorithms, it is necessary to prepare a training sampling.

2.1. Data markup

Training sampling consists of data presented in a certain way and labeled with various attributes of CI. The text is divided into tokens, represented by words, which are assigned a tag (label) denoting belonging to a certain type of information.

Tags can be placed according to one of the following schemes:

- BIO/IOB, where B (Begin) is the beginning of the entity; I (Inside) is the continuation of the entity; O (Outside) is not related to the entity;
- BILUO/BILOU [8], where L (Last) is the end of an entity; U (Unit) is a single token entity; B, I, and O are decoded as in the BIO/IOB scheme.

Since the BIO scheme is more commonly-used, it is used in the present work.

Tagging of tokens may differ depending on the problem to be solved. In the Nested Named Entity Recognition (Nested NER) task [9, 10], two tags are assigned to each token: a summary tag and a nested tag. An example of markup is shown in Table 1. Tags manually applied by a qualified expert typically contain an abbreviated meaningful description of the information contained in the token being tagged. For example, the tag B-SNM is an abbreviation of Surname.

Table 1. Token tagging to recognize nested named entities

Token	Consolidated Tag	Nested Tag
Sidorov	B-PERSON	B-SNM
Ivan	I-PERSON	B-FNM
Petrovich	I-PERSON	B-PNM
has concluded	O	O
the contract	O	O
with	O	O
OOO	B-ORG	B-OPF
Romashka	I-ORG	B-ORG_NAME

When recognizing discontinuous named entities (Discontinuous NER) [11], tagging can be represented as a table where the number of columns depends on the maximum number of discontinuities for the discontinuous entity. Thus, the first word in a discontinuous entity is tagged with prefix B, while all subsequent tags are shifted by one column to the right at each discontinuity and prefixed with I (in the case of the BIO scheme).

Since the present work is aimed at solving the classical problem of named entities recognition, so the training sample is divided into words. Each word is matched with a label indicating that it belongs to one or another type of CI. The data set used contains various regulatory documents, memos and other documents involved in the production activities of the organization, which will later be depersonalized.

By tagging the data and training an artificial intelligence (AI) model, it is possible to recognize CI automatically, which in turn opens up the possibility for subsequent automatic depersonalization.

3. DEPERSONALIZATION

Once detected in the CI text, it can be depersonalized in a reversible or irreversible way. In most cases, depersonalization means irreversible implementation; if necessary, it is possible to save the substitution table in a protected loop to obtain reversible depersonalization.

The following depersonalization algorithms are possible in any implementation:

- setting to zero—deleting all or a significant part of the original value;
- replacement by constant—replacement of the original value by a non-zero constant;
- replacement with a value from the reference book—replacement of the original value with a random different value from the reference book, corresponding to the data type to be replaced;
- replacement by a character set—converting each character of the original value into a random character that matches the data type;
- shuffling—shuffling of individual values or groups of values of attributes of personal data in the array of personal data;

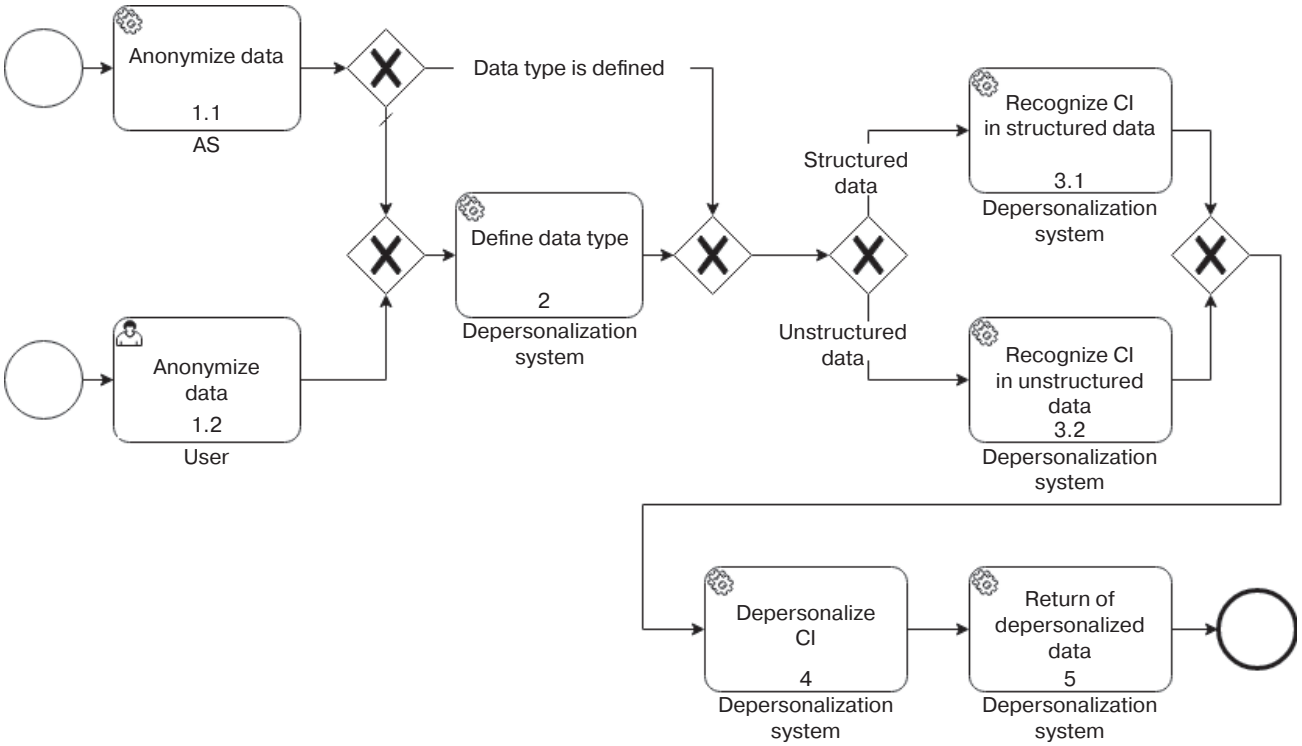

Fig. 1. CI depersonalization process

- blurring the sum and the date—replacing the original value by a random value close to the impersonal value;
- transformation based on a given expression—transformation of the initial value by an expression containing both constants and variables;
- masking—replacing part of the original value with a special character or a set of characters (mask);
- replacement by a random value—replacement of the original value with a randomly generated value;
- generation of pseudo-meaningful meanings—creation of text on the basis of language model or given expressions, allowing the correct text from the point of view of the basic linguistic norms and data parameters to be received. In addition to this method, we can refer generation of photos, taking into account gender and age of the person.

When choosing an approach to depersonalization of personal data, it is worth considering the guidance published by Roskomnadzor[2], according to which the main methods of depersonalization include: the method of introducing identifiers (replacing part of the data by identifiers and creating a table of matching identifiers with the original data); the method of changing the composition or semantics of personal data by replacing them with the results of statistical processing, transformation, summarization, or deleting parts of data; decomposition method (dividing the set of personal data).

[2] Order of Roskomnadzor dated September 05, 2013 No. 996 "On approval of requirements and methods for depersonalization of personal data" (in Russ.). https://rkn.gov.ru/docs/6_Trebovanija_i_metody_po_obezlichivaniju_personal6nykh_dannykh.docx. Accessed February 09, 2023.

Taking into account the recommendations of Roskomnadzor, the most suitable algorithms are pseudo-value generation and constant replacement. To implement reversible depersonalization, it is necessary to create a table of matching source data; here, it should be noted that the table itself should be stored separately from the depersonalized data, with only persons authorized to work with personal data in open form having access to it.

Depending on the problem to be solved, various algorithms may be used. For example, if it is necessary to unambiguously determine that an anonymization was performed and to understand what type of information was removed, a constant replacement algorithm is the best choice. If it is necessary to preserve the length of the value to be replaced at the same time as determining that an anonymization was performed, the partial masking algorithm will handle this task well. In the case where the depersonalized data needs to be used in almost the same way as the original data, for example, for training AI models, the best choice would be an algorithm for generating pseudo-meaningful values.

4. AUTOMATIC DEPERSONALIZATION SYSTEM

In order to work with CI as safely as possible, it is necessary to develop a system of automatic depersonalization. The process of automatic depersonalization by means of the system implemented by the authors of the present work consists of the following tasks (Fig. 1).

1.1 Request to the system for depersonalization according to API from a third-party AS.

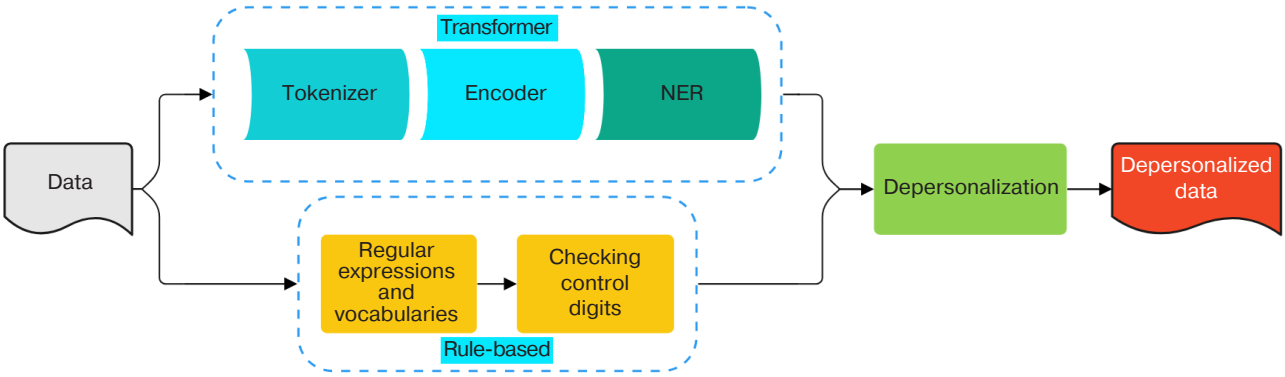

Fig. 2. Data processing by the depersonalization system

1.2 Request to the system for depersonalization through the interface from the user.
2 Data type definition and preprocessing.
3.1 CI recognition in structured data.
3.2 CI recognition in unstructured data.
4 Depersonalization of the recognized CI by the most suitable algorithm.
5 Return of a depersonalized document or data stream.

The need to separate the recognition of personal data in structured and unstructured information in the system arises due to the fact that different machine learning algorithms are used, in particular those using considering and not-considering syntactic features.

4.1. Data preparation

In total, about one million tokens, represented by individual words, were manually tagged by the authors for model training. For the markup we used service documents containing personal data, banking secrets and other CI. A BIO scheme was chosen as the markup scheme, where the first token within a confidential entity is prefixed with B, and all subsequent tokens are prefixed with I. This approach allows most pre-trained architectures to be compared and used, which simplifies the process of creating an AI model, at least in terms of reducing the time to train it.

Resulting set of marked data is divided into 3 parts, where 80% of the data is used to train the model, 10% is used to validate it, and 10% is used to calculate the metrics of the trained model. This is the ratio used, not 60/20/20, because some types of CI in the data set are not sufficient, and it would be irrational to further reduce their number in the training set.

When splitting text into tokens, it is necessary to save the indices of the splitting boundaries in order to anonymize it strictly within the specified boundaries following CI recognition.

4.2. Model training

Most advanced results in the tasks of named entities recognition are shown by neural networks based on transformer architecture. Transformers pre-trained on a large corpus of data are well reused in the tasks of natural language processing [12]. For this purpose, it is sufficient to fine-tune the model on its own data, thereby adjusting weights in order to better take into account the semantics of the input data.

The pre-trained rubert-base-cased model [13] is used as the basis, the use of other suitable architectures does not significantly affect the performance of the model. This is primarily due to the similarity of various transformers used to solve the NER problem, such as BERT [13], RoBERTa [14], and spaCy [15]. The pre-trained model comprises a Tokenizer and Encoder to which a NER classifier is added. In order to improve accuracy and reduce false positives, rule-based recognition algorithms are used, in particular, regular expressions and check digit checking [16]. The results of data processing by neural networks and rules are then summarized into a general assumption that the text belongs to one of the types of CI. A schematic representation of the data depersonalization process by the proposed system is shown in Fig. 2.

Due to the lack of context when processing structured data, preference is given to the rule-based recognition module.

A rule-based CI recognition model without neural networks and a PyTorch model based on a recurrent neural network (RNN) [17] are also implemented in the depersonalization system for the purposes of comparison.

Since most existing depersonalization systems are rule-based and have similar implementations, the rule-based model for CI recognition without neural networks serves to provide a baseline metric against which other solutions can be compared. Comparing an implemented depersonalization system with other implementations will knowingly present the proposed solution in a better light, since third-party implementations were designed for a different, most often structured, data set [18–20]. For example, some third-party systems work only with personal data and do not support bank secrecy depersonalization.

Since all personal data must be recognized and anonymized in the context of this task, the recall metric is important; however, precision is also important to ensure that the number of false positives does not undermine trust in the system. For this reason, the F_1-measure is used, which takes both of these metrics into account, and is calculated by the formula

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

The metrics are calculated using the test part of the marked dataset described in Section 4.1. To begin with, a confusion matrix is constructed, in which the horizontal axis contains the true tags from the markup and the vertical axis contains the tags predicted by the AI model. Then the number of true CI attributes (TP, true positive), the number of true unrecognized attributes (TN, true negative), the number of false attributes recognized (FP, false positive) and the number of false unrecognized attributes (*FN*, false negative) are counted from the error matrix. After that, recall and precision are calculated using the formulas

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

and

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

and then their average harmonic F_1-measure is determined [21].

Table 2 shows the main attributes of CI and calculated weighted average F_1-measure by different models: rule-based model, recurrent neural network and BERT model. It is worth noting that the rule-based implementation works only on the basis of rules, while the other implementations use neural networks together with regular expressions and other rule-based algorithms.

Table 2. Main attributes of CI and calculated weighted average F_1-measure

CI attribute	F_1 (rule-based)	F_1 (RNN)	F_1 (BERT)
Surname	0.804	0.911	0.931
Name	0.819	0.876	0.929
Patronymic	0.874	0.883	0.943
Passport serial number	0.883	0.907	0.906
Authority that issued the passport	0.701	0.794	0.899
Phone number	0.959	0.969	0.967
E-mail	0.955	0.959	0.964
IP-address	0.929	0.932	0.930
Geolocation	0.904	0.919	0.922
Address	0.809	0.810	0.912
Date of birth	0.813	0.837	0.915
TIN	0.918	0.915	0.919
IIAN	0.931	0.935	0.934
OMI policy number	0.921	0.914	0.921
Bank account number	0.937	0.929	0.936
Bank card number	0.967	0.959	0.965
Military ID number	0.892	0.880	0.889
Primary State Registration Number of the Individual Entrepreneur	0.910	0.909	0.919
Job position	0.812	0.820	0.873
Organization name	0.817	0.899	0.951
Average weighted F_1-measure	0.878	0.898	0.926

The RoBERTa and spaCy models were additionally compared. These showed metrics similar to the BERT model with a scatter of F_1-measure values less than 0.01. In this regard, the BERT model was chosen because it is smaller than the RoBERTa model at the same time as having more flexible settings than the spaCy model; this becomes an important factor when implementing an industrial version of the model in a system.

As shown in Table 2, the rule-based solution is significantly inferior to machine learning models in terms of the values of the F_1-measure. The effect is especially noticeable in string data types, where context plays a significant role. Due to the heterogeneous set of documents, the recurrent neural network RNN also performs worse than BERT. Based on the values of the F_1-measure metric presented in Table 2, and the fact that transformer models have wide potential for development and reuse, the BERT model outperforms the other solutions by an average of 4%, for which reason it was selected in the final solution.

The main advantage of the depersonalization system using the BERT model over other solutions is the use of the self-attention mechanism, which allows better detection of CI through the analysis of context and importance of words in the text. The attention mechanism used in the model can be expressed by the formula

$$\text{attention} = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right)\mathbf{V},$$

where **Q** is the query vector; **K** is the key vector; **V** is the value vector; d_k is the dimensionality of vectors.

Vectors **Q**, **K**, and **V** are obtained by multiplying the token embedding by the corresponding matrices obtained by pre-training the rubert-base-cased model taken as the basis. Since in reality calculations performed over vector representations of several tokens **Q**, **K**, and **V** are matrices; therefore, before calculating the product of **Q** and **K**, the matrix **K** must be transposed [7]. In a practical implementation, the key vector and values are the same vector and serve to represent a token, while the query vector shows the significance of a given token with respect to other.

The Softmax function is expressed by the formula

$$\sigma(\mathbf{Z})_i = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}},$$

where i and j are indices of a vector element in the range from 1 to N serves for normalization, i.e., it converts a vector **z** of dimension N to a vector **σ** of the same dimension, where all coordinates of the normalized vector $\boldsymbol{\sigma}_i$ are expressed by a number in the range from 0 to 1, and their sum is equal to one.

CI recognition in unstructured data, represented by images and audio recordings, is reduced to the processing of unstructured texts. For this purpose, the Optical Character Recognition (OCR) [22] and Automatic Speech Recognition (ASR) tasks [23] are solved in advance. [23].

4.3. Depersonalization

Having recognized a CI, the system depersonalizes it using one of the selected algorithms. The choice of algorithm is possible both at the level of the whole document or data set, as well as that of the separate type of CI. The system presented by the authors implements depersonalization algorithms based on the following methods:

- replacement with a constant (placeholder) of the form {Attribute_CI};
- masking to the * symbol;
- pseudo-meaningful value generation, including replacement with a value from a reference, conversion based on a given expression, and date blurring.

For example, having recognized by the AI model in the sentence “Alexander Sidorov (TIN 503199560259) received a transfer to the card 4561 2612 1234 5467” the CI represented by the surname, name, taxpayer identification number (TIN) and the bank card number, the system user can choose one of the depersonalization algorithms described above. When replaced with a placeholder, the sentence in question will take the following form: “{Name} {Surname} (TIN {TIN}) received a transfer to card {Bank Card Number},” where the CI is replaced with constants indicating what type of information was previously in the sentence. With partial masking, the CI is replaced by a mask, and the sentence in question will take the following form: “********* S****** (TIN 50**********) received transfer 4561 26** **** 5467,” where the parts of words that are not dangerous for the identification of the data subject, but which allow the indirect attributes, for example, the bank that issued the card, to be preserved. When substituted with pseudo meanings, the sentence in question will take the following form: “Vladislav Lazarev (TIN 503195234624) received a transfer to card 4561 2698 5513 5467.” The latter algorithm, unlike the previous two, works more slowly, as it generates pseudo meaningful data, but generates a fully meaningful text that can be used, for example, in machine learning tasks.

The selection of the depersonalization algorithm used depends on the task to be solved and is left to the user or the AS.

CONCLUSIONS

A total of about one million tokens are marked for training the AI model, so that a large number of data representation methods containing CI are covered. When the number of types of depersonalized documents is small, it is sufficient to partition a small set of data that includes all the necessary types of CI for model pre-training. Since, due to the use of transformer models, model retraining is not required in most cases, the developed system can be reused in different organizations "as is" or with adjustments on a small volume of data. The use of neural networks permits the removal of huge directories of surnames and names, as well as other data entities used to identify a person. Regular expressions, in turn, take into account structural features, such as existing series, codes and bank identification numbers, which makes it possible to detect even those data on which the model has not previously been trained.

The distinguishing advantage of the depersonalization system presented by the authors from the existing ones is the support of both structured and unstructured data. Moreover, in most known systems, depersonalization is performed in a destructive way, after which the data become unusable for many tasks, for example, for machine learning.

Average weighted F_1-measure of the implemented CI recognition model exceeded 0.9, indicating the high quality of the depersonalization system, which effectively eliminates the need for human labor for CI detection.

Implemented algorithms of depersonalization based on the method of constant replacement, masking and generation of pseudo-meaningful values cover all basic tasks of depersonalization: depersonalization with the possibility of unambiguous determination of the fact of masking, synonymous depersonalization, irreversible and reversible depersonalization, etc. The specified algorithms can also be used to automatically depersonalize the recognized confidential data. The practical value of the automatic depersonalization system developed by the authors lies in the fact that the depersonalized confidential data can be used similarly to the original data, but without the risk of violating cybersecurity requirements. Due to the automation of the process, the cost for depersonalization procedures can be practically reduced to zero.

Confidential data depersonalization system contains at least one processor and one memory connected to the processor, which contains machine-readable instructions. In addition, the depersonalization system may be run on a server, a programmable logic controller, or any other devices capable of executing a given sequence of instructions.

The proposed automatic depersonalization system can be used to automatically recognize and depersonalize personal information at almost any cycle associated with its transfer and processing. Thanks to this, the risk of identity disclosure in case of data leakage is reduced. For example, automatic depersonalization can be used in banking, government services, the data science community [24], and other entities related to the processing of CI, in particular, personal data.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Shabrova A.I., Terenin A.A., Babak N.G. Methodology for risk assessment from confidential information disclosure in data sources using data mining. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. 2022;18(3):666–679 (in Russ.). https://doi.org/10.25559/SITITO.18.202203.666-679
2. Stolbov A.P. De-identification of personal data in health care. *Vrach i informacionnye tekhnologii = Medical Doctor and Information Technologies*. 2017;3:76–91 (in Russ.). Available from URL: https://elibrary.ru/zgyvot
3. Spevakov A.G., Kalutskiy I.V., Nikulin D.A., Shumailova V.A. Depersonalization of personal data during processing of information in automated systems. *Telekommunikatsii = Telecommunications*. 2016;10:16–20 (in Russ.). Available from URL: https://www.elibrary.ru/wwvxmt

СПИСОК ЛИТЕРАТУРЫ

1. Шаброва А.И., Теренин А.А, Бабак Н.Г. Методика оценки риска от разглашения конфиденциальной информации в источниках данных с использованием интеллектуального анализа данных. *Современные информационные технологии и ИТ-образование*. 2022;18(3):666–679. https://doi.org/10.25559/SITITO.18.202203.666-679
2. Столбов А.П. Обезличивание персональных данных в здравоохранении. *Врач и информационные технологии*. 2017;3:76–91. URL: https://elibrary.ru/zgyvot
3. Спеваков А.Г., Калуцкий И.В., Никулин Д.А., Шумайлова В.А. Обезличивание персональных данных при обработке в автоматизированных информационных системах. *Телекоммуникации*. 2016;10:16–20. URL: https://www.elibrary.ru/wwvxmt

4. Oleksy M., Ropiak N., Walkowiak T. Automated anonymization of text documents in Polish. *Procedia Computer Science*. 2021;192(1):1323–1333. https://doi.org/10.1016/j.procs.2021.08.136
5. Saluja B., Kumar G., Sedoc J., Callison-Burch C. Anonymization of Sensitive Information in Medical Health Records. In: *CEUR Workshop Proceedings*. 2019;2421:647–653. Available from URL: https://ceur-ws.org/Vol-2421/MEDDOCAN_paper_2.pdf
6. Roy A. *Recent Trends in Named Entity Recognition* (*NER*). arXiv. 2021. https://doi.org/10.48550/arxiv.2101.11420
7. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. In: *Advances in Neural Information Processing Systems*. 2017. https://doi.org/10.48550/arXiv.1706.03762
8. Ratinov L., Roth D. Design Challenges and Misconceptions in Named Entity Recognition. In: *Proceedings of the Thirteenth Conference on Computational Natural Language Learning* (*CoNLL 2009*). 2009. P. 147–155. Available from URL: https://aclanthology.org/W09-1119.pdf
9. Fisher J., Vlachos A. *Merge and label: A novel neural network architecture for nested NER*. arXiv. 2019. https://doi.org/10.48550/arXiv.1907.00464
10. Fu Y., Tan C., Chen M., Huang S., Huang F. Nested named entity recognition with partially-observed TreeCRFs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021;35(14):12839–12847. https://doi.org/10.1609/aaai.v35i14.17519
11. Dai X., Karimi S., Hachey B., Paris C. *An effective transition-based model for discontinuous NER*. arXiv. 2020. https://doi.org/10.48550/arXiv.2004.13454
12. Lothritz C., Allix K., Veiber L., Klein J., Bissyande T.F.D.A. Evaluating pretrained transformer-based models on the task of fine-grained named entity recognition. In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020. P. 3750–3760. http://doi.org/10.18653/v1/2020.coling-main.334
13. Kuratov Y., Arkhipov M. *Adaptation of deep bidirectional multilingual transformers for Russian language*. arXiv. 2019. https://doi.org/10.48550/arXiv.1905.07213
14. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzman F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. *Unsupervised cross-lingual representation learning at scale*. arXiv. 2020. https://doi.org/10.48550/arXiv.1911.02116
15. Patel A.A., Arasanipalai A.U. *Applied Natural Language Processing in the Enterprise*. O'Reilly Media, Inc.; 2021. 336 p. ISBN 978-1-4920-6257-8. Available from URL: https://spacy.io/universe/project/applied-nlp-in-enterprise/
16. Singco V.Z., Trillo J., Abalorio C., Bustillo J.C., Bojocan J., Elape M. OCR-based Hybrid Image Text Summarizer using Luhn Algorithm with Finetune Transformer Models for Long Document. *Int. J. Emerging Technol. Adv. Eng*. 2023;13(02):47–56. http://doi.org/10.46338/ijetae0223_07
17. Soltau H., Shafran I., Wang M., Shafey L.E. *RNN Transducers for Nested Named Entity Recognition with constraints on alignment for long sequences*. arXiv. 2022. https://doi.org/10.48550/arXiv.2203.03543

4. Oleksy M., Ropiak N., Walkowiak T. Automated anonymization of text documents in Polish. *Procedia Computer Science*. 2021;192(1):1323–1333. https://doi.org/10.1016/j.procs.2021.08.136
5. Saluja B., Kumar G., Sedoc J., Callison-Burch C. Anonymization of Sensitive Information in Medical Health Records. In: *CEUR Workshop Proceedings*. 2019;2421:647–653. URL: https://ceur-ws.org/Vol-2421/MEDDOCAN_paper_2.pdf
6. Roy A. *Recent Trends in Named Entity Recognition* (*NER*). arXiv. 2021. https://doi.org/10.48550/arxiv.2101.11420
7. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. In: *Advances in Neural Information Processing Systems*. 2017. https://doi.org/10.48550/arXiv.1706.03762
8. Ratinov L., Roth D. Design Challenges and Misconceptions in Named Entity Recognition. In: *Proceedings of the Thirteenth Conference on Computational Natural Language Learning* (*CoNLL 2009*). 2009. P. 147–155. URL: https://aclanthology.org/W09-1119.pdf
9. Fisher J., Vlachos A. *Merge and label: A novel neural network architecture for nested NER*. arXiv. 2019. https://doi.org/10.48550/arXiv.1907.00464
10. Fu Y., Tan C., Chen M., Huang S., Huang F. Nested named entity recognition with partially-observed TreeCRFs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021;35(14):12839–12847. https://doi.org/10.1609/aaai.v35i14.17519
11. Dai X., Karimi S., Hachey B., Paris C. *An effective transition-based model for discontinuous NER*. arXiv. 2020. https://doi.org/10.48550/arXiv.2004.13454
12. Lothritz C., Allix K., Veiber L., Klein J., Bissyande T.F.D.A. Evaluating pretrained transformer-based models on the task of fine-grained named entity recognition. In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020. P. 3750–3760. http://doi.org/10.18653/v1/2020.coling-main.334
13. Kuratov Y., Arkhipov M. *Adaptation of deep bidirectional multilingual transformers for Russian language*. arXiv. 2019. https://doi.org/10.48550/arXiv.1905.07213
14. Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzman F., Grave E., Ott M., Zettlemoyer L., Stoyanov V. *Unsupervised cross-lingual representation learning at scale*. arXiv. 2020. https://doi.org/10.48550/arXiv.1911.02116
15. Patel A.A., Arasanipalai A.U. *Applied Natural Language Processing in the Enterprise*. O'Reilly Media, Inc.; 2021. 336 p. ISBN 978-1-4920-6257-8. URL: https://spacy.io/universe/project/applied-nlp-in-enterprise/
16. Singco V.Z., Trillo J., Abalorio C., Bustillo J.C., Bojocan J., Elape M. OCR-based Hybrid Image Text Summarizer using Luhn Algorithm with Finetune Transformer Models for Long Document. *Int. J. Emerging Technol. Adv. Eng*. 2023;13(02):47–56. http://doi.org/10.46338/ijetae0223_07
17. Soltau H., Shafran I., Wang M., Shafey L.E. *RNN Transducers for Nested Named Entity Recognition with constraints on alignment for long sequences*. arXiv. 2022. https://doi.org/10.48550/arXiv.2203.03543

18. Abirkhaev E.A., Erokhin A.F., Pushkin P.Yu. Methods of depersonalizing data: overview and analysis. *Naukosfera*. 2021;6(2):57–31 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=46561812
19. Seryshev A.S., Krotov A.D., Efanova N.V. Development of an application for personal data depersonalization. In: *Digitalization of the Economy: Directions, Methods, Tools: Proceedings of the 3rd All-Russian Scientific and Practical Conference*. Krasnodar: Kuban State Agrarian University; 2021. P. 294–297 (in Russ.). ISBN 978-5-9074-3005-1. Available from URL: https://www.elibrary.ru/item.asp?id=44891383
20. Fot U.D., Korobova E.O. Depersonalization of personal data in the personnel management system of oil and gas sector enterprises. In: *The Role of the Oil and Gas Sector in the Technical and Economic Development of the Orenburg Region: Proceedings of the scientific-practical conference. Saratov: Amirit;* 2021. P. 161–168 (in Russ.). ISBN 978-5-0014-0888-8. Available from URL: https://www.elibrary.ru/item.asp?id=48392659
21. Williams C.K.I. The effect of class imbalance on Precision-Recall Curves. *Neural Computation*. 2021;33(4): 853–857. https://doi.org/10.1162/neco_a_01362
22. Du Y., Li C., Guo R., Yin X., Liu W., Zhou J., Bai Y., Yu Z., Yang Y., Dang Q., Wang H. *PP-OCR: A practical ultra lightweight OCR system*. arXiv. 2020. https://doi.org/10.48550/arXiv.2009.09941
23. Pan J., Shapiro J., Wohlwend J., Han K.J., Lei T., Ma T. *ASAPP-ASR: Multistream CNN and self-attentive SRU for SOTA speech recognition*. arXiv. 2020. https://doi.org/10.48550/arXiv.2005.10469
24. Ryffel T., Trask A., Dahl M., Wagner B., Mancuso J., Rueckert D., Passerat-Palmbach J. *A generic framework for privacy preserving deep learning*. arXiv. 2018. https://doi.org/10.48550/arXiv.1811.04017

18. Абирхаев Е.А., Ерохин А.Ф., Пушкин П.Ю. Методы обезличивальных данных: обзор и анализ. *Наукосфера*. 2021;6(2):57–31. URL: https://www.elibrary.ru/item.asp?id=46561812
19. Кротов А.Д., Серышев А.С., Ефанова Н.В. Разработка приложения для обезличивания персональных данных. В сб.: *Цифровизация экономики: направления, методы, инструменты*: *сб. материалов III всероссийской научно-практической конференции*. Краснодар: Кубанский государственный аграрный университет; 2021. С. 294–297. ISBN 978-5-9074-3005-1. URL: https://www.elibrary.ru/item.asp?id=44891383
20. Фот Ю.Д., Коробова Е.О. Обезличивание персональных данных в системе управления персоналом предприятий нефтегазового сектора. В сб.: *Роль нефтегазового сектора в технико-экономическом развитии Оренбуржья: сб. трудов научно-практической конференции*. Саратов: ООО «Амирит»; 2021. С. 161–168. ISBN 978-5-0014-0888-8. URL: https://www.elibrary.ru/item.asp?id=48392659
21. Williams C.K.I. The effect of class imbalance on Precision-Recall Curves. *Neural Computation*. 2021;33(4): 853–857. https://doi.org/10.1162/neco_a_01362
22. Du Y., Li C., Guo R., Yin X., Liu W., Zhou J., Bai Y., Yu Z., Yang Y., Dang Q., Wang H. *PP-OCR: A practical ultra lightweight OCR system*. arXiv. 2020. https://doi.org/10.48550/arXiv.2009.09941
23. Pan J., Shapiro J., Wohlwend J., Han K.J., Lei T., Ma T. *ASAPP-ASR: Multistream CNN and self-attentive SRU for SOTA speech recognition*. arXiv. 2020. https://doi.org/10.48550/arXiv.2005.10469
24. Ryffel T., Trask A., Dahl M., Wagner B., Mancuso J., Rueckert D., Passerat-Palmbach J. *A generic framework for privacy preserving deep learning*. arXiv. 2018. https://doi.org/10.48550/arXiv.1811.04017

About the authors

Nikita G. Babak, Postgraduate Student, Department of Computing Machines, Systems and Networks, Institute of Information Technologies and Computer Science, National Research University MPEI (14/1, Krasnokazarmennaya ul., Moscow, 111250 Russia); Chief Data Protection Officer, Cybersecurity Department, Sberbank of Russia (19, Vavilova ul., Moscow, 117312 Russia). E-mail: nikita.enrollee@gmail.com. ResearcherID HHY-9372-2022, RSCI SPIN-code 3687-6548, https://orcid.org/0000-0001-7129-1018

Leonid Yu. Belorybkin, Director of Data Protection Projects, Cybersecurity Department, Sberbank of Russia (19, Vavilova ul., Moscow, 117312 Russia). E-mail: lbelorybkin@gmail.com. https://orcid.org/0000-0002-8575-5773

Shamil A. Otsokov, Dr. Sci. (Eng.), Professor, Department of Intelligent Information Security Systems, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: shamil24@mail.ru. Scopus Author ID 57212622267, https://orcid.org/0000-0001-7451-5443

Alexey A. Terenin, Cand. Sci. (Eng.), Managing Director, Cybersecurity Department, Sberbank of Russia (19, Vavilova ul., Moscow, 117312 Russia). E-mail: aaterenin@yandex.ru. http://orcid.org/0000-0002-6242-6117

Anastasia I. Shabrova, Data Protection Architect, Cybersecurity Department, Sberbank of Russia (19, Vavilova ul., Moscow, 117312 Russia). E-mail: shabrova1113@gmail.com. https://orcid.org/0000-0002-4315-3061

Об авторах

Бабак Никита Григорьевич, аспирант, кафедра вычислительных машин, систем и сетей Института информационных и вычислительных технологий ФГБОУ ВО «Национальный исследовательский университет «МЭИ» (111250, Россия, Москва, Красноказарменная ул., д. 14, стр. 1); главный эксперт по защите данных, Департамент кибербезопасности ПАО «Сбербанк России» (117312, Россия, Москва, ул. Вавилова, д. 19). E-mail: nikita.enrollee@gmail.com. ResearcherID HHY-9372-2022, SPIN-код РИНЦ 3687-6548, https://orcid.org/0000-0001-7129-1018

Белорыбкин Леонид Юрьевич, директор проектов по защите данных, Департамент кибербезопасности ПАО «Сбербанк России» (117312, Россия, Москва, ул. Вавилова, д. 19). E-mail: lbelorybkin@gmail.com. https://orcid.org/0000-0002-8575-5773

Оцоков Шамиль Алиевич, д.т.н., профессор, кафедра КБ-4 «Интеллектуальные системы информационной безопасности» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: shamil24@mail.ru. Scopus Author ID 57212622267, https://orcid.org/0000-0001-7451-5443

Теренин Алексей Алексеевич, к.т.н., управляющий директор, Департамент кибербезопасности ПАО «Сбербанк России» (117312, Россия, Москва, ул. Вавилова, д. 19). E-mail: aaterenin@yandex.ru. http://orcid.org/0000-0002-6242-6117

Шаброва Анастасия Игоревна, архитектор по защите данных, Департамент кибербезопасности ПАО «Сбербанк России» (117312, Россия, Москва, ул. Вавилова, д. 19). E-mail: shabrova1113@gmail.com. https://orcid.org/0000-0002-4315-3061

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt