The object of this study is the process of identifying implicit relationships between public services and the process of forming portfolios and groups of services determined on the basis of the formulated principle of combining services. The task to identify groups of interconnected services, which is relevant in view of the processes of digital transformation in the field of public services, was addressed. Expanding portfolios of public services thru the formed groups of interconnected services could make it possible to better take into account the interests of users, as well as simplify the process of development and reengineering of public services. The proposed approach to form groups of services would reduce the cost and complexity of reengineering services or converting them into electronic form.

The result of research is a new principle of combining or grouping services, which is based on the methods of intelligent analysis. The basis of the proposed principle of forming groups of services is a simple and effective mathematical apparatus of applying similarity coefficients to establish the similarity of sets, which has not yet been used in the field of public services. The presence of interrelationships between services was established based on the analysis of Jaccard and Sorensen similarity coefficients. An algorithm for forming groups of services based on a new principle has been developed. The relationship between services was established as a result of the analysis of the composition of services, that is, the documents or data necessary for the provision of these services. This allows applying the proposed approach without restrictions and additional conditions. The developed algorithm also allows for «What-if» analysis of the obtained results. An illustrative example was considered: for a set of 10 services, 4 groups of services are formed. Suggestions for applying the proposed approach are provided

Keywords: information technology of intelligent analysis, algorithm of grouping public services, service portfolio, Jacquard similarity coefficient, Sorensen similarity coefficient

UDC 004.021:004.9
DOI: 10.15587/1729-4061.2023.280218

ESTABLISHING THE GROUPING PRINCIPLE OF PUBLIC SERVICES BASED ON THE ANALYSIS OF SIMILARITY COEFFICIENTS

Olena Gavrilenko
PhD, Associate Professor*
Oleksandr Khomenko
Postgraduate Student*
Oksana Zhurakovska
PhD, Associate Professor*
Alla Kohan
Corresponding author
PhD, Associate Professor*
Roman Matviichuk**
Andrii Piskun**
Yuliia Khavikova**
*Department of Information Systems and Technologies
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»
Beresteiskyi ave., 37, Kyiv, Ukraine, 03056
**«EU4DigitalUA» e-Governance Academy (EGA)
Baseina str., 21 a, Kyiv, Ukraine, 01004

Received date 17.03.2023
Accepted date 25.05.2023
Published date 30.06.2023

How to Cite: *Gavrilenko, O., Khomenko, O., Zhurakovska, O., Kohan, A., Matviichuk, R., Piskun, A., Khavikova, Y. (2023). Establishing the grouping principle of public services based on the analysis of similarity coefficients. Eastern-European Journal of Enterprise Technologies, 3 (3 (123)), 22–29. doi: https://doi.org/10.15587/1729-4061.2023.280218*

1. Introduction

One of the development priorities of EU countries is digitalization of society and economy. The development of e-governance and digital economy, the issue of digital transformation is receiving more and more attention. Digitalization of public services is an important component of e-government. Research in the field of digitization of public services and the implementation of the results of these studies will make public services more accessible to the population and business, improve the quality of public services, which will increase the competitiveness of the country's economy as a whole.

Basic concepts in the field of public service provision are public, electronic public, and complex services.

According to [1], a public service is defined as an action that has legal or social significance. As a result of this action, the subject provides a public service. This service is provided upon application (appeal) or the absence of such an application from the subject of the application. As a result of this action, the subject of the appeal acquires rights and/or obligations, or such rights are transferred or terminated. In addition to the rights of the subject of appeal, material or immaterial benefits may be provided as a result of the specified action.

Law [2] defines electronic public services provided by state bodies, local self-government bodies, enterprises, institutions and organizations under their management.

The basis for providing a public or electronic public service is an event that represents some life situation or business event.

To receive public services, users contact service centers (in some cases – directly the entities providing these services) or the corresponding portal or mobile application. They submit an application (if the service is not proactive), fill out forms, and submit documents for verification. In some cases, users are forced to separately apply for several services related to one life or business event or situation. Therefore, the provision of such services to the user on the basis of one application, for one

request, greatly simplifies the process of providing services, and to ensure the convenience of users, complex services are created. Complex services, according to [2], contain a set of services related to one life (business) event or situation and are provided on the basis of one application or request. In various countries, the approach of combining services based on life events or situations is successfully applied [3].

Conducted studies, which analyzed data on the provision of public services, show that there are cases of users applying for several services at the same time. These services are not related to a certain life event or situation, or the list of documents submitted by applicants for different services is the same or very similar [4, 5]. The detection of such services is possible based on the analysis of statistical information on the provision of services and the use of mathematical apparatus to detect hidden dependences between services. We shall use the term «group of services» for such interrelated services, united according to any other principle than a life (business) event or situation. Groups of services formed in this way can be included in the portfolio of services, in case of finding practical utility for the applicants. Otherwise, the formation of such groups of services will significantly simplify the process of reengineering public services.

The formation of a set of connected services is a task of finding hidden dependences in large data sets, which can be solved by methods of statistical analysis and data mining.

Therefore, the establishment of principles that make it possible to identify interdependent services, group them and, if necessary, include them in a service portfolio, is an urgent task.

2. Literature review and problem statement

The development of the country's economy is directly related to the processes of digital transformation in all its sectors. In particular, in the field of public services, the process of digital transformation involves the transfer of public services online, which is one of the most priority tasks in this field.

Work [6] revealed the meaning of the concept of digitalization in the context of state power and public administration, established its interdisciplinary nature and relatively short history of application. In [7] it is shown how digital technologies can be implemented in the activities of companies. Paper [8] provides an analysis of the implementation and prospects of digitalization, as well as its features in the field of public administration. But it should be noted that these works provide analysis and recommendations of a general nature, while the field of public services needs additional research so that the implementation of digitalization in it is as effective as possible.

Works [9, 10] provide a system of indicators for analyzing the state of digital development in various areas of the economy. Average indicators for the European Union, calculated for 28 member states, are given in [9]. In [10], it was noted that all EU member states have made progress in the field of digitalization but the overall picture across member states is ambiguous, and it is necessary to make joint efforts to achieve the goals by 2030. In [11], it is emphasized that accelerating the pace of digital innovations and their implementation in Europe requires promoting digitalization and overcoming imbalances in digitalization between regions and countries. In [12], the relevance of the implementation of the digitalization process in all spheres of the economy, in particular in the sphere of public services, is substantiated. This follows from the fact that the task of transferring all public services online is one of the most important goals of the EU countries in the field of digital transformation. But it should be noted that in these works the specified problem is considered at a general level. Specific ways of solving it are not given, and an important aspect of transferring public services online, namely the issue of creating comprehensive services, is not investigated.

In [13] it is noted that the development of the field of public services, the transfer of services to electronic form also involves the creation of complex services that combine services based on one life situation. However, other possibilities of grouping services in the process of reengineering and creation of new services are not considered.

In the modern user-oriented model of the electronic state, which is used in various countries of the world, for example, in the United States of America, Great Britain, Denmark, the approach of combining services by life event is successfully used [3]. The same approach is also at the basis of the formation of complex services in Ukraine and Estonia: eMalyatko, ID14, E-Business, and E-Tax are examples of successfully functioning integrated services in these countries. The formation of complex services based on the principle of unification based on a life event or situation is defined by laws [1, 2]. The reports of international experts of the EU4DigitalUA project (Ukraine & Estonia) [14] consider the service design method based on life events, as well as the task of forming a branch portfolio of services. However, it should be noted that the possibility of combining services according to other principles was not considered in the studies, and the issue of creating integrated services was generally not given enough attention.

In works [4, 5], in addition to the principle of combining services by life event, it is proposed to apply other principles of combining services into a portfolio. Thus, in [4], the principle of combining services on the basis of statistical information on the provision of services is considered. The combination of services is carried out on the basis of the analysis of correlations between them and the selection of statistically significant connections. The proposed principle underlies the formation of recommendations regarding the composition of the service portfolio. In [5], the principle of combining services based on associative rules is considered. The application of the apparatus of associative rules made it possible to formulate recommendations regarding the composition of the portfolio of services, taking into account implicit, hidden dependences between services. However, it should be noted that the limitations of the approaches proposed in both papers are the need to collect and use statistical data on the provision of services. The completeness and quality of this information significantly affects the quality of the obtained solutions.

Therefore, it can be concluded that the task of identifying new principles of combining services on the basis of hidden dependences between them requires research. On the basis of these principles, groups of services will be formed, which, if appropriate, will be included in the portfolio of services.

3. The aim and objectives of the study

The purpose of the study is to formulate the principle of grouping services based on the identification of hidden relationships between services as a result of the analysis of the composition of services, i.e., the set of documents required for the provision of these services. This will make it possible to form groups of services, applying the principle (criterion) of connectivity of services, in case of expediency, to introduce portfolios of services based on the formed groups of services

according to the new principle. This will ensure the benefits of using integrated services, namely the simplification of the process of providing public services. In addition, the grouping of services according to the new principle will increase the efficiency of the processes of service reengineering and the transfer of services into electronic form by simplifying the process of developing the corresponding software.

To accomplish the aim, the following tasks have been set:

– to solve the problem of establishing the existence of a relationship between services based on the analysis of the composition of services;

– to develop an algorithm for grouping services based on the analysis of their composition;

– to perform intelligent analysis for the generated dataset.

4. The study materials and methods

The object of our study is the process of grouping services according to the new principle of unifying services based on the analysis of information about the composition of services, namely the set of documents (hereinafter – data) necessary for obtaining each service.

The information needed to analyze the presence of relationships between services from the point of view of their composition is simple, does not require additional processing, and is available. For this, it is necessary to know what documents the user must provide in order to receive the corresponding service. Or what documents (information, data) necessary for obtaining a public service are in the possession of public authorities, enterprises, institutions and organizations belonging to the sphere of their management. Thus, each service can be represented by a set of documents included in its composition, and the analysis of relationships between services consists in determining the similarity score of sets based on the use of similarity coefficients.

The application of a mathematical apparatus using similarity coefficients makes it possible to use information about the composition of services to determine the relationships between services from the point of view of the sets of documents necessary to obtain them. Therefore, to form a set of data (dataset), it is necessary to select a set of services. Each service must be provided with a set of documents that are necessary for its receipt, i.e., actually constitute the composition of this service.

Data for the formation of the dataset can be obtained on the portal «Guide to public services» [15]. They include information about the set of documents that the user must submit to the center for the provision of administrative services (entity for the provision of administrative services) in order to receive the corresponding service.

10 services containing a total of 11 documents were selected for analysis. A dataset was built; the general view is given in Table 1.

Table 1

Input

No.	Service name	Composition of the service (documents that must be submitted to receive the service)
1	Service 1	Document 1, document 2, document 3, document 4, document 6
2	Service 2	Document 1, document 2, document 3, document 4, document 5
3	Service 3	Document 1, document 2, document 3, document 4, document 7
4	Service 4	Document 1, document 2, document 8, document 11
5	Service 5	Document 1, document 2, document 3, document 4, document 5, document 6
6	Service 6	Document 3, document 4, document 5, document 6
7	Service 7	Document 3, document 4, document 5, document 6, document 8, document 9
8	Service 8	Document 3, document 4, document 5, document 6, document 7, document 9
9	Service 9	Document 4, document 7, document 8, document 9, document 10
10	Service 10	Document 1, document 4, document 8, document 9, document 10

The rows of Table 1 provide data on the composition of each service. Column 2 lists the names of the analyzed services, column 3 lists the documents required to obtain this service.

The dataset formed in this way is the input information for the service grouping algorithm.

Fig. 1 shows a model of conducting research for the purpose of grouping interrelated services.

The basis of the proposed principle of grouping services is the establishment of similarities between services using methods of intelligent analysis. The methods of intelligent data analysis involve the discovery of hidden patterns in the data. In the study, it is proposed to apply an algorithm for finding relationships between services based on similarity coefficients.

Jaccard and Sorensen similarity coefficients were used to test the presence of a relationship between each pair of services. The inclusion of services in a group is based on the analysis of pairwise values of the corresponding similarity coefficients, and the formation of groups takes place in two stages, at each of which different similarity coefficients are applied. The preliminary selection of candidates for merging is based on the analysis of the values of the Jaccard coefficient, and the final formation of groups from services with a high degree of similarity is carried out on the basis of the analysis of the values of the Sorensen coefficient. In addition, at the second stage, groups of services are formed, which can be recommended for inclusion in the portfolio of services, in case of finding practical usefulness for applicants.

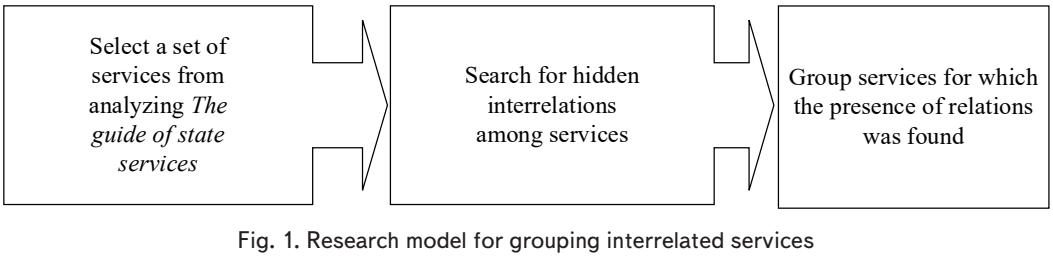

Fig. 1. Research model for grouping interrelated services

The use of the specified methods of intelligent analysis guarantees obtaining reliable results when grouping services. No hypotheses and assumptions were used to conduct research since the method of finding interrelated groups of services based on the composition of these services using similarity coefficients is universal and can be applied to any set of services without any restrictions or assumptions.

5. Research results of the problem of analysis of similarity coefficients for the formation of groups of services

5. 1. Solving the problem of establishing the existence of a relationship between services based on the analysis of the composition of services

Mathematical statement of the problem: a set of services $P=\{P_j\}$, $j=\overline{1,m}$. is given. Each service P_j from the set P is represented by a set of documents included in its composition:

$$P_j=\left\{P_{j1},\ \ldots,P_{ji},\ \ldots,P_{jn_j}\right\},$$

where P_{ji} is the i-th document that is part of the documents required to provide the service P_j, $j=\overline{1,m}$, $i=\overline{1,n_j}$, n_j is the number of documents included in the service P_j, m is the number of services.

It is necessary to establish the relationship between the services of the set P on the basis of information about the composition of services.

Theoretical justification of the approach: each service P_j from the set P is a set of documents that the user must provide to receive this service. Then the task of establishing the relationship between services is the problem of establishing the similarity of sets. Therefore, the presence of a relationship between two sets P_i and P_j can be established based on the analysis of the similarity coefficients of these sets.

The similarity coefficient (similarity measure, similarity index) is a dimensionless indicator of the similarity of the compared objects. Consider the similarity coefficients of Jaccard and Sorensen [16, 17]. The Jaccard similarity coefficient is used to evaluate the similarity of finite sets in ecology, in computer science, to search for similar documents, plagiarism, etc. The Sorensen coefficient is used to solve the problems of evaluating the similarity of sets in medicine, computer science, ecology, and image segmentation. The Jaccard similarity coefficient for a pair of sets P_i, P_j can be calculated using the formula:

$$J\left(P_i,P_j\right)=\frac{\left|P_i\cap P_j\right|}{\left|P_i\cup P_j\right|}=\frac{\left|P_i\cap P_j\right|}{\left|P_i\right|+\left|P_j\right|-\left|P_i\cap P_j\right|},\qquad(1)$$

where $\left|P_i\right|$ is the power of the set P_i.

In general, the value of the Jaccard similarity coefficient is within $0\le J(P_i, P_j)\le 1$. If the sets are equal, then $J(P_i, P_j)=1$, and if the sets do not have any element in common, then $J(P_i, P_j)=0$.

The Sorensen similarity coefficient for a pair of sets P_i, P_j can be calculated using the formula:

$$S\left(P_i,P_j\right)=\frac{2\left|P_i\cap P_j\right|}{\left|P_i\right|+\left|P_j\right|},\qquad(2)$$

where $|P_i|$ is the power of the set P_i.

In general, the value of the Sorensen similarity coefficient is within $0\le S(P_i, P_j)\le 1$. If the sets are equal, then $S(P_i, P_j)=1$, and if the sets do not have any element in common, then $S(P_i, P_j)=0$.

To apply both coefficients in order to identify the similarity of services, which are considered as a set of documents, it is advisable to determine the ratio between these coefficients.

Assertion. For the similarity coefficients of Jaccard and Sorensen, the following inequality holds:

$$J\left(P_i,P_j\right)\le S\left(P_i,P_j\right).\qquad(3)$$

Proof. We introduce the following notation:

$$\left|P_i\cap P_j\right|=a,$$

$$\left|P_i\cup P_j\right|-\left|P_i\cap P_j\right|=b.$$

Then execute:

$$\left|P_i\cup P_j\right|=a+b.$$

According to the inclusion-exclusion formula:

$$\left|P_i\cup P_j\right|=\left|P_i\right|+\left|P_j\right|-\left|P_i\cap P_j\right|.$$

So, then it is executed:

$$\left|P_i\right|+\left|P_j\right|=2a+b.$$

We write the coefficients of Jacquard (1) and Sørensen (2), using the introduced notation:

$$J\left(P_i,P_j\right)=\frac{a}{a+b},\qquad(4)$$

$$S\left(P_i,P_j\right)=\frac{2a}{2a+b}=\frac{a}{a+\frac{b}{2}}.\qquad(5)$$

Therefore, it follows from equalities (4) and (5) that the value of the Jaccard coefficient will not exceed the corresponding value of the Sorensen coefficient, that is, inequality (3) is valid. Exactly what had to be proved.

Thus, services for which the values of the Jaccard or Sorensen similarity coefficients are as close as possible to 1, i.e., 0.5 and above for the Jacquard coefficient and 0.7 and above for the Sorensen coefficient, should be considered interconnected or similar. Let's enter the notation:

– $t_J=0.5$ – the threshold value of the Jaccard similarity coefficient;

– $t_S=0.7$ – the threshold value of the Sorensen similarity coefficient.

It should be noted that the indicated threshold values for both coefficients may vary depending on the set goals.

5. 2. Development of an algorithm for grouping services based on the analysis of their composition

The basis of the formation of groups of services is the verification of the existence of interrelationships between services based on the analysis of their composition. This can be established by examining the values of the similarity coefficients for pairs of services. At the same time, at the first stage, the selection of interconnected services is carried out as a result of the analysis of the values of the Jaccard coefficient (1) for pairs of services. At the second stage, these results are refined based on the analysis of the values of the Sorensen coefficient (2) for pairs of services, and the final formation of the groups of services selected after the first stage is carried out. At the same time,

pairs of services that were selected at the first stage but were not included in the groups at the second stage, are included in the list of recommendations. They are proposed for consideration, from the point of view of the expediency of including relevant services in groups.

The algorithm for forming groups of interconnected services uses the following rules:

Rule 1. If F_i is a group of services, and P_k is some service that is a candidate for inclusion in this group, then if the condition is fulfilled:

$$\forall P_j \in F_i \left[S\left(P_j, P_k\right) \geq t_S \right], \tag{6}$$

where $S(P_j, P_k)$ is the value of the Sorensen coefficient for a pair of services P_j, P_k, determined by relation (2), t_S is the threshold value of the Sorensen similarity coefficient, it is necessary to include the service P_k in the F_i group.

Rule 2. If F_i is a group of services, and P_k is some service that is a candidate for inclusion in this group, and condition (6) is not fulfilled, then it is necessary to check the fulfillment of the condition:

$$\exists F' \subset F_i \left[\begin{array}{l} \left(\forall P_j \in F' \left[S\left(P_j, P_k\right) \geq t_S \right] \right) \wedge \\ \wedge \left(\forall P_j \in F_i \setminus F' \left[S\left(P_j, P_k\right) < t_S \right] \right) \end{array} \right], \tag{7}$$

where $S(P_j, P_k)$ is the value of the Sorensen coefficient for a pair of services P_j, P_k, which is determined by relation (2), t_S is the threshold value of the Sorensen similarity coefficient.

If condition (7) is fulfilled, it is necessary to form a new group of services $F_{i+1} = F' \cup \{P_k\}$.

Fig. 2 shows the scheme of the proposed service group formation algorithm.

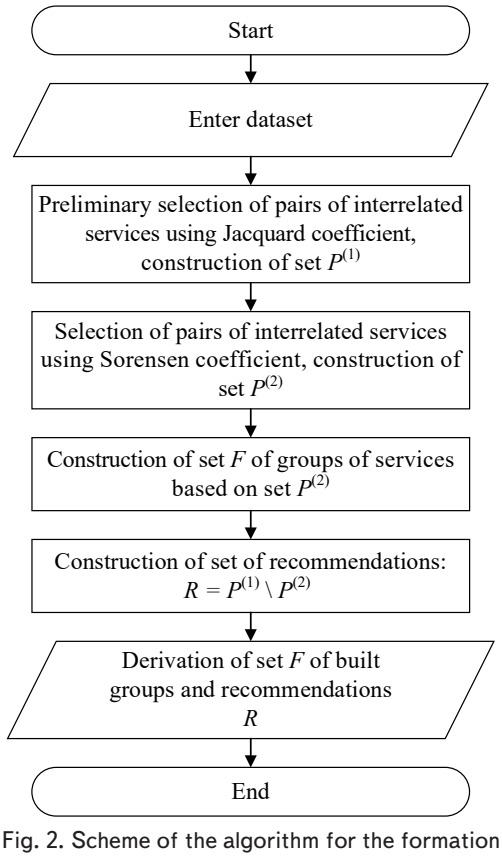

Fig. 2. Scheme of the algorithm for the formation of groups of services

We shall describe the main steps of the algorithm:

Step 1. Formation of the dataset is carried out on the basis of the Register of administrative services posted on the «Guidance on public services» [15]. The name of the services is given in the «Name» field, and the «Input» field contains a list of documents required for the provision of services, on the basis of which a corresponding set $\left\{P_{j1}, \ldots, P_{ji}, \ldots, P_{jn_j}\right\}$ is formed for each service $P_j \in P$.

Step 2. Pre-selection of pairs of interrelated services is carried out based on the analysis of the values of the Jaccard coefficient (1), as a result of which the set $P^{(1)}$ is formed:

$$P^{(1)} = \left\{ \left(P_i, P_j\right) \middle| J\left(P_i, P_j\right) \geq t_J, P_i, P_j \in P, 1 \leq i < j \leq m \right\},$$

where $J(P_i, P_j)$ is the value of the Jaccard coefficient for a pair of services (P_i, P_j), t_J is the threshold value of the Jaccard coefficient, m is the number of services.

Step 3. The selection of pairs of interconnected services for the formation of service groups is carried out based on the analysis of the values of the Sorensen coefficient (2), resulting in the formation of a set $P^{(2)}$:

$$P^{(2)} = \left\{ \begin{array}{l} \left(P_i, P_j\right) \middle| S\left(P_i, P_j\right) \geq t_S, \\ \left(P_i, P_j\right) \in P^{(1)}, 1 \leq i < j \leq m \end{array} \right\},$$

where $S(P_i, P_j)$ is the value of the Sorensen coefficient for a pair of services (P_i, P_j), t_S is the threshold value of the Sorensen similarity coefficient.

Step 4. The formation of the set F – groups of services is carried out on the basis of the set $P^{(2)}$ by applying rules 1 and 2.

Step 5. The formation of the set of pairs recommended for consideration is carried out according to the ratio:

$$R = P^{(1)} \setminus P^{(2)}.$$

As a result of the operation of the algorithm, a set of groups of services F is formed, as well as R – a set of pairs of services that are offered for consideration and that can be grouped, if it is deemed appropriate.

In addition, if necessary, the resulting set can be expanded or narrowed, for which it is necessary to re-apply the proposed algorithm with other values of the thresholds t_J and t_S.

So, to expand the set of solutions, it is necessary to reduce the threshold value t_S (or both values t_J and t_S), but in such a way that the condition $t_J \leq t_S$ is fulfilled.

To narrow down the set of solutions, it is necessary to increase the threshold values of t_J and t_S, but in such a way that the condition $t_J \leq t_S < 1$ is fulfilled.

5. 3. Conducting intelligent analysis for the generated dataset

To establish the existence of a relationship between each pair of services, matrices of paired similarity coefficients were constructed. The input dataset is given in Table 1.

To form the set $P^{(1)}$ at the previous stage of selection, the values of the Jaccard coefficient were calculated (Table 2).

It should be noted that for service 4, the value of pairwise Jaccard coefficients with all other services is less than 0.5, so the set $P^{(1)}$ will not contain any pair containing service 4. Therefore, the set $P^{(1)}$ takes the form:

$$P^{(1)} = \left\{ \begin{array}{l} (1,2),(1,3),(1,5),(1,6),(2,3),(2,5),(2,6),(3,5), \\ (5,6),(5,7),(5,8),(6,7),(6,8),(7,8),(9,10) \end{array} \right\}.$$

To form the set $P^{(2)}$, the values of the Sorensen coefficient were calculated (Table 3).

Table 2

Matrix of Jacquard coefficient values

Service No.	1	2	3	4	5	6	7	8	9	10
1	1	2/3	2/3	2/7	5/6	1/2	3/8	3/8	1/9	1/4
2	2/3	1	2/3	2/7	5/6	1/2	3/8	3/8	1/9	1/4
3	2/3	2/3	1	2/7	4/7	2/7	2/9	3/8	1/4	1/4
4	2/7	2/7	2/7	1	1/4	0	1/9	0	1/8	2/7
5	5/6	5/6	4/7	1/4	1	2/3	1/2	1/2	1/10	2/9
6	1/2	1/2	2/7	0	2/3	1	2/3	2/3	1/8	1/8
7	3/8	3/8	2/9	1/9	1/2	2/3	1	5/7	3/8	3/8
8	3/8	3/8	3/8	0	1/2	2/3	5/7	1	3/8	2/9
9	1/9	1/9	1/4	1/8	1/10	1/8	3/8	3/8	1	2/3
10	1/4	1/4	1/4	2/7	2/9	1/8	3/8	2/9	2/3	1

Table 3

Matrix of Sørensen coefficient values

Service No.	1	2	3	5	6	7	8	9	10
1	1	4/5	4/5	10/11	2/3	6/11	6/11	1/5	2/5
2	4/5	1	4/5	10/11	2/3	6/11	6/11	1/5	2/5
3	4/5	4/5	1	8/11	4/9	4/11	6/11	2/5	2/5
5	10/11	10/11	8/11	1	4/5	2/3	2/3	2/11	4/11
6	2/3	2/3	4/9	4/5	1	4/5	4/5	2/9	2/9
7	6/11	6/11	4/11	2/3	4/5	1	5/6	6/11	6/11
8	6/11	6/11	6/11	2/3	4/5	5/6	1	6/11	4/11
9	1/5	1/5	2/5	2/11	2/9	6/11	6/11	1	4/5
10	2/5	2/5	2/5	4/11	2/9	6/11	4/11	4/5	1

Based on the data from Table 3, the set $P^{(2)}$ is constructed:

$$P^{(2)} = \left\{\begin{matrix}(1,2),(1,3),(1,5),(2,3),(2,5),\\(3,5),(5,6),(6,7),(6,8),(7,8),(9,10)\end{matrix}\right\}.$$

Based on the set $P^{(2)}$, after applying rules 1 and 2, a set of service groups is formed:

$F=\{F_1, F_2, F_3, F_4\}$,

where $F_1=\{1, 2, 3, 5\}$, $F_2=\{5, 6\}$, $F_3=\{6, 7, 8\}$, $F_4=\{9, 10\}$.

In addition, a number of pairs of services were obtained, which are a recommendation for consideration and possible grouping of services:

$R=\{(1,6), (2,6), (5,7), (5,8)\}$.

Thus, as a result of the operation of the algorithm, four groups of services were formed for the threshold values $t_J=0.5$ and $t_S=0.7$:

– group 1: service 1, service 2, service 3, service 5;
– group 2: service 5, service 6;
– group 3: service 6, service 7, service 8;
– group 4: service 9, service 10.

We also received a recommendation to consider the feasibility of grouping services into four groups based on the established relationship between services: $R=\{(1,6), (2,6), (5,7), (5,8)\}$.

If the threshold values $t_J=0.7$ and $t_S=0.9$ are set, two groups of services will be formed as a result:

– group 1: service 1, service 5;
– group 2: service 2, service 5.

It is recommended to consider the feasibility of creating a group based on the established relationship between services: $R=\{(7,8)\}$. When forming a portfolio of services, it is necessary to take into account the received groups of services.

6. Discussion of results of investigating the problem of forming groups of services based on the analysis of similarity coefficients

The choice of the type of information and the structure of the formed dataset (Table 1) is due to the methods of intelligent analysis that allow solving the task set.

Solving the problem of establishing the existence of a relationship between services using the analysis of Jacquard and Sorensen similarity coefficients is due to the fact that each service is considered as a set of documents included in its composition. Then the mathematical apparatus for determining the similarity of sets, which is based on the application of similarity coefficients, can be applied to the solution of the given problem.

The developed scheme of the service group formation algorithm includes the stage of preliminary selection of interrelated services based on the analysis of Jaccard similarity coefficients (1). The next stage of the final formation of groups is based on the analysis of Sorensen similarity coefficients (2). This can be explained by the fact that the established ratio between the similarity coefficients (3) made it possible to conclude on the expediency of using the Jaccard coefficient in the preliminary selection of interconnected services. For the final formation of groups of services that have a high degree of similarity, it is advisable to use the Sorensen coefficient.

As a result of the intelligent analysis of the dataset, sets of services with a high degree of similarity are formed. This is due to checking the values of the Jaccard (1) and Sorensen (2) similarity coefficients that exceed the set threshold values t_J and t_S for these coefficients.

As a result of applying the developed algorithm to the dataset (Table 1), four groups of services were formed:

– group 1: service 1, service 2, service 3, service 5;
– group 2: service 5, service 6;
– group 3: service 6, service 7, service 8;
– group 4: service 9, service 10.

This composition of groups can be explained by the application of rule 1 and rule 2, which are designed specifically for the formation of groups of services based on information about the similarity of services.

In addition, as a result of the service group formation algorithm, recommendations were provided in the form of a list of interrelated services. They can be taken into account after analyzing the expediency of including these services in groups based on the analysis of their composition. The content of the received recommendations is due to their inclusion of pairs of services for which only one of the conditions is met – the Jaccard coefficient exceeding the set threshold value.

A number of advantages of our study should be noted.

The basis of the proposed principle of forming groups of services is a simple and effective mathematical apparatus of applying similarity coefficients to establish the similarity of sets, which has not yet been used in the field of public services. It was used mainly in computer science, botany, ecology, etc. [16, 17].

The advantage of the proposed approach is also that, unlike the current principle of forming complex services based on life events [1–3], its application can potentially allow expanding the list of existing complex services. It will also allow the

introduction of new electronic services (portfolios of services) that have a similar composition, that is, a list of documents required for their receipt. In the future, this will allow users to receive several services on the basis of one application, which corresponds to one of the main principles of the development of public services – orientation to the interests of users. It is important to note that the proposed approach can reduce the cost of reengineering services or conversion to electronic form. It will also make it possible to simplify the development process due to the fact that service groups are formed as a result of combining services that have a similar composition. The proposed principle of grouping services can be used in the creation of software that will allow analysis of services in the process of their reengineering. This will simplify and speed up the processes of reengineering and creation of public services.

Another advantage of the approach of grouping services based on the analysis of their composition is the information necessary for its implementation. In this case, when grouping services, it is necessary to obtain a list of documents necessary for its provision for each service. This information is simple and accessible, in contrast to the principles of grouping services based on statistical information about the provision of services. Thus, information on the number of orders for each service for a certain period of time cannot be obtained from open access, and the result may vary depending on the region for which the analysis is carried out. Information about the order of services by different users for a certain period of time is difficult to access, and the result may also depend on the region for which the analysis is conducted. In contrast to these approaches, when applying the proposed principle of forming portfolios from groups of services, the relevance of the received recommendations does not depend on the quality and quantity of statistical data.

In addition, an additional advantage of the proposed principle is the possibility of conducting a «What-if» analysis, which involves the possibility of analyzing how changes in the parameters of the group formation algorithm affect the solution. This is implemented by using threshold values for similarity coefficients. By changing the specified threshold values, it is possible to influence the result by expanding or reducing the composition of the groups, and also increase the degree of similarity of the services that should be combined into groups, if necessary.

A significant advantage of the proposed approach is also the possibility of its application in the case when each service is represented not by a set of documents but by a set of data and information necessary for the provision of this service. Such a transition from documents to data is foreseen within the methodology of reengineering the system of public (administrative) services [14].

It should be noted that the use of similarity coefficients makes it possible to establish links between services that are similar in terms of the documents that the user must provide to receive a specific service. This principle of combining services is the basis of the developed algorithm for forming groups of services. It allows you to conduct an intelligent analysis and form groups of services, as well as recommendations for including services in groups and form a portfolio of services from the received groups, if necessary.

As a limitation of the proposed approach, it can be noted that the results obtained in the form of sets of related services are recommendations, and the final decision on the formation of portfolios of services should be made by authorized public authorities.

As a disadvantage of the proposed approach, it is advisable to note the possibility of not obtaining a result at sufficiently high (close to one) threshold values for similarity coefficients. But in this case, to obtain a non-empty set of formed groups of services, it is necessary to repeat the analysis with lower threshold values of similarity coefficients.

As a development of the proposed approach, it is advisable to investigate the issue of establishing the relationship between services using other similarity coefficients: Simpson, Kulchytskyi, Brown-Blanke, and Otiai [17]. This will make it possible to increase the relevance and variety of groups of services received within the framework of this approach.

Further research in this area involves consideration of other principles of service grouping.

7. Conclusions

1. The solution to the problem of establishing the existence of a relationship between services is based on the analysis of the similarity of the composition of services. It is proposed to consider services as a set of documents, which allows applying a mathematical apparatus for determining the similarity of sets, which is based on the application of Jaccard and Sorensen similarity coefficients. The practical significance of the obtained result is in the use of Jaccard and Sorensen similarity coefficients in the service grouping algorithm based on combining services with a similar composition.

In addition, the relationship between the specified coefficients was analyzed. Based on this, it was concluded that the Jacquard similarity coefficient should be used to establish the relationship between services if more «coarse» recommendations for grouping are to be made. The purpose of this is to reduce the search space or reduce the size of the initial dataset. And for more accurate recommendations, it is worth using the Sorensen coefficient, which was reflected in the developed algorithm.

It should be noted that the establishment of the relationship between services based on the analysis of similarity coefficients is carried out on the basis of simple and publicly available information about the composition of services. This is an additional advantage, as opposed to establishing the relationship between services based on the analysis of statistical information on the provision of services, which is not publicly available.

2. The developed algorithm solves the problem of grouping services by identifying hidden relationships between services. For this purpose, it is proposed to apply the methods of intelligent analysis. The principle of taking into account hidden dependences between services used in the algorithm allows combining or grouping services that are not taken into account when creating complex services based on life events. They also cannot be detected based on the analysis of statistical information on the provision of services.

As a result of the algorithm, groups of services are formed that have a high degree of similarity in terms of their composition, which guarantees the correctness and reliability of the results. Recommendations for grouping services are formed on the basis of the identified connections. At the same time, the desired degree of similarity between services can be adjusted by changing the threshold values for the similarity coefficients used in the algorithm.

The practical value of the obtained results is that the formation of service groups using the proposed algorithm allows you to use them at the stage of service development or reengineering. Formation, in the case of practical expediency, of portfolios from identified groups of services will enable users to order services included in the group, provide documents once and receive service results faster. This will relieve the work of authorities that provide services, as well as significantly simplify

the process of obtaining service results for users. In addition, the grouping of services based on this principle can significantly simplify the process of service development and reengineering, since in this case services with a similar composition will be considered. Thus, it will make it possible to increase the efficiency of the processes of provision and reengineering of public services.

The proposed principle of grouping services can be used in the creation of software that will allow analysis of services in the process of their reengineering. This will simplify and speed up the processes of reengineering and creation of public services.

3. The intelligent analysis of information about the composition of services carried out as part of the study is the basis for the formation of recommendations for grouping services, in contrast to the world practice of forming complex services based on life events. Relationships between services are established based on the analysis of Jaccard and Sorensen similarity coefficient values. The choice of a high threshold value for the specified coefficients justifies the correctness and reliability of the obtained results.

Acknowledgments

The research reported in this paper was conducted within the framework of the International Technical Assistance Project «EU4DigitalUA: Interoperability, e-services, cyber security». We express our gratitude to the «EU4DigitalUA» project for providing the necessary data and assistance in conducting research.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study and the results reported in this paper.

Funding

The research was carried out with funds from the International Technical Assistance Project «EU4DigitalUA: Interoperability, e-services, cyber security».

Data availability

All data are available in the main text of the manuscript.

References

1. Pro administratyvni posluhy: Zakon Ukrainy vid 06.09.2012 No. 5203-VI. Verkhovna Rada Ukrainy. Available at: https://zakon.rada.gov.ua/go/5203-17
2. Pro osoblyvosti nadannia publichnykh (elektronnykh publichnykh) posluh: Zakon Ukrainy vid 15.07.2021 No. 1689-IX. Verkhovna Rada Ukrainy. Available at: https://zakon.rada.gov.ua/go/1689-20
3. Persaud, A., Persaud, P. (2013). Rethinking E-Government Adoption: A User-Centered Model. International Journal of Electronic Government Research, 9 (4), 56–74. doi: https://doi.org/10.4018/ijegr.2013100104
4. Gavrilenko, O., Zhurakovska, O., Kohan, A., Matviychuk, R., Piskun, A., Khavikova, Y., Khalus, O. (2022). The principle for forming a portfolio of public services based on the analysis of statistical information. Eastern-European Journal of Enterprise Technologies, 3 (3 (117)), 57–64. doi: https://doi.org/10.15587/1729-4061.2022.260136
5. Gavrilenko, O., Khomenko, O., Zhurakovska, O., Kohan, A., Piskun, A., Khalus, O. (2022). Application of association rules for formation of public (administrative) services portfolio. Advanced Information Systems, 6 (4), 63–68. doi: https://doi.org/10.20998/2522-9052.2022.4.09
6. Najafli, E. (2021). Understanding the concept of «digitalization» in the context of government. Law and Society, 5, 201–214. doi: https://doi.org/10.32842/2078-3736/2021.5.27
7. Zhosan, H. (2020). Development of digitalization in Ukraine. Economic Analysis, 30 (1), 44–52. doi: https://doi.org/10.35774/econa2020.01.02.044
8. Ivanovych, L. M. (2020). Digitalization in Ukraine: concept, problems, prospects and tasks of public administration. Naukovyi visnyk: Derzhavne upravlinnia, 2 (4), 202–213. doi: https://doi.org/10.32689/2618-0065-2020-2(4)-202-213
9. Digital Economy and Society Index (DESI) 2020. Available at: https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2020
10. Digital Economy and Society Index 2021: overall progress in digital transition but need for new EU-wide effort. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_5481
11. *Digitalisation in Europe 2021-2022: Evidence from the EIB Investment Survey. European Investment Bank*. Available at: https://www.eib.org/en/publications/digitalisation-in-europe-2021-2022.htm
12. How broadband, digitization and ICT regulation impact the global economy. Available at: https://www.itu.int/hub/publication/d-pref-ef-bdr-2020/
13. Tyshchenkova, I. O. (2017). Elektronni posluhy u diyalnosti publichnoi administratsiyi Ukrainy. Dnipro: DDUVS, 156. Available at: https://er.dduvs.in.ua/bitstream/123456789/935/1/Монография_Тищенкова%20I.O.pdf
14. EU4DigitalUA. Available at: https://eu4digitalua.eu/
15. Hid z derzhavnykh posluh. Available at: https://guide.diia.gov.ua/
16. Leskovec, J., Rajaraman, A., Ullman, J. D. (2014). Mining of Massive Datasets. Cambridge University Press. doi: https://doi.org/10.1017/cbo9781139924801
17. Magurran, A. (2013). Measuring Biological Diversity. Wiley. Available at: https://www.perlego.com/book/1006601/measuring-biological-diversity-pdf