

Mapping of Source and Target Data for Application to Machine Learning Driven Discovery of IS Usability Problems

Oksana Ņikiforova¹, Vitaly Zabiniako^{2*}, Jurijs Kornienko³, Madara Gasparoviča-Asīte⁴, Amanda Siliņa⁵

¹Riga Technical University, Riga, Latvia

²⁻⁴ABC software Ltd., Riga, Latvia

⁵SEB, Riga, Latvia

Abstract – Improving IS (Information System) end-user experience is one of the most important tasks in the analysis of end-users behaviour, evaluation and identification of its improvement potential. However, the application of Machine Learning methods for the UX (User Experience) usability and efficiency improvement is not widely researched. In the context of the usability analysis, the information about behaviour of end-users could be used as an input, while in the output data the focus should be made on non-trivial or difficult attention-grabbing events and scenarios. The goal of this paper is to identify which data potentially can serve as an input for Machine Learning methods (and accordingly graph theory, transformation methods, etc.), to define dependency between these data and desired output, which can help to apply Machine Learning / graph algorithms to user activity records.

Keywords – Data mapping, IS usability, Machine Learning, User Experience (UX).

I. INTRODUCTION

According to the System Theory [1], which serves as a formal foundation for IS development and maintenance, any system is a combination of two interrelated aspects – data (structure) and processes (behaviour), where an adequate organisation of both aspects is the basis of an effectively functioning system. In this article, the focus is on the behavioural aspect, whereas improving IS end-user experience is one of the most important tasks in the analysis of end-users behaviour, evaluation and identification of its improvement potential. This task can be hindered by insufficient user involvement, lack of user feedback, incomplete and changing user requirements, etc. These are some of the critical reasons, which could potentially reduce the number of IS users or business customers.

Moreover, according to a Forrester Research study, a well-designed UI (User Interface) could increase site benefits by up to 200 %, and better UX design could increase benefits by up to 400 % [2]. Related study states that it is possible to use EUEM (End-User Experience Management) to reduce

technology-related disruptions to employee productivity, track and quantify employee technology experience, and collect qualitative feedback so that I&O (Infrastructure and Operations) professionals can use this knowledge to their benefit, to understand the value they can expect from EUEM. In [3], authors state that “...nowadays, many enterprises consider UX an organisation-wide priority”. Authors of research [4] mention that “... the service life of software depends on user experience”.

Analysis of IT system user behaviour and improvement of business process efficiency is usually accomplished through the collection and compilation of information using a variety of qualitative and quantitative methods, including: interviews; business and process research; analysis of daily work content and user experience. The focus of this process is on a systematic approach to data collection and interpretation, where the main challenge is the structured and methodological choice of analysis and research tools. One of the advanced modern research directions is Machine Learning and the application of appropriate methods for solving various tasks. However, the application of Machine Learning methods for the UX analysis is not widely researched. Few existing recent surveys on potential abilities to analyse UX with automated tools show that there is no universal method based on existing data, which would not require time-consuming processing of existing information.

Nevertheless, the challenge of improving the usability of Information Systems is increasing, taking into account increasing complexity and the range of available services, especially during the impact of the COVID-19 pandemic, when online and remote solutions are becoming much more important and widely used.

Based on the System Theory [1], solution of any problem can be viewed as a function that processes certain input data and produces the expected output results. In the context of the discovering usability problems, the information about behaviour of end-users could be used as an input, while the output data should focus on non-trivial or difficult attention-

* Corresponding author's email: Vitalijs.Zabinako@abcsoftware.lv

grabbing events and scenarios in UX. Thus, the research question of this paper is defined as shown in Fig. 1.

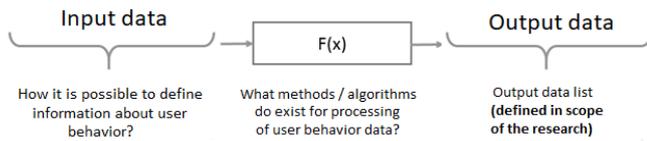


Fig. 1. Definition of the research question of this paper.

The paper is structured as follows. The next section outlines background and related work in the respective knowledge domain. The third section is concentrated on revealing suitable output data for UX analysis. The fourth section analyses potential sets of input data and states the list of source data useful for target data retrieval. The fifth section demonstrates an example of the application of Machine Learning methods for the pair of source/target data example as an experiment on potential usage of such an algorithm for UX data specifics. The sixth section specifies the mapping expected target data retrieval from the defined source data to ensure that the defined lists are complete for the application of Machine Learning methods. Finally, conclusions are made on the result of current research and potential future research areas.

II. BACKGROUND AND RELATED WORK

The [5] defines UX as “a person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service”. The source [6] elaborates this definition as “... every aspect of the user’s interaction with a product, service, or company that make up the user’s perceptions of the whole. User experience design (UCD) as a discipline is concerned with all the elements that together make up that interface, including layout, visual design, text, brand, sound and interaction. UX works to coordinate these elements to allow for the best possible interaction by users”. The source [7] additionally mentions that “UX is multidimensional and multi-faceted due to the many different types of experience, including social and empathic experiences, that users can live when using a product”.

Also, UX is closely related to UCD (User Experience Design) and Usability, which is defined by ISO/IEC 9241 as “the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”. The [8] states that “... the principles of the user centred design process are: Early focus on user tasks, empirical measurement, and iterative design”.

As for the usage of Machine Learnings for UX improvement, the study [2] states that “... we can use ML and its potential to help make good decisions in design... To better understand the perception in the field, we have focused on practitioners to identify promising directions for the application of ML to UXR... Combining the practices of UX and ML may yield positive effects in both directions... Our findings indicate that

using ML for the evaluation of a product’s UX may be a promising field for future research... Our survey indicated that the disciplines of ML and UX are expected to overlap and that UX practitioners see promising use cases of applying ML to UXR”. Here the User Experience Rating (UXR) is a metric that correlates IT end-user experience with application wait times. While [9] states that “... Machine Learning is now a fairly established technology, and User Experience designers have begun to integrate ML services into the things that they design”.

The most part of available information in public space describes the usage of “general AI/ML” for improving the user experience, for example:

- analysis and personalised offer of products, automatic generation of emails and search results (e.g., Google, YouTube, etc.);
- Natural Language Processing (NLP), chat-bots (e.g., Alexa, Google Home, etc.), voice recognition;
- IoT (Internet of Things), AR/VR (Augmented/Virtual Reality), analysis of facial expressions of users with neural networks and help of video capturing devices;
- Machine Learning – predictive analysis, data mining, e-learning and data modelling;
- image tagging, object identification, sentiment analysis, Optical Character Recognition (OCR) and person tracking.

Although these sources do mention the potential of using Machine Learning in UX improvement, realistically, there are no studies or results where the precise list of input data, formats of these, etc. would be defined together with the desired output data. Mapping of them does not define precise implementation algorithms; however, some studies exist, which propose, for example, a mathematical model for the innovation methodology, Design Thinking, using the Markov process [10]. Thus, no product/tools have been offered yet, which could be realistically applied to UX assessment and identification of problems, as well as finding potential solutions to the problems.

Therefore, the question remains as to what the input (or source) / output (or target) data could be applicable to Machine Learning methods and how these can be correlated with each other to analyse UX.

III. TARGET DATA DEFINITION BASED ON UX/CX EVALUATION HEURISTICS

In the introduction of the article and the task statement, the solution is defined as a function, which, by processing certain source data, reveals the required output data. In the context of improving UX, it is not obvious that we can expect the desired data as a result of the function. To identify the appropriate dataset, it is necessary to analyse the UX context, aspects and related factors.

In the scope of this research, the following UX aspects are of great interest [11]:

- usability – an indicator of the quality of the user experience;
- utility – the usefulness depends on the functionality and power of the back-end software that allows the user to get the job done. Products (or services) are useful if they

create benefits for the consumer. This is the real root cause of the product or system;

- emotional impact – it encompasses how users feel emotionally about interactions, including user satisfaction;
- significance – it arises from the user’s personal relationship with the product being used over a period of time. While usability and emotional impact relate to prime influence factors, the significance is that which occurs over time. These are when a product becomes an important attribute in the life of a particular user.

Moreover, the modern trends in customer experience are the following [12]:

- real-time insight in Customer Experience;
- journey automation and orchestration;
- personalised UI and content;
- user sentiment monitoring;
- AI powered chat-bots.

Excelling at CX – and maximising the revenue and efficiency benefits – requires a consistent, shared understanding of [13]:

- customers;
- their wants and needs;
- their perception of their interactions with your brand;
- the way to keep customers on the ideal journey – with their preferred channels.

The goal of the user experience engineer is to make the product on display user-oriented. This means that the product must be user-focused and the system must be as intuitive and user-friendly as possible. UX serves as a link between the goals of the host/product and the needs of the user. As a result of not following the user experience, the system in use may incorporate good technical features and be in line with the business strategy of the selector, but may be completely inefficient and “unpleasant” from the user’s point of view.

In user-oriented design, usability or usability analysis takes place with the involvement of end users. These user requirements are recognised and defined through methods such as focus groups, usability testing, card sorting, participatory design, questionnaires and interviews. The following areas are typically analysed to get a better idea of what end users want:

- person – at the beginning of the process, a person is created, who represents the target audience for the problem to be solved. A person symbolises a group of users who have similar needs, goals, skills, attitudes, and so on. A person ensures the right decisions with regard to product features, navigation, interactions, visual design and aspects;
- scenario – it is the “daily life” of a person. It is about the problems that a person has. Both emotional and physical details are important here;
- use case – a series of actions taken by a person to achieve a goal.

These factors suggest that target data classification can be expected under two types of information: 1) user characteristics (see section IV); and 2) user actions – this information needs additional research and usability heuristics – see below.

Nielsen in [14] defined 10 heuristics for designing user interface interactions, which still have not lost their relevance. They are the following [14]:

- 1) system status visibility;
- 2) the real-world relevance of the system;
- 3) user freedom and control of the system;
- 4) consistency and standards;
- 5) mistake-free;
- 6) recognition rather than cancellation;
- 7) flexibility and efficiency of use;
- 8) aesthetic and minimalist design;
- 9) assistance in identification, diagnosis, and recovery from errors;
- 10) help and documentation.

These are called “heuristics” because they are practical tips for developing usability. Heuristics are offered in the context of user-centred design, as they help develop a user-friendly interface that provides a pleasant experience when interacting with the system.

Other literature sources also concentrate on searching for and evaluating UX output heuristics. Authors in [15] state that “... in this article, we propose a novel AUX (Anticipated User eXperience) and UX evaluation method through which developers can obtain a knowledge base to create their designs... Intended Use: This criterion refers to the usage purpose of each user tool. For example, the way the user believes that a chat should work... Positive Anticipated Emotion: It concerns pleasant emotions that a user anticipates experiencing as a consequence of interacting with a user tool... ”.

In [7], “... the following facets are composed of the following UX properties: Economical facet: (i) Entertaining; (ii) Pleasantness; (iii) Productivity; (iv) Usefulness... ”.

“The literature already offers several UX evaluation aids, from simple questionnaires, like the User Experience Questionnaire – UEQ and the IPTV-UX (User Experience for Interactive Television) Questionnaire, up to more articulated and complete methods, like the Layered Emotions Measurement Tool and the Valence method... The research described in this paper suggests a UX evaluation method based on the exploitation of both emotions and mental models” [16].

“... Various methods have been developed for evaluating usability of interactive systems” [17].

All mentioned literature sources outline multiple factors related to UX, although these do not provide unified list (for example, without duplicates, contradictions, fuzzy, etc.) which could be used with precise AI/ML methods and algorithms. As a result, by analysing the mentioned information and authors’ experience working in IS development and support in the context of UX/UI analysis, the complete and consistent list of output metrics is proposed. The expected target data (listed in alphabetical order) are formal enough, do focus on the factors that can influence user actions, and also are appropriate for the use with AI/ML methods (refer to section V).

The list is the following:

- atypical action within the particular scenario;
- average time interval between actions;
- longest delay in particular action;
- longest path from source to target;
- not visited step (no longer used);
- percentage of actions used in the user session relative to number of all available actions in the IS;
- repeated action within the particular scenario;
- search content;
- shortest path from source to target;
- shortest time performing particular step (potentially redundant action);
- the most often visited step (action);
- the most popular route (sequence of actions);
- the most rarely visited step (action);
- typical behaviour scenarios for different user groups;
- typical behavioural scenarios to achieve the set goal;
- user emotions and sentiments;
- wrong transition from one step to another (with return to previous step).

IV. SOURCE DATA DEFINITION BASED ON ANALYSIS OF POTENTIALLY ACCESSIBLE INPUT INFORMATION ABOUT USER ACTIVITIES

The conclusion from the previous chapter is that we need statistical input data that contain information about the user's characteristics/attributes and his actions in IS (with timestamps and other appropriate metadata). Thus, it would be necessary to evaluate such possible sources that reflect the dynamics of user's behaviour. Examples of potential candidates widely used during IS analysis, development and maintenance are as follows.

1. State diagrams (not really suitable, as these are needed to be created separately, which may not be appropriate or even contradict the relevant IS documentation).
2. Scenarios / use cases / user stories (not suitable because these are designed before the system implementation and due to potential inappropriate support may not contain all the necessary information).

Therefore, it is necessary to look for some type of information that can be generated from the operation of IS in real time and environment and does not require additional efforts from IS support. IS audit records of user's actions can be used instead.

A real example of an audit trail that can be mapped to atomic (indivisible) data items usually contains a sequence of audited business actions (steps), such as:

- 1) login/logout;
- 2) entering main menu;
- 3) creating/editing/deleting records;
- 4) records of specifics to problem domain functional activities.

Considering the potential of AI/ML for data processing, it is possible to track, evaluate and use a sets of crucial UX metrics

(which help to develop a clear picture of user behaviours and interests).

One of such sets is defined in [18]:

- 1) device users are on when they visit a website;
- 2) location of users;
- 3) session time;
- 4) session length;
- 5) pages visited;
- 6) categories/products viewed;
- 7) bounce rates;
- 8) exit pages;
- 9) user flow.

One more source of the set of initial data potentially suitable for retrieving the set of data expected to analyse UX can be a set of data about IS user and his actions defined by *Google Analytics* – it can be used as potential content of audit records (for full list of appropriate data – refer to Section V).

In real IS, such audit data are often very general and presented in the form not suitable for direct analysis – either too much technical data are logged (such as each inner call of low-level web services), or almost no data are logged at all, for example, only logins and logouts. Therefore, the current auditing logic should be setup accordingly. Additionally, audit data are required to be extended with a high-quality meta-data set, which helps fully analyse user's activities. The source of such data may be even logging of user's actions in the Internet browser (but only by complying with the GDPR General Data Protection Regulation – by recording only the detection and timing of the execution of common operations, and not the detailed attributes and content of each operation that may contain sensitive user data). An example of such information retrieval could be data retrieved from the Internet browser, which presents a set of characteristics stored about user's session. Thus, an extended detailed logging of user activities is required.

Based on the generalization of data sets described above, as far as UX analysis focus is on dynamic states of user activities and their attributes in the context of time, the authors of this paper define the following list of attributes suitable for our task as source data (listed in alphabetical order):

- age;
- application name;
- “bounce rates”;
- client ID;
- device;
- document title;
- event action;
- event category;
- gender;
- geographical override;
- IP override;
- item name;
- pages visited;
- product category;
- screen resolution;
- session length;

- session time;
- social network;
- transaction ID;
- user ID;
- user language;
- user timing time.

Integration of the list of initial data potentially taken into consideration for retrieving of target data is based on combination of differently named, but similar by their nature attributes, e.g. Geographical override defined in Google Analytics is combined with Location of users defined in the set of crucial UX metrics, the same, Event Action is combined with User flow, Product Category is combined with categories/products viewed. Moreover, such attributes as Document location URL, Social action, Exit pages, etc., are eliminated from the final list of source data required due to being insufficient in the scope of target data identification.

V. POTENTIAL APPLICATION CASE OF AI/ML-BASED ANALYSIS OF USER BEHAVIOUR

In the previous sections, we have defined a set of source data, which can be processed to get the expected target data. We see this as a graph transformation problem, in which it is necessary to analyse the transition between vertices with a certain probability, regularity and duration. Suitable methods for solving such tasks are the methods of Machine Learning, which the authors of this article have already successfully applied in solving the cyber security problem [19].

In order to test the hypothesis of this article that the expected target data can be obtained from the defined source data, the pair of input data – “atypical action within the particular scenario” (as target data) and “pages visited” (as source data) – are taken for a detailed analysis. The general sequence of the

transition from the defined source data to the expected target data is shown in Fig. 2.

The historical source data of visited pages are received as a set of audit records from IS (A1). The set of audit data (A1) is then captured as an array of user “steps” (B1) via data adapter, which processes and stores audit data. The initial graph of Markov Chain is built (C1), which captures the number of individual transitions (“steps”) between each pair of input data elements. The initial Markov Chain (C1) is converted to its final form by assigning the appropriate probabilities to the defined transitions (C2), which is strictly mathematical process. This is the user’s behavioural profile, which can be used for further analysis of new user’s sessions, in order to detect target “atypical actions”. Each new user’s sequence of actions is obtained from the current audit records of actions (A2). The sequence of actions (B2) is compared with the Markov Chain of behavioural profile (C2) – step by step, in order to evaluate statistical metric of similarity (D1) (depicted with light grey colour in Fig. 2) or difference (D2) (depicted with dark grey colour) of these steps, based on probability analysis. If a majority of user’s actions/steps comply with his behavioural profile – his session is identified as “typical”, otherwise it is “atypical”. The session records are presented in a list of actions highlighted as “typical”/“atypical” (E). The results of ML data analysis is visualised in convenient form for comprehension – as GUI data with appropriate charts (G1/G2).

This case shows the example of AI/ML usage for converting initial source/input data to expected target/output data. Similar approach to data transformation can be applied to other pairs of source and target data, where these data pairs should be considered in the context of mapping of these data in different combinations. By applying the mapping of these data, it could be possible to detect and evaluate potential improvements of UX.

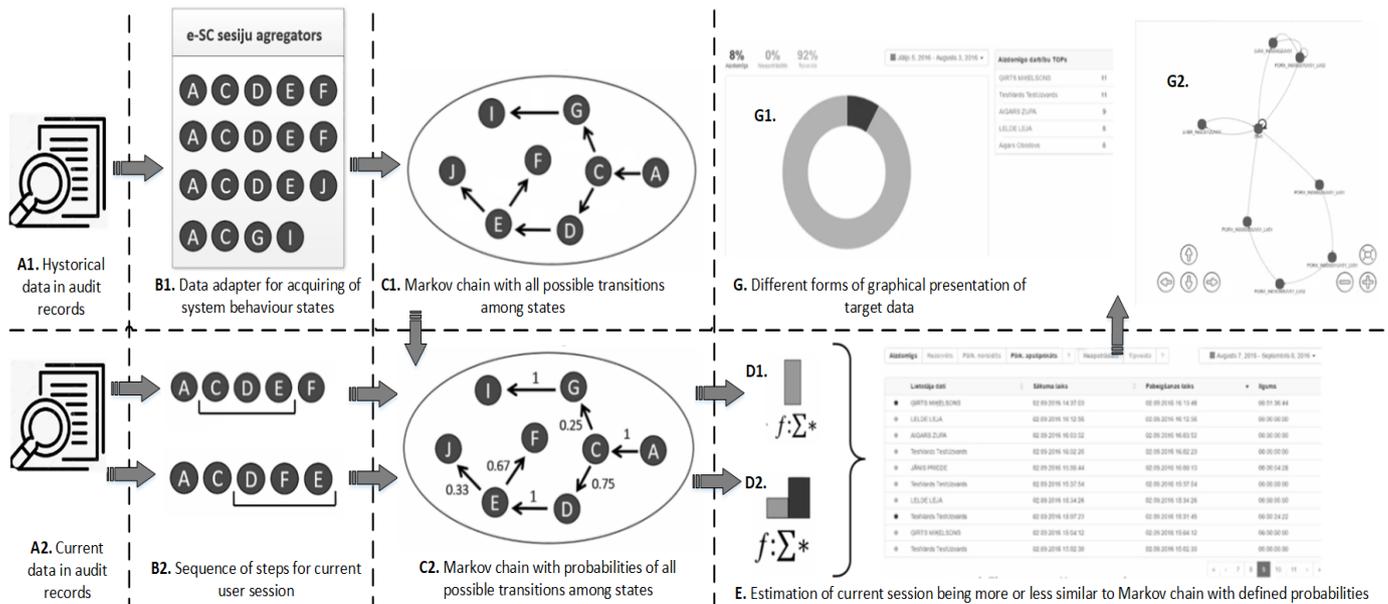


Fig. 2. The general sequence of the transition from the defined source data to the expected target data.

VI. MAPPING BETWEEN THE SOURCE AND TARGET DATA FOR UX ANALYSIS

As it is stated in the introduction, the proposed approach to UX analysis requires transformation of source data (which is specified in Section IV) into target data (see Section III). Section V demonstrates an ability to use Markov chains as one of the ML methods for such data transformation. This section is devoted to the analysis of the audit trail items – what data are recorded, and what combinations or transformations could be to gain potentially richer range of target data list.

For the UX analysis, attributes of user and session, which are defined in this article as an integrated list of audited business actions, a set of crucial UX metrics and attributes defined by Google analytics can be used as the source data. Thus, the full mapping table (Table I) shows an analysis of source data suitable to judge which of target data can be obtained.

Table I shows that some source data are used to obtain expected targets, some are not used. Moreover, some source data are used for identification of all targets.

Otherwise, some source data are not suitable as an input parameter to conclude about output data, but are important as a particular aspect for target specification. These aspects and such characteristics, like the context of UX, session identification and its attributes, which give an ability to identify target data focused on user activities, give an ability to classify the source in the order of their priorities and context of usage.

So far, the classified list of source data reordered by their types is the following.

1. Primary for all targets:
 - [S1] – Client ID
 - [S2] – User ID
 - [S3] – Pages visited
 - [S4] – Transaction ID
2. Primary for all targets specific to time:
 - [S5] – User timing time
 - [S6] – Session time
 - [S7] – Session length
3. Primary for all targets specific to action:
 - [S8] – Application name
 - [S9] – Item name
 - [S10] – Event action
 - [S11] – Event category
 - [S12] – Product category
 - [S13] – Document title
 - [S14] – Bounce rates
 - [S15] – Social network
4. Secondary for all targets:
 - [S16] – Device
 - [S17] – Screen resolution
 - [S18] – Geographical override
 - [S19] – IP override
 - [S20] – Age
 - [S21] – Gender
 - [S22] – User language

TABLE I
MAPPING OF SOURCE ITEMS DEFINED IN SECTION IV AND TARGET “MONITORING ITEMS” DEFINED IN SECTION III

Input/output	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17
S1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S3	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S5												x	x	x			
S6												x	x	x			
S7												x	x	x			
S8																x	x
S9						x	x	x	x	x	x						
S10						x	x	x	x	x	x						
S11						x	x	x	x	x	x						
S12						x	x	x	x	x	x						
S13						x	x	x	x	x	x						
S14										x	x						
S15	x	x	x	x	x							x	x	x			x
S16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S18	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S19	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S20	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S21	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
S22	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
																x(*)	x(*)

The target data can be classified to which context of source data is the most suitable for their retrieval, which, in turn, gives an ability to define target classification. The list of target data given in the order of their classification is the following.

1. Targets specific to step sequence/route length:
 - [T1] – Shortest path from source to target
 - [T2] – Longest path from source to target
 - [T3] – The most popular route (sequence of actions)
 - [T4] – Typical behaviour scenarios for different users
 - [T5] – Typical behavioural scenarios to achieve the goal
2. Targets specific to looking for one particular step (action):
 - [T6] – The most often visited step (action)
 - [T7] – The most rarely visited step (action)
 - [T8] – Not visited step (no longer used)

- [T9] – Repeated action within the particular scenario
 - [T10] – Wrong transition from one step to another (with return to previous step)
 - [T11] – Atypical action within the particular scenario
3. Targets specific to time spent:
 - [T12] – Shortest time performing the particular step (potentially redundant action)
 - [T13] – Longest delay in the particular action
 - [T14] – Average time interval between actions
 4. Other:
 - [T15] – Percentage of actions used in the user’s session relative to a number of all available actions in the IS
 - [T16] – Search content
 - [T17] – User emotions and sentiment

TABLE II

DEPENDENCY DEFINITION BETWEEN SOURCE AND TARGET DATA COMBINATIONS WITH COMMENTS FOR THEIR IMPLEMENTATION

	Target data	Set of corresponding source data	Comments on implementation
1.	Shortest path from source to target	Client ID, User ID, Pages visited, Transaction ID	Implementable by algorithms for graph search
2.	Longest path from source to target		
3.	The most popular route (sequence of actions)		
4.	Typical behaviour scenarios for different user groups	Device, Screen resolution, Geographical override, IP override, Age, Gender, User language	Implementable by Markov chains
5.	Typical behavioural scenarios to achieve the set goal	Social network	
6.	The most often visited step (action)	Client ID, User ID, Pages visited, Transaction ID	Implementable by algorithms for graph search
7.	The most rarely visited step (action)		
8.	Not visited step (no longer used)		
9.	Repeated action within the particular scenario	Device, Screen resolution, Geographical override, IP Override, Age, Gender, User language	Implementable by algorithms for graph search
		Item name, Event action, Event category, Product category, Document title	
10.	Wrong transition from one step to another (with return to previous step)	Client ID, User ID, Pages visited, Transaction ID	Implementable by Markov chains
		Device, Screen resolution, Geographical override, IP override, Age, Gender, User language	
11.	Atypical action within the particular scenario	Item name, Event action, Event category, Product category, Document title	Implementable by algorithms for graph search
		Bounce rates	
12.	Shortest time performing the particular step (potentially redundant action)	Client ID, User ID, Pages visited, Transaction ID	Implementable by algorithms for weighted graph
13.	Longest delay in the particular action	Device, Screen resolution, Geographical override, IP override, Age, Gender, User language	
14.	Average time interval between actions	User timing time, Session time, Session length	
15.	Percentage of actions used in the user’s session relative to number of all available actions in the IS	Client ID, User ID, Pages visited, Transaction ID	Implementable by algorithms for graph search
		Device, Screen resolution, Geographical override, IP override, Age, Gender, User language	
16.	Search content	Client ID, User ID, Pages visited, Transaction ID	(*) Additional source data required to identify the specific target
		Device, Screen resolution, Geographical override, IP override, Age, Gender, User language	
		Application Name	
17.	User emotions and sentiments	Client ID, User ID, Pages visited, Transaction ID	(*) Additional source data required to identify the specific target
		Device, Screen resolution, Geographical override, IP override, Age, Gender, User language	
		Application name	
		Social network	

By merging a similar set of source data and the same-typed target data for better visualisation, Table II shows dependence of similar/different source/target data pair combinations with comments on implementation.

Table II is structured in such a way that both target and source data are grouped according to the combinations of data required for implementation and potentially used implementation methods. It can be concluded that by excluding two “special” target cases (which were also classified as being within “Other” group, namely – “Search content” and “User emotions and sentiments”), it is possible to determine both the source data set and the transformation algorithm for the other expected targets.

Moreover, secondary data are not considered as an input parameter that allows deducing, for example, the shortest/longest path, but instead – as an additional attribute that allows deducing the determination of the shortest/longest path, for example, for people speaking different languages. It would be interesting to evaluate, for example, the “Wrong transition from one step to another” target from the statistical point of view in order to understand whether it is more common for men or women.

Another eye-catching aspect could be the most popular path or target for people of different ages/genders. One more case for analysis can be related to the attribute of geographical position in the context of the most/less frequently used action, etc.

However, care must be taken while interpreting the data, because specific uncommon cases might arise, for example, when grandson performs the necessary activities instead of his grandmother. This can negatively affect both the statistics and “noise” the data analysis.

VII. CONCLUSIONS AND FUTURE RESEARCH

This paper focuses on trending topics in the modern IT world and addresses the question on applying ML to the relationships between sources and targets of the UX. In the task of UX analysis, the focus is on user’s behaviour and its characteristics. The article identifies the expected results, which can provide a basis for UX experts to analyse user behaviour and compare the use of different system components to analyse and identify the differences and typical shortcomings in their development. For the processing and analysis of such information, the authors of the article have defined a list of input data – audit records can be used for this purpose.

Audit records have been selected based on the fact that these may already be available in the typical IS and do not require for additional information to be obtained or for the IS to be adjusted. This, in turn, reduces obstacles to the implementation of the proposed user behaviour analysis procedures. This study has examined whether all required target data can be based on a potentially available input data set.

Thereby, the research questions defined in the introduction (see Fig. 1) are answered with the results of the paper and can be expressed as shown in Fig. 3.

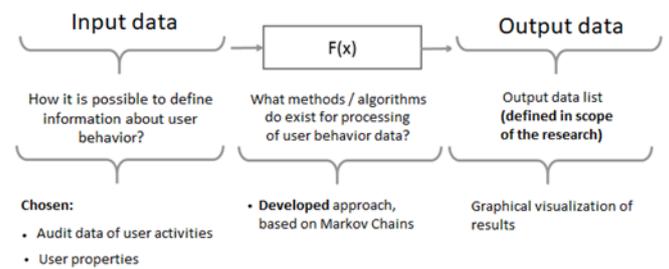


Fig. 3. Depiction of the result of this paper regarding original research question.

The main conclusions are the following.

1. It is possible to utilise AI/ML automatization benefits for analysis and potential improvement of UX of any existing IS, which is especially important in the state of COVID pandemic.
2. Our study of related works has allowed identifying potentially useful source/target data for this task and performing appropriate mapping, which can be used by UX experts in combination with AI/ML for guided well-grounded discovering of usability problems of each particular IS, based on the analysis of specific real use cases.
3. The proposed transformation algorithm, which is based on Markov Chains model, has proven that such mapping between source and target data is effectively possible. Further research is required to enhance the obtained results to ensure full data mapping support; this is planned to be done in the scope of next article, which will highlight results of our further research.

One of the potential areas of future research would be to practically define the transformation of source/target data combinations and to propose the forms of visual reporting of the analysis results, with the possibility for the user to modify the necessary reports by combining and comparing data.

VIII. ACKNOWLEDGMENT

The research leading to these results has received funding from the research project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/18/A/003 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 1.11 "An Analysis of IT Systems Users Behaviour and Improvement of Efficiency of Implemented Business Processes by Using AI/ML".

REFERENCES

- [1] L. v. Bertalanffy, *General System Theory*. New York: George Brazillier, 1969.
- [2] Spencer Kinney Digital Marketing Agency, "Why Smooth User Experience (UX) is Vital For Business" [Online]. Available: <https://www.spencerkinney.com/why-smooth-user-experience-ux-is-vital-for-business/> [Accessed: 18 Dec. 2020].
- [3] M. Chromik, F. Lachner, and A. Butz, "ML for UX? – An Inventory and Predictions on the Use of Machine Learning Techniques for UX Research", in *NordiCHI'20: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, Oct. 2020, Article No. 57 pp. 1–11. <https://doi.org/10.1145/3419249.3420163>

- [4] Q. Yu, X. Che, Y. Yang, and L. Wang, "A Transfer Learning Based Interpretable User Experience Model on Small Samples", in *2019 IEEE 19th International Conference on Software Quality, Reliability and Security*, Jul. 2019, pp. 186–196. <https://doi.org/10.1109/QRS.2019.00035>
- [5] J. Oh and U. Lee, "Exploring UX Issues in Quantified Self Technologies", in *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, Jan. 2015, pp. 53–59. <https://doi.org/10.1109/ICMU.2015.7061028>
- [6] C. Y. Wong, K. Chu, and M. Pauzi, "Advocating UX Practice in Industry: Lessons Learnt from UX Innovate Bootcamp", in *2016 4th International Conference on User Science and Engineering (i-USER)*, Aug. 2016, pp. 204–209. <https://www.doi.org/10.1109/IUSER.2016.7857961>
- [7] M. Pallot, K. Pawar, P. Krawczyk, M. Topolewski, A. Lecossier, and A. R. Abdel Razek, "Evaluating User eXperience as a Means to Reveal the Potential Adoption of Innovative Ideas", in *2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, June 2020. <https://www.doi.org/10.1109/ICE/ITMC49519.2020.9198368>
- [8] H. Degen, G. Guillen, and H. Schmidt, "Effectiveness and Cost-Benefit Ratio of Weekly User Group Sessions", in *Design, User Experience, and Usability. Design Philosophy and Theory*, A. Marcus, W. Wang, Eds. Springer Nature Switzerland AG, 2019, HCII 2019, LNCS 11583, pp. 208–221. https://doi.org/10.1007/978-3-030-23570-3_16
- [9] G. Dove, K. Halskov, J. Forlizzi, and J. Zimmerman, "UX Design Innovation: Challenges for Working with Machine Learning as a Design Material", in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, May 2017, pp. 278–288. <https://doi.org/10.1145/3025453.3025739>
- [10] L. J. Lopez, D. H. Garcia, M. Talancon, M. A. Camacho, A. F. Valentin, "Mathematical Modeling for Design Thinking Innovation Method Based on Markov Chain Theory", *NISCAIR-CSIR*, pp. 135–138, 2019.
- [11] R. Hartson, and P. Pyla, *The UX Book*. 2019, 916 p.
- [12] S. Khan, and M. Iqbal, "AI-Powered Customer Service: Does it Optimize Customer Experience?", in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, Jun. 2020. <https://www.doi.org/10.1109/ICRITO48877.2020.9198004>
- [13] N. Behare, S. Waghulkar, and S. Shah, "A Theoretical Perspective on Customer Experience (CX) in Digital Business Strategy", in *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*, Aug. 2018. <https://www.doi.org/10.1109/RICE.2018.8509079>
- [14] J. Nielsen, "Enhancing the Explanatory Power of Usability Heuristics," in *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1994, pp. 152–158. <https://doi.org/10.1145/191666.191729>
- [15] L. Sanchez-Adame, S. Mendoza, B. A. Gonzalez-Beltran, J. Rodriguez, and A. Meneses Viveros, "AUX and UX Evaluation of User Tools in Social Networks", in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, Dec. 2018, pp. 104–111. <https://www.doi.org/10.1109/WI.2018.0-101>
- [16] S. Filippi and D. Barattin, "User Experience (UX) Evaluation Based on Interaction-Related Mental Models", *Advances in Usability and User Experience*, pp. 634–645, June 2018. https://www.doi.org/10.1007/978-3-319-60492-3_60
- [17] P. Buono, D. Caivano, M. F. Costabile, G. Desolda, and R. Lanzillotti, "Towards the Detection of UX Smells: The Support of Visualizations", *IEEE Access*, vol. 8, pp. 6901–6914, Dec. 2019. <https://www.doi.org/10.1109/ACCESS.2019.2961768>
- [18] C. Haughey, "How to Improve UX with AI and Machine Learning" [Online], July 2019. Available: <https://www.springboard.com/blog/improve-ux-with-ai-machine-learning/> [Accessed: 14 Dec. 2020].
- [19] P. Osipov, L. Aleksejeva, A. Borisov, Y. Chizhov, T. Zmanovska, and V. Zabiniako, "Implementation and Operation Aspects of a System for Detecting Abnormally Level of User Activity", *Automatic Control and Computer Sciences*, vol. 51, no. 6, pp. 417–425, Nov. 2017. <https://doi.org/10.3103/S0146411617060050>



Oksana Nīkiforova received the Doctoral degree in Information Technologies (system analysis, modelling and design) from Riga Technical University, Latvia, in 2001. She is presently a Professor at the Department of Applied Computer Science, Riga Technical University. Her current research interests include model driven software development and agile software development project management methods and tools.
E-mail: oksana.nikiforova@rtu.lv
ORCID ID: <https://orcid.org/0000-0001-7983-3088>



Vitalijs Zabiniako received the Doctoral degree in Information Technologies (system analysis, modelling and design) from Riga Technical University, Latvia, in 2012. He is presently a Lead Researcher at ABC software Ltd. His current research interests include data visualisation, cyber security and graph theory.
E-mail: vitalijs.zabinako@abcsoftware.lv



Jurijs Kornienko received the Doctoral degree in Information Technologies (system analysis, modelling and design) from Riga Technical University, Latvia, in 2007. He is presently a Deputy Chairman and Head of Microsoft Technology Department / Project Manager at ABC software Ltd. His current research interests include cyber security, data mining and acquisition.
E-mail: jurijs.kornijenko@abcsoftware.lv



Madara Gasparoviča-Asīte received the Doctoral degree in Information Technologies (system analysis, modelling and design) from Riga Technical University, Latvia, in 2016. She is presently a User Support Specialist at ABC software Ltd. Her current research interests include decision support systems, data mining tasks and modular rules.
E-mail: madara.gasparovica-asite@abcsoftware.lv



Amanda Siliņa received the Bachelor degree in financial engineering from Riga Technical University, Latvia, in 2020. She is presently digital analyst at SEB and master's student in financial engineering mathematics from Riga Technical University. Her current research interests include data interpretation and model based decisions.
E-mail: amanda.silina@seb.lv