Ministry of Education and Science of Ukraine
National Aviation University
Software Engineering Department

International Conference on Software Engineering

April 12 – 14, 2021

Proceedings

Kyiv 2021

Conference organizers and partners

	National Aviation University (Kyiv, Ukraine) www.nau.edu.ua
	Software engineering department www.sed.nau.edu.ua
	Artificial Intelligence and Knowledge Engineering Research Lab (Cairo, Egypt): http://aiasulab.000webhostapp.com/Ambassadors.html
	The Association of Developers and Users of Intelligent Systems http://www.aduis.com.ua/
	Faculty of Computer and Information Sciences, Ain Shams University (Cairo, Egypt): http://cis.asu.edu.eg/
ITHEA®	The Institute for Information Theories and Applications FOI ITHEA® (Bulgariya) http://www.ithea.org/index.html
ITHEA® International Scientific Society	International Scientific Society ITHEA® ISS http://www.ithea.org/iss/iss.html
	V. M. Glushkov Institute of Cybernetics of the NAS of Ukraine (Kyiv, Ukraine) http://incyb.kiev.ua/?lang=en

NATIONAL AVIATION UNIVERSITY

(www.nau.edu.ua)

For almost 80 years of its history more than 200,000 highly skilled professionals have been trained in this higher aviation educational institution. Among them there are well-known scientists, heads of aviation companies, enterprises, organizations and institutions that provide aircraft flights, their maintenance and repair, transportation of passengers and cargo.

Strong scientific and pedagogical schools enable to train not only engineers but also economists, lawyers, environmentalists, translators, psychologists, sociologists and others.

SOFTWARE ENGINEERING DEPARTMENT

www.sed.nau.edu.ua

Software engineering department studies more than 900 students. Students obtain fundamental knowledge and practical skills in software development area as well as theoretical basic in software development life-cycle management.

AIN SHAMS UNIVERSITY AI AND KNOWLEDGE ENGINEERING LAB

http://aiasulab.000webhostapp.com/

The Ain Shams University Artificial Intelligence and Knowledge Engineering Lab was founded in 2005 and belongs to the Computer Science department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.

The Lab is currently composed of 3 full Profs., 3 Assistant Prof., 15 Drs (PhD holders) 5 of them from Arabic countries, 10 PhD students and 5 Msc students. Laboratory is working in the following disciplines:

- Intelligent Information Processing, Data Mining.
- Knowledge Engineering and Expert Systems.
- Image Processing and Pattern Recognition.
- Data Science and Big Data Analytics
- Intelligent Education and Smart Learning.
- Medical Informatics, Bio-informatics and e-Health.
- Computing Intelligence and Machine learning
- Internet of Things (IoT),
- Technologies for HealthCare;
- Space Science and Cosmic Rays Big Data Analytics.

In addition, the lab includes 17 scientific international ambassadors from Germany, Canada, Italy, Georgy, Ghana, Cyprus, Ucraine, Romania, Greece, Japan, Mexico, and Belgium

THE ASSOCIATION OF DEVELOPERS AND USERS OF INTELLIGENT SYSTEMS

http://www.aduis.com.ua/

Association of Developers and Users of Intelligent Systems (ADUIS) consists of about one hundred members including ten collective members. The Association was founded in Ukraine in 1992. The main aim of **ADUIS** is to contribute to the development and application of the artificial intelligence methods and techniques.

Association has long-term experience in collaboration with teams, working in different fields of research and development. Methods and programs created in Association were used for revealing regularities, which characterize chemical compounds and materials with desired properties.

FACULTY OF COMPUTER AND INFORMATION SCIENCES, AIN SHAMS UNIVERSITY (Cairo, Egypt):

http://www.shams.edu.eg/598/news#

Faculty of computer and information sciences was founded in 1996 and consists of 5 departments, namely; Computer Science, Information Systems, Computers Systems, Scientific Computing, and Basic Sciences.

Faculty of computer and information sciences announces it's distinct programs by credit hours system:

- Software Engineering;
- Artificial Intelligence:
- Cyber Security;
- Digital multimedia.

Faculty of computer and information sciences also provides Bioinformatics program which accepts Thanawya Amma science section

INSTITUTE FOR INFORMATION THEORIES AND APPLICATIONS FOI ITHEA® (BULGARIA)

http://www.ithea.org/

The Institute for Information Theories and Applications FOI ITHEA® is an international nongovernmental organization functioning since 2002 year. ITHEA® is aimed to support international scientific research through international scientific projects, workshops, conferences, journals, book series, etc.

INTERNATIONAL SCIENTIFIC SOCIETY

ITHEA® ISS (INTERNATIONAL SCIENTIFIC SOCIETY)

http://www.ithea.org/iss/iss.html

ITHEA® International Scientific Society (**ITHEA® ISS**) is aimed to support growing collaboration between scientists from all over the world.

Till now, the **ITHEA® International Scientific Society** has been joined by more than **4000** members from **53** countries from all over the world:

ITHEA® Publishing House (**ITHEA® PH**) is the official publisher of the works of the ITHEA®ISS. The scope of the issues of the ITHEA® ISS covers the area of Informatics and Computer Science. ITHEA® PH welcomes scientific papers and books connected with any information theory or its application.

ITHEA® ISS has four International Journals, established as independent scientific printed and electronic media, published by ITHEA® Publishing House:

- International Journal "Information Theories and Applications" (IJ ITA), since 1993;
- International Journal "Information Technologies and Knowledge" (IJ ITK), since 2007;
- International Journal "Information Models and Analyses" (IJ IMA), since 2012;
- International Journal "Information Content and Processing" (IJ ICP), since 2014.

Every year, ITHEA® ISS organizes many scientific events called **"ITHEA® ISS Joint International Events of Informatics" (ITA**). ITHEA® ISS has organized 135 International Conferences and workshops and has published 45 scientific books: monographs, thematic collections and proceedings.

All publications of ITHEA® ISS are accessible freely via ITHEA® main site www.ithea.org as well as via ITHEA® Digital Library (IDR) http://idr.ithea.org.

The great success of ITHEA® International Journals, Books, and Conferences belongs to the whole of the **ITHEA® International Scientific Society.** Membership in ITHEA® ISS is free.

V.M.Glushkov Institute of Cybernetics of the NAS of Ukraine

http://incyb.kiev.ua/?lang=en

M.Glushkov Institute of Cybernetics of the NAS of Ukraine is a Scientific Centre widely known in Ukraine and beyond it engaged in the solution of fundamental and applied issues of informatics and computer science as well as introduction of their methods and means to various spheres of human activities.

Today the main areas of the research of the Institute are:

Development of the general theory and methods of system analysis, mathematical modelling, optimization and artificial intelligence;

Development of the general theory of management, methods and means of arrangement of intelligent control systems of different levels and purposes

Creation of the general theory of computing machines and development of perspective computing equipment, artificial intelligence and informatics;

Creation of perspective systems of mathematical support of general and applied purposes;

Development of new information technologies and intelligent systems; and

Solving of fundamental and applied problems of informatization of the society.

Scientific schools of computer mathematics and discrete optimization, mathematical theory of computing systems and artificial intelligence, system analysis and stochastic programming, mathematical theory of reliability and programming theory were forming for fifty years at the V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine and they gained world's recognition.

Conference topics

Theoretical Basics of Software Engineering

AGILE and Model-Driven Approaches to Software Development.

Approaches to Software Development Life Cycle Processes Improvement;

Business Process Management and Engineering

Empirical Software Engineering;

Formal Foundations of Software Engineering;

Frameworks and Middleware;

Model-Driven Engineering;

Software Development Technologies;

Software Engineering Standards;

Software Patterns and Refactoring;

Applied Aspects of Software Engineering

Approaches and Models to Estimate Expenses of Future Software Project;

Component-Based Software Engineering;

Requirements Engineering;

Service-Oriented Software Engineering;

Software Designing;

Software Quality;

Software Maintains and Evolution;

Software Testing.

Software Engineering Application Domains

Agent and Multi-Agent Systems;

Biotechnologies and Smart Health Technologies;

Cloud Computing;

Data Bases and Knowledge Bases;

Image Processing and Computer Vision;

Internet;

Security;

Software Development for Mobile Operation Systems.

Real-Time Systems;

Network and Data Communications;

User-Centered Software Engineering.

Steering committee

Abdel-Badeeh M. Salem (Ain Shams University, Egypt);

Albert Voronin (National Aviation University, Ukraine);

Krassimir Markov (Institute of Information Theories and Applications, Bulgaria);

Olena Chebanyuk (National Aviation University, Ukraine);

Serhii Zybin (National Aviation University, Ukraine);

Vitalii Velychko (V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine).

Program committee

Antinisca Di Marco (University of L'Aquila, Italy);

Besik Dundua (Tbilisi State University, Georgia);

Eugene Machusky (National Technical University of Ukraine Igor Sikorsky Kyiv Polytechnic Institute, Ukraine);

George Totkov (University of Plovdiv, Bulgaria);

Krassimira Ivanova (University of Telecommunications and Post, Bulgaria);

Michael Gr. Voskoglou (Graduate Technological Educational Institute of Western Greece School of Technological Applications, Greece);

Mohamed Ismail Roushdy (Future University, Egypt);

Mykola Guzii (National Aviation University, Ukraine);

Nagwa Badr (Ain Shams University, Egypt);

Sedat Akleylek (Ondokuz Mayıs University, Turkey);

Vladimir Talalaev (National Aviation University, Ukraine).

Organization committee

Mykyta Krainii (National Aviation University, Ukraine);

Tetiana Konrad (National Aviation University, Ukraine);

Yana Bielozorova (National Aviation University, Ukraine);

Yuliia Bezkorovaina (National Aviation University, Ukraine).

Table of contents

SECTION: THEORETICAL BASICS OF SOFTWARE ENGINEERING

CONCEPT OF DATA SETS REUSE IN SOFTWARE PRODUCT LINE APPROACH

Olena Chebanyuk

Abstract*: Approach of data reuse to estimate reliability of software components before adding them to repository is proposed in this this paper.*

Keywords*: Software Product Line, Repository of Software Components, Systematic Reuse.*

Introduction and State of Art in Reuse Problem

Software Product Line is perspective and modern approach focused on systematic software development artifacts reuse. Modern tendencies and approaches are focused mostly on reuse of software modules and design solutions.

Following this tendency one of the perspective and practical aspects of Software Product Line approach is designing of software components repositories. From the theoretical point of view, such repositories simplify the procedure of software development, namely such activities as software construction, designing and development (IBM, 2016).

When new software module of project is added to repository of software components – the actual question about preliminary software quality estimation becomes an actual. Importance of this question is explained by the next factors: IT company, then customer, then software user must be sure that functionality is realized correctly, and hidden mistakes of execution of software bugs will not fail the correct execution of software.

Modern approaches are focused on new testing techniques usage (Chebanyuk and Palahin, 2016), for example Test-Driven Development (TDD). Traditionally TDD is involved in forward engineering, when data sets (test cases and test suites) are prepared before realizing project features. Preparing data and expectation of needed results in

forward engineering allows increasing the speed and accuracy of the process selecting from variety of software modules those, which satisfies to the requirements of the new software project in the best way.

Contribution of this thesis: this thesis proposes general idea of software modules quality estimation approach that considers data sets reuse. Data sets –are test cases and test suites that are accumulated from previous projects in software product line approach.

Task to propose a concept of data sets reuse in software product line approach for estimating software components quality.

The essence of the proposed approach is represented in the figure 1.

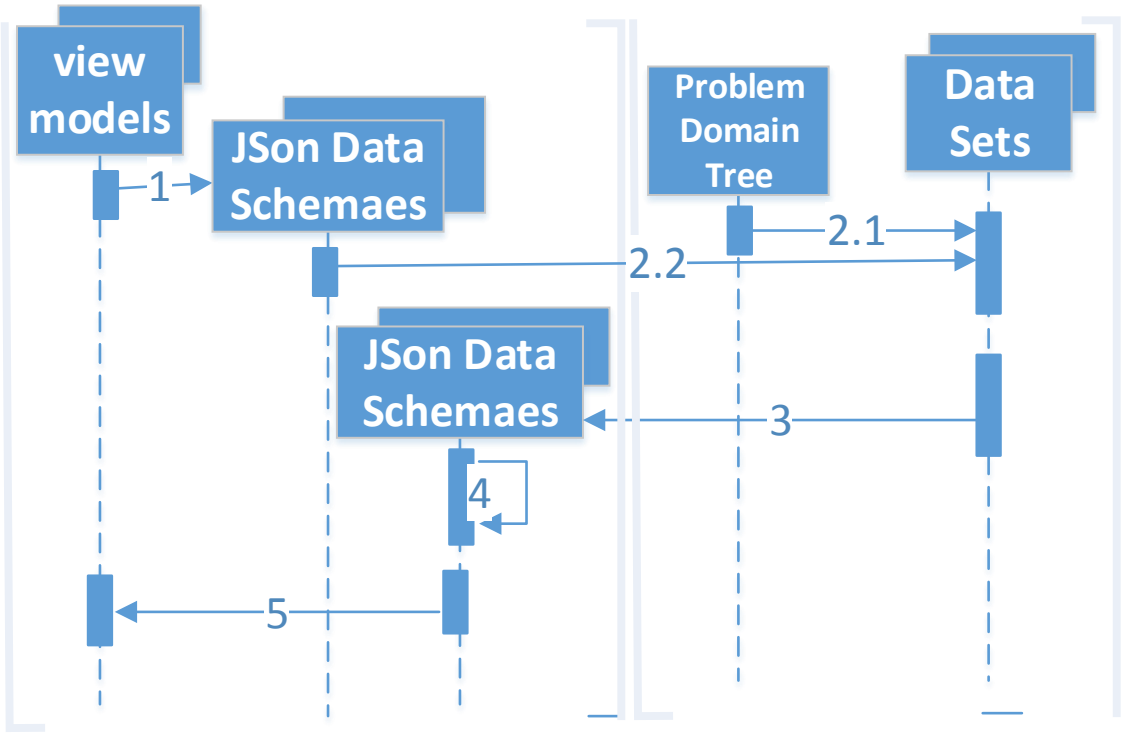

Figure 1. General scheme of the data sets reuse in product line approach

Reuse concept is based on the following actions:

1. Problem domain tree is designed (Chebanyuk and Palahin, 2019). Data sets that are represented as test suites are gathered from the previous projects, performed in software company. They are matched to problem domain tree processes.

2. View models of MVC project are investigated (Chebanyuk and Palahin, 2018). Then JSON (or XML) schemas, representing data structures, namely project Data Models are extracted (message 1 in the figure 1). Good example of data structures processing representing is proposed in book (Seriy at el,, 2016).
3. If project topic matches to problem domain processes (message 2.1 in the figure 1) request to Software Product Line repository of data sets is performed (message 2.2. in figure 1).
4. The closest data schemas to existing View Models are returned (message 3 in figure 1)
5. Then data for automatic testing are refined (Message 4 in figure 1). Consider possible example of such a refinements: in View Model five data fields are represented. The closest data scheme contains four fields that matched to View Model. Request can fill extra data field by default values (for example “null” or zero) or perform a search in other data schemes field that is match to View Model. Then missing field is merged with existing data model.
6. Automatic test are executed (Chebanyuk and Palagin, 2016). Then the request to extort test results to components repository is performed. Performing similar procedures (messages 3 and 4 in figure 1) testing results are extracted from the repository and the further comparison operation of expected and obtained results is performed.

Bibliography

1. Chebanyuk O. & Palahin O. (2016) Model-Driven Test Approach in Agile Based on User-Stories Analysis. *International journal “Informational Technologies and Knowledge”*, *Vol 10, number 4, 2016,* 303-316
2. Chebanyuk O. & Palahin O. (2018) A multi-layer approach of view models designing. *International journal “Informational Content and Processing”, 2018*, 203-216.
3. Chebanyuk O. & Palahin O. (2019) Domain Analysis Approach. *International journal “Informational Content and Processing”. Volume 6, Number 2, 2019*, 3-20.

4. Alex Seriy, Bhargav Perepa, Christian E. Loza, Christopher P. Tchoukaleff, Gang Chen, Ilene Seelemann, Kurtulus Yildirim, Rahul Gupta, Soad Hamdy, Vasfi Gucer (2016) Getting Started with IBM API Connect: Scenarios Guide, 1-150.

Authors' Information

Olena Chebanyuk *– D.Sc., Professor of Software Engineering Department, National Aviation University, Kyiv, Ukraine, e-mail: chebanyuk.elena@ithea.org*

Major Fields of Scientific Research: *Model-Driven Architecture, Model-Driven Development, Software architecture, Mobile development, Software development*

NORMALIZATION OF LIFE CYCLE MODELS IN TASKS PLANNING SOFTWARE PROJECTS

Vladimir Talalaev, Larysa Postavna

***Abstract**: An approach to solving the problem of choosing a life cycle model for software project management and a step-by-step algorithm for its implementation are proposed. The approach is based on the application of the procedure of normalization of life cycle models using a structural template, as well as the implementation of the procedure of expert-analytical comparison of normalized models with the profiles of the context of the software project.*

***Keywords**: Software Engineering, Life Cycle Models, Software Project.*

Introduction

The success of any software development project can be assessed within a set of metrics agreed with key stakeholders, among which the three-dimensional metric "quality - cost - time" occupies a dominant position. Also known as the "iron triangle" problem, this metric allows you to formulate a general criterion for the effectiveness and success of the project. A brief definition of this criterion can be summarized as follows: “A software project can be considered successful if such static and dynamic project parameters are selected and implemented that improvement of any parameter of the “iron triangle” sides cannot be achieved without losing the balance of interests of any of them key "stakeholders of the project". The implementation of this approach involves the presence in the arsenal of project participants at least a complete set of project characteristics, in which the processes of local optimization of the multicriteria task of selecting the best by the given criterion give the desired effect in achieving global optimum. Among this set of project characteristics, an important role is given to the choice of life cycle model. This is one of the key decisions determines the optimal choice of structural and functional parameters and characteristics of all processes of the project

life cycle, and, consequently, significantly affects the achievement of the optimal global criterion of

The following solution to the problem of choosing a life cycle model for a software project is based on the concept of "life cycle model" in the narrow sense, when the main processes of the model are only the processes of the software product development stage. This approach is due to the widespread use in software engineering practice of descriptions of typical life cycle models, which are limited to the stage of development. Of course, this approach narrows the scope of the results, but usually leads to more adapted to the practical application of research results.

In the problem under consideration, the life cycle model is defined as a description of a pattern (template) that summarizes a particular practice of software development and reproduces the key properties of life cycle processes in structural-functional, temporal, spatial and behavioral projections. Since the basis of software projects are the processes of software development (design, construction, implementation of testing, etc.), the correct choice of life cycle model is crucial in the planning of software projects.

Analysis of information sources on software project planning, in most best practices, the decision to choose a life cycle model is chosen by project team management based on their experience and intuition by analytically comparing the context of software development with information on the advantages and disadvantages of a known life model cycle. Along with the subjectivity of this choice, there is a significant dependence on the completeness of the knowledge base for best design practices, as well as a significant dimension of the problem of analytical comparison of project context profiles with descriptions of life cycle model options. The synergistic effect of these negative factors significantly impairs the effectiveness of design decisions and increases their risk. Reducing the risk of life cycle model selection procedures and reducing the undesirable consequences of their impact is possible through their formalization and further automation within the use of software CASE-design tools.

In the proposed approach, a step-by-step selection algorithm is implemented, the main operating blocks of which are:

1. Formation of a template for normalization of the description of models of a life cycle of software products.

2. Collection of information on best practices in the use of life cycle models in the practice of software development and systems.

3. Normalization of life cycle models from the knowledge base in accordance with the created template and the formation of a database of models.

4. Development of a structural template to describe the context profile of the software project.

5. Implementation of expert-analytical procedure for comparing the profile of the software project with normalized life cycle models. Formation of a set of dominant variants of life cycle models for selection of the best.

6. Application of the expert procedure of the final choice of the life cycle model and its application for planning the working structure of the project.

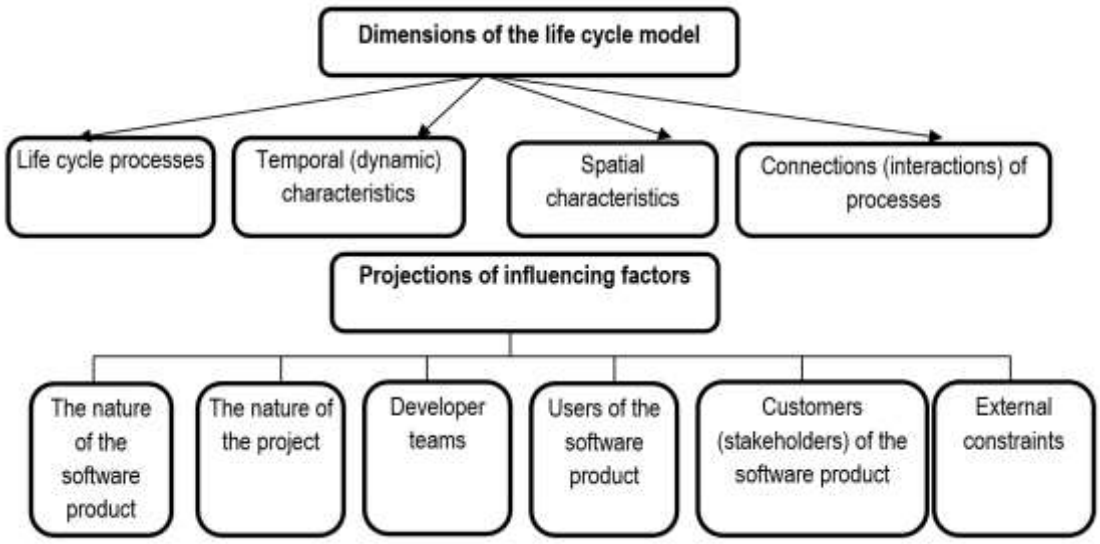

Fig. 1 Templates for normalizing life cycle models and building context profiles of software projects

Figure 1 shows the structure of life cycle model normalization templates and the context profile structure, according to which more than 20 descriptions of typical life cycle models were processed and a database for software projects was created.

Conclusion

The application of the proposed approach, as well as its implementation in the form of an appropriate step-by-step selection algorithm can significantly improve the process of planning software projects, reduce the risks of their implementation and create a strong knowledge base for effective software project management.

Bibliography

1. Gruber T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing // International Journal Human-Computer Studies. –1995. – Vol. 43(5–6). – P. 907–928.
2. Shamova TI, Davydenko TM, Rogacheva NA Adaptive school: problems and prospects. - Arkhangelsk - M., 1995. - 160 p.

Authors' Information

Vladimir Talalaev *– Assoc. Prof. of Software Engineering Department, National Aviation University, Kyiv, Ukraine.* ***E-mail:*** *VATalalaev@gmail.com)*

Major Fields of Scientific Research*: Model-Driven Architecture, Software architecture, Software development, Software risks, business process modeling*

Larysa Postavna *– assistant of Software Engineering Department, National Aviation University, Kyiv, Ukraine.* ***E-mail:*** *larysa.postavna@npp.nau.edu.ua (postavnalarisa@gmail.com)*

AUTOMATIC ONTOLOGY CREATION FROM NATURAL LANGUAGE TEXT

Anna Litvin, Vitalii Velychko, Vladislav Kaverinsky

***Abstract**: An approach and its software implementation was created to build an OWL ontology using as an input a natural language text or set of texts. The approach is primarily aimed but not limited to inflected languages. The method for the moment is acceptable for tagged or regularly structured texts. The created ontologies are tested as valid using Protégé and RDFlib and are useful for dialog and reference systems.*

***Keywords**: NLP, NLU, ontology, automatic ontology creation, OWL.*

***ITHEA Keywords**: H.3.4 Systems and Software.*

***Conference and topic**: Applied Aspects of Software Engineering: Software Designing.*

Introduction

Manually database creation and fulfillment could be a labor-intensive and rather boring process. Thus its automation is highly desirable. An information source for this purpose could be a natural language text that contains some information. Considered here database type is a graph database that is useful for ontologies – data structures that represent entities of some subject area and relationships between them. One of the popular languages for ontologies construction is OWL [Antoniou, 2016], which was used in this work. An OWL ontology can be easily converted to an RDF-triples set for which SPARQL queries can be used for information obtaining.

Some methods for ontologies creation automation have been proposed, for instance, in works [Young, 2018; Elnagar, 2020; Balakrishna, 2008]. But the method and the goal of a certain approach are highly dependent on the destination of the ontology and its usage, which affect the structure of the ontology. Thus the task of new systems building for ontologies automatic creation aimed at the certain case to construct the most useful and suitable in structure and content graph database seems to be actual and important.

The considered ontology structure

In the proposed approach all the entities in the ontology are devised in two main groups – classes and properties. The relationship types are limited to the following: “Subclass” – links more certain entity to more abstract, “Subproperty” - links more certain property to more abstract one, “Domain” – links to a property an entity (or set of entities) which has the property, “Range” - links to a property an entity (or set of entities) which is the value of the property. In this way two entities (or sets of entities) can be linked with a property where one entity is its “Domain” and the other is “Range”. So “Domain/Range” linking could be used for a horizontal not hierarchical binding of entities. These relationship types are standard for OWL and are rather enough for most cases.

In the proposed ontology structure the main abstract predefined classes are: “Action”, “Adjective”, “Adverb”, “Number”, “Object” (nouns and name groups), “Marker” (question and negotiation words), “ReadyAnswerLink” (links to large text blocks and media information), “Count” (denoting the number of possible occurrence multiplicity of a term in the text), “Quantity” (the actual multiplicity of occurrence of a term in the text or considered part of the text), “ConditionIntersection” (combinations or intersections of concepts from a potential user phrase that determines a certain response of the system). The predefined main properties in the considered ontology structure are following: “MainProperty” – its descendent's link combinations of concepts to descendants of the “ConditionIntersection” class; “BindQuantityToInteraction” - binds reference to text blocks - descendants of the class “ReadyAnswerLink” to the multiplicity of occurrence in their text of certain concepts – descendants of the class “Quantity”; “IntersectionCount” - binds the actual multiplicity of occurrence of the concept in the text block – descendants of the class “Quantity” to indicate the number of the multiplicity of occurrence of the concept in the certain text block – descendants of the class “Count”.

Subclasses of the “Object” class can have their descendants, which are name groups. The hierarchy of descendants of the “Object” class is as follows: the highest are the individual nouns that are the main in name groups and define the most general variant of

the concept then the hierarchy is followed by name groups as words grow in them, which define more specific entities. The ontology structure could be supplemented with the other concepts specific to the considered subject area, for example: “Date”, “Person”, “Location”, etc. These ones can also have predefined descendants, for instance: “Author” – a descendant of “Person”, “CreationDate” – a descendant of “Date”, etc. The predefined ontology entities and structures form the framework basing on which further automatic graph creation becomes possible and moreover the methodology of the following usage of the ontology becomes more clear, because the basic structure of the SPARQL queries for information obtaining becomes more predictable.

The ontology creation technique

It is assumed that the input text is tagged or regularly structured, which makes it easier to obtain a good quality of the resulting ontology. A tagged text is actually a set of texts each of them has at least a title and main parts. Such set of texts could be obtained for example by web pages scraping an informational resource. On each page there could be found a title of the article, its text and may be some other information in certain places: date, author, location, subject area, article type etc. An example of regularly structured text could be a collection of letters of or to some person. For each of the letters, there could be found in the same or approximately the same parts such information tokens like “destination address”, “sender address”, “dates of sending or/and receiving”, “sender’s name”, “addressee”, “the main text” etc. Additional semantic analysis of the main text block and the title (if exist) could be used to estimate the topic, type and subject of it.

The main text blocks which could also contain images and media content are stored in a separate document base. In OWL ontology only links on them are to be present. To this links are bind directly obtained information blocks (such as dates, locations, topics). And they become ontology classes of a certain predefined ancestor. If the information is repeated only one point is needed to be stored in the ontology. So multiplied links are created. To create such links specific properties are used. For example, there is a sending date and a letter text. A property “SendingDates” is to be created. It has

decedents each of them has a certain date as its “Domain” field value and list of letters sent on this date.

In other way is performed linking a title to the main information block. The titles are not stored in the ontology. They are tokenized to separate words and name groups, which are stored in the ontology according to their role in the sentence and without repeating. The links to main texts are bind to the entities sets (intersection) according to the meaningful terms from the corresponding title.

If a concept (only for nouns and noun groups) of those included in the intersections of concepts associated with a particular text, is included in the text more than once, it is considered that this concept is more important and decisive. Therefore, the corresponding subclasses of “Count” and “Quantity” are created for them. The same subclass “Count” can be associated with different subclasses of Quantity, which avoids duplication – for example, three times different concepts may occur in different texts. In turn, a subclass of the “Quantity” class can be linked to multiple texts links – the same concept can be included, for example, twice in different texts. This information is useful to range the ontology responses for queries by their relevancy and informative value.

Conclusion

Such automatically created ontologies could be useful for dialog and reference systems. For them, it is rather easy to create a limited number of unified formal query templates, input information for which can be concepts obtained from the user’s phrase in such manner as they obtained from titles and texts during ontology creation. Such system become rather reliable: even in case of lack of some concepts in the user’s phrase or their super saturation the rather relevant response could be obtained.

Some examples of software that implement the proposed approach was created for Ukrainian, English and Norwegian languages. As information sources for different ontologies were used: finance consulting information (Norwegian), climate change questions (English), articles by literary critics about the work of Oles Honchar (Ukrainian),

correspondence of Oles Gonchar (Ukrainian), pedagogical works of Vasily Sukhomlinsky (Ukrainian), medical rehabilitation glossary (Ukrainian). Testing of the dialog systems

Bibliography

1. [Antoniou, 2016] Antoniou G. Semantic Web. Moscow: DMK-Press, 2016.
2. [Young, 2018] B. Young, A. JungHyen. Methodology for Automatic Ontology Generation Using Database Schema Information. Mobile Information Systems, 2018. pp. 1 – 13.
3. [Elnagar, 2020] S. Elnagar, V. Yoon, M. A. Thomas. An Automatic Ontology Generation Framework with An Organizational Perspective. Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 pp. 4860 – 4869.
4. [Balakrishna, 2008] M. Balakrishna, M. Srikanth. Automatic Ontology Creation from Text for National Intelligence Priorities Framework (NIPF). Proceedings of 3rd International Ontology for the Intelligence Community (OIC) Conference. 2008. pp. 8 – 12.

Authors' Information

Anna Litvin *– Glushkov Institute of Cybernetics of NAS of Ukraine; postgraduate student. 40 Glushkova ave., Kyiv, Ukraine, 03187;* ***e-mail:*** *litvin_any@ukr.net*

Major Fields of Scientific Research*: natural language processing for analysis*

Vitalii Velychko *– Glushkov Institute of Cybernetics of NAS of Ukraine; Ph.D. Senior Researcher. 40 Glushkova ave., Kyiv, Ukraine, 03187;* ***e-mail:*** *aduisukr@gmail.com*

Major Fields of Scientific Research*: information systems with the processing of natural language objects*

Vladislav Kaverinsky *– Institute for Problems of Materials Science, Ph.D. Senior Researcher, 3 Krzhizhanovsky st., Kyiv, Ukraine;* ***e-mail:*** *insamhlaithe@gmail.com*

Major Fields of Scientific Research*: mathematical modeling*

APPLICATION FEATURES OF AGILE APPROACHES IN SOFTWARE ENGINEERING

Iryna Gruzdo, Ihor Michurin

***Abstract**: The features of agile approaches in software engineering and points of control of the schedule in the development of software products are highlighted.*

***Keywords**: Software Engineering, AGILE, SCRUM, development, business.*

Introduction

The use of agile approaches in software development has had a significant impact on the development of the IT business. therefore, highlighting the features of the agile methodology and how it differs from traditional development approaches is important for effective software development. It is also necessary to highlight the features of the control of the project schedule when using agile methodologies, which is usually performed by the project manager. The correct application of agile methodologies allows you to achieve high efficiency in software product development. The purpose of the article is to systematize the main distinctive features of the agile methodology and the features of planning the project development schedule when using them.

General features of agile approach

The emergence and development of Agile approaches is associated with the need to respond to the rapid changes in customer requirements. It is important to note that effective implementation of these approaches requires a high level of interaction between the development team and the customer. Also, each iteration in the Agile method approach takes 2-4 weeks, that is, it represents a very fast implementation of iterative and incremental methods. Moreover, each iteration is usually fixed in terms of time and total development costs. During each iteration of the project development, the team can perform several tasks simultaneously. For example, during the initial iterations, the team's focus is primarily on planning and analysis tasks. The entire content of the project is divided into a list of requirements. The request log contains a list of outstanding work,

that is, work that has not yet been completed. At the beginning of each iteration, the project development team determines from the requirements log those tasks that will be completed during the iteration and the implementation of which is currently the most important for the project. At the end of the iteration, the tasks should be completed and, if necessary, shown to the customer. Typically, customers or customer representatives must participate in the project development process at all times to interact with the development team and provide feedback. In turn, the presence of feedback allows the team to properly meet the needs of the customer. This allows you to quickly respond to changing customer needs, which can be useful for both the development team and the customer. During each iteration, when applying agile approaches, it is important to control the schedule of tasks performed by the team. To do this, the project manager needs to determine the current size of the completed work with the planned amount of work. At the end of each iteration, it is necessary to perform a retrospective analysis in order to analyze the completed iteration and possible improvement of the software product development processes. In addition, the control of the schedule is associated with the analysis of the priorities of the work plan that have not yet been completed, and their change, if necessary. Determining the level of productivity of the software product development team is important for effectively scheduling tasks. It is necessary to determine the effective duration of one iteration of the project development process, which can last from two to four weeks. In the event of changes, the task of determining their level and impact on the pace of project development becomes urgent [Hamed and Abushama, 2013], [Yashyna et al, 2017].

Specifics of using agile methodologies in companies of different sizes

Today Agile methodology is used not only in IT but also in classic business. The most popular methodologies included in Agile is: Scrum, SCRUM of SCRUMs, Scrum@Scale, Disciplined Agile Delivery (DAD), Scaled Agile Framework (SAFe), Large Scale Scrum (LeSS), Agile Unified Process (AUP), Crystal Clear Method, Dynamic Systems Development Method (DSDM), Feature-Driven Development (FDD), Scaled Agile Framework (SAFe), etc. One of the popular Agile methods is Kanban. Its simplicity

allows it to be used effectively, especially in small teams. In addition, there is no need to use additional staff to monitor the implementation of this methodology. But in medium and large teams, its use is ineffective due to the large number of tasks that the team performs. Another disadvantage of Kanban is the lack of a backlog. One of the most popular Agile methods is Scrum. It consists of iterations and allows to develop software efficiently and systematically. The backlog allows to see all the tasks and control the development process. Scrum is primarily based on self-organization and team skills. In general, Scrum can be used for teams of all sizes. In the case of a large project, a Scrum of Scrums can be applied. On the other hand, misusing Scrum or using Scrum on an inexperienced team can lead to failure. A Scrum Master is also often needed to supervise the correct application of the Scrum methodology. Scaled Agile Framework. The Scaled Agile Framework is an agile methodology that is primarily targeted at large companies with many teams. This methodology allows to organize an effective development process in large companies, but in medium and small companies, its application can be ineffective. During the study, the Scrum methodologies that are used in the development of software were studied, it can be concluded that the choice of a specific methodology depends on the project itself, its size and time required for its implementation, finances, as well as the quality of the team. But it should be noted that all Scrum technologies allow you to achieve high-quality results in a short time and get finished software. It can be concluded that when companies are faced with large and complex projects with constantly changing requirements or goals, they turn to more flexible approaches.

Conclusion

Agile methodology is most effective when implemented correctly and can increase the productivity of the software development team, accelerate customer satisfaction and reduce costs. Scrum Master Trends report [2019 Scrum Master Trends Report] from the year 2019 shows that 81% of the respondents claimed to use Kanban in addition to Scrum. Kanban, Scrum, Scrum of Scrums and the Scaled Agile Framework are effective

agile methodologies for software development in IT companies. These methodologies are the most used in software development practice.

Bibliography

1. [Hamed and Abushama, 2013] Hamed, A. M. M., Abushama, H., Popular agile approaches in software development: Review and analysis. International conference on computing, electrical and electronic engineering (ICCEEE), Khartoum, Sudan, 26-28 Aug. 2013, DOI: https://doi.org/10.1109/icceee.2013.6633925
2. [Yashyna et al, 2017] Yashyna, K. V., Yalova, K. M., Sugal, E. O., Software development agile methods overview. Collection of scholarly papers of Dniprovsk State Technical University (Technical Sciences), Vol. 1 No 30., 2017. Pp 153-156. http://sj.dstu.dp.ua/article/view/145898
3. [2019 Scrum Master Trends Report] https://www.scrum.org/resources/2019-scrum-master-trends-report

Authors' Information

Iryna Gruzdo *– Assoc. Prof. of the Department of Software Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. E-mail: irina.gruzdo@nure.ua.*

Major Fields of Scientific Research: *Information Technology; Plagiarism; Software Project Management; Business Intelligence*

Ihor Michurin *– student of the Department of Software Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. E-mail: ihor.michurin@nure.ua*

Major Fields of Scientific Research: *Information Technology; Software Engineering; Artificial Intelligence*

AN APPROACH OF DESIGNING RESTFULL API MODEL FOR DISTRIBUTED WEB APPLICATIONS

Yurii Milovidov

***Abstract**: Paper proposes an extensibility of RESRFULL API approach to perform more reliable web sites and steps towards web development.*

***Keywords**: Model-View Controller, REST, HTTP requests.*

Introduction

REST approach is widespread used in modern web-development activities. Summarizing practices if web-development companies the sense of this approach is the next: sorting of Controller methods by HTTP requests for more quick understanding of Controller structure [Milovidov Yu., 2019]. Thus, such an approach is aimed to perform cognitive and communicative functions in web-development [Chebanyuk & Markov, 2015].

Contribution of this thesis: Extended approach considering cognitive and communicative functions of development team is proposed. Advantages of this approach. Proposed model can be basic for implementing practical approaches for development of logical and automated analysis of web-page lifecycle with open extensible mode to add for example security functions to web-page lifecycle.

Proposed Approach

It is proposed to give universal manes of different Controller method analyzing whole live cycle of web-page.

Summarizing table is represented below

Init stage		
Request type		
GET Function Init(dataset)	To obtain initial data to download View Model	Generated markup and initial sets of data and

		metadata for starting operations
GET Add listener1 … listenern	To set up basic functionality (event processing) of View Model components	To create interactive instances of View Model User Interface components
GET Error processing functions	To provide secure and stable working mode for web-page	Add security functions to input fields and
optional GET refine	To refine MarkUp	To add extra settings to *.html page
Working stage		
Post Renew parts of *.html page	Provide interactive answers to User requests	Send data from html form
Delete Delete some resources DeleteResourceNAME	Delete some resource	
Put UpdateResourceName	Renew of some resource	

Conclusion

Proposed table summarizes stages of web-page lifecycle. Recommended names of methods can be base to analyze Controller web-pages by means of static analysis tools

and verify whether all necessary patterns of web-page lifecycle are realized in web-application.

Bibliography

1. [Chebanyuk E. & Markov Kr. 2015] Chebanyuk E. & Markov Kr. Software model cognitive value. International Journal “Information Theories and Applications”, Vol. 22, Number 4, 2015, 338-355
2. [Milovidov Yu. 2019] Yurii Milovidov. Comparison of web services creation technologies using testing application SoapUI. International Journal "Information Content and Processing", Vol. 6, Number 2, 2019, 60-67

Authors' Information

***Yurii Milovidov**, National University of Life and Enironmental Sciences of Ukraine, Kyiv, scientific Department of Programming Technologies. Senior Lecturer. Email: milovidov.email.ua*

Major Fields of Scientific Research: Programming technologies, web-design, Internet – technologies

SECTION: APPLIED ASPECTS OF SOFTWARE ENGINEERING

THE METHOD MULTICRITERIAL EVALUATION OF VERIFY THE QUALITY PROGRAM SYSTEMS IN USE

Yuliia Bezkorovaina

Abstract: *The main purpose of verifying quality is to ensure that the software systems are meeting the requirements and implied needs of the users. The verification of quality includes defining purpose and measures, their evaluation, and decide on it. One of the tasks is what method of evaluation to select, and develop automatic calculations. This thesis offers a solution to the problem on the example of the quality model of software systems in use and method multicriterial evaluation.*

Keywords: *verification, software, quality in use, multicriterial evaluation, quality model, SQuaRE.*

ITHEA Keywords: *D.2.1 Requirements/Specifications, D.2.4 Software/Program Verification, D.2.8 Metrics*

Conference and topic: *International Conference on Software Engineering "SoftEngine - 2021", Approaches to Software Development Life Cycle Processes Improvement, http://www.ithea.org/softengine/*

Introduction

The software systems pass through the process of the life cycle. The standard [ISO12207, 2017] described phases of the development. The software life cycle is allocated to the process: agreement, organizational project-enabling, technical management, and technical. To initiate the process, supplier and customer must agree on a purpose, outcomes, and activities and tasks [ISO12207, 2017], [IEEE1012, 2017]. This thesis will be about the verification of quality and acceptance of software systems.

There are a series of standards about quality verification the software systems. The part of those standards describes different quality models [ISO25010, 2011]. There is a quality in use model [ISO25022, 2016], quality product model, quality data model, quality

services model, and so on. This thesis is dedicated to verifying quality in the use of the final software system.

The purpose of evaluating the quality of the final software product is being to decide on the acceptance of the product and determine the time of the product release [ISO25040, 2011]. To assess the quality of the finished software system during use, you need to determine the requirements for it. The requirements specification considers the budget, target dates, and the purpose of the assessment. The specification can change or improve during the development of software systems [ISO25040, 2011].

Related work

There is an empirical analysis of quality standards concerning usability characteristics and another software system quality characteristics [Bevan et al, 2016], the implementation of a product quality model according to the standard [Kazuhiro, 2013], [Yusupova et al, 2016]. Other sources are considered [Danyk et al, 2016], [Pysarchuk and Pinchuk, 2006], [Pysarchuk et al, 2019], which prove in practice the effectiveness of the proposed method.

Task and challenges

The complexity of the quality evaluation task is justified by the variety of stakeholders and users, areas of application of quality models, the complexity of software systems and the variety of their life cycle, quality management tools, etc. The main task of research is to prove the effectiveness of the multicriteria method for the quality model in use evaluation of the software systems.

Body

The quality model in use is composed of the standard [ISO25022, 2016] and evaluation of the model [ISO25040, 2011]. Quality in use model composed of five characteristics that relate to the outcome of interaction when a product is used in a particular context of use. So, the quality in use model (characteristics, subcharacteristics, and measures) represent in table 1. Each measure required the calculation of metrics. The value of each

criterion affects the software system quality in use. So, we can say that this is a problem with many diverse criteria.

Table 1. Model of quality in use

Characteristics	Subcharacteristics	Measures
Effectiveness	-	Tasks completed
		Objectives achieved
		Errors in a task
		Tasks with errors
		Task error intensity
Efficiency	-	Task time
		Time efficiency
		Cost-effectiveness
		Productive time ratio
		Unnecessary actions
		Fatigue
Satisfaction	Usefulness	Overall satisfaction
		Satisfaction with features
		Discretionary usage
		Feature utilization
		The proportion of users complaining
		The proportion of user complaints about a particular feature
	Trust	User trust
	Pleasure	User pleasure

Characteristics	Subcharacteristics	Measures
	Comfort	Physical comfort
Freedom from risk	Economic risk mitigation	Return on investment
		Time to achieve a return on investment
		Business performance
		Benefits of IT Investment
		Service to customers
		Website visitors converted to customers
		Revenue from each customer
		Errors with economic consequences
	Health and safety risk mitigation	User health reporting frequency
		User health and safety impact
		Safety of people affected by the use of the system
	Environmental risk mitigation	Environmental impact
Context coverage	Context completeness	Context completeness
	Flexibility	Flexible context of the use
		Product flexibility
		Proficiency independence

As we can see, there is a problem to decide on the readiness of the software system. The standard [ISO25040, 2011] proposes different methods of evaluation quality. But in this thesis, the author proposes to evaluate the final version of the software system with the method of multicriteria evaluation using nested convolutions according to a nonlinear scheme of compromises [Voronin, 2017], [Kharchenko and Pysarchuk, 2014].

The quality model transforms into an infographic model of factors, indicators, and criteria without subcharacteristics that are not required by the standard [ISO25010, 2011]. By heuristic analysis of the quality model in use formed an infographic model of factors, indicators, and criteria of the quality in use (Table 2).

Table 2. Infographic model factors, indicators, and criteria

№	Factors	№	Factors in the group	Indicator	Criteria
1	Effectiveness	1	Tasks completed	C_{et}	→max
		2	Objectives achieved	OA_e	→max
		3	Errors in a task	E_{ef}	→min
		4	Tasks with errors	T_{ee}	→min
		5	Task error intensity	E_{et}	→min
2	Efficiency	6	Task time	T_{et}	→min
		7	Time efficiency	U_{ee}	→min
		8	Cost-effectiveness	EC_e	→min
		9	Productive time ratio	T_{epr}	→max
		10	Unnecessary actions	UN_a	→min
		11	Fatigue	F_e	→min
3	Satisfaction	12	Overall satisfaction	S_o	→max
		13	Satisfaction with features	S_f	→max
		14	Discretionary usage	U_{sd}	→max

№	Factors	№	Factors in the group	Indicator	Criteria
		15	Feature utilization	F_{su}	→max
		16	The proportion of users complaining	P_{cc}	→min
		17	The proportion of user complaints about a particular feature	P_{cf}	→min
		18	User trust	T_s	→max
		19	User pleasure	P_s	→max
		20	Physical comfort	C_s	→max
4	Freedom from risk	21	Return on investment	ROI_e	→max
		22	Time to achieve a return on investment	T_{roi}	→min
		23	Business performance	P_b	→max
		24	Benefits of IT Investment	B_i	→max
		25	Service to customers	S_c	→max
		26	Website visitors converted to customers	V_c	→max
		27	Revenue from each customer	R_c	→max
		28	Errors with economic consequences	E_{ec}	→min
		29	User health reporting frequency	F_{uhr}	→max
		30	User health and safety impact	IMP_{uhs}	→max
		31	Safety of people affected by the use of the system	S_{as}	→max

№	Factors	№	Factors in the group	Indicator	Criteria
		32	Environmental impact	Im_e	→max
5	Context coverage	33	Context completeness	C_{cc}	→max
		34	Flexible context of the use	F_{fc}	→max
		35	Product flexibility	P_{fc}	→max
		36	Proficiency independence	P_{ic}	→max

The formation of the decision-making model consists of aggregating the list of contradictory partial criteria to the generalized assessment. It was therefore selected for convolution nonlinear compromise scheme [Voronin, 2017], [Kharchenko and Pysarchuk, 2014].

The main idea of the approach is in the following sequence:

1. Forming a system of partial indicators of quality;
2. Determining the values of partial indicators of decision quality;
3. Formation of a generalized quality indicator from a system of partial and bringing it to a relative value.
4. Analysis of evaluation and decision-making on the compliance of the software system in use.

Interpretation of the received decision consists of the resulted values of the integrated estimation to a uniform scale of change - from 0 to 1. The received numerical estimation can be reduced to a linguistic category according to a fundamental scale of estimation (Table 3).

Table 3. Fundamental rating scale

Integrated quality assessment	Linguistic category of quality
1.0-0.7	High

0.7-0.5	Good
0.5-0.4	Satisfactory
0.4-0.2	Low
0.2 and less	Unsatisfactory

So, after calculation, the model can help us decide about the software systems quality in use. If we had received a value of 0.6, it means that the software system has a good quality. That quality depends on a set of measurements. Supplier and customer may decide or to accept as a final version, or develop onwards a software system.

Conclusion

This thesis considered the quality model in use by the standard [ISO25010, 2011], [ISO25022, 2016], [ISO25040, 2011]; method of evaluation of the quality in use [Danyk et al, 2016], [Pysarchuk and Pinchuk, 2006], [Pysarchuk et al, 2019]. The next steps in research are to evaluate the quality product model by standard ISO/IEC 25023 “Measurement of system and software product quality”, compare the effectiveness of evaluation methods and identify the most effective.

Bibliography

1. [ISO12207, 2017] ISO/IEC/IEEE 12207:2017. Systems and software engineering — Software life cycle processes. ISO, Switzerland, Geneva, 2017, 145 p.
2. [IEEE1012, 2017] IEEE Std 1012:2017. Standard for System, Software, and Hardware Verification and Validation, in IEEE Std 1012-2016 (Revision of IEEE Std 1012-2012/ Incorporates IEEE Std 1012-2016/Cor1-2017) - Redline, 2017, 465 p.
3. [ISO25010, 2011] ISO/IEC 25010:2011. Systems and software engineering. Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models. ISO, Switzerland, Geneva, 2011, 34 p.
4. [ISO25022, 2016] ISO/IEC 25022:2016. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of quality in use. ISO, Switzerland, Geneva, 2016, 41 p.
5. [ISO25040, 2011] ISO/IEC 25040:2011. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process. ISO, Switzerland, Geneva, 2011, 45 p.

6. [Bevan et al, 2016] Bevan N., Carter J., Earthy J., Geis T., Harker S.. New ISO Standards for Usability, Usability Reports and Usability Measures. Conference: International Conference on Human-Computer Interaction, Theory, Design, Development and Practice. Vol. 9731, pp 268-278. DOI: 10.1007/978-3-319-39510-4_25
7. [Yusupova et al, 2016] Yusupova N., Gvozdev V., Yanchek K., Smetatina O., Morozov A. MODEL'' Kachestva SQUARE programmnogo producta. INFORMATION TECHNOLOGIES FOR INTELLIGENT DECISION MAKING SUPPORT (ITIDS'2016). Ufa, 17–19 maja, 2016, pp. 54-59
8. [Kazuhiro, 2013] Kazuhiro Esaki. Verification of Quality Requirement Method Based on the SQuaRE System Quality Model. American Journal of Operations Research, 2013, 3, 70-79. DOI: 10.4236/ajor.2013.31006
9. [Danyk et al, 2016] Danyk Yu., Pysarchuk O., Lagodniy O., Vyporhoniuk O. Matemetichna model bagatokriterialnogo otsiniuvania efektyvnosti internet-saitiv tsilovogo spriamuvania. Ingeneria programnogo zabezpechenia. Visnik, Issue 1, Vol. 76. 2016, pp. 114-120
10. [Pysarchuk et al, 2019] Pysarchuk O., Bezkorovaina Yu., Dyshlevyi O., Skalova V. Metodyka bahatokryterialnogo ocinyuvannya vidpovidnosti prohramnoho zabezpechennya vymoham zamovnyka. Naukoyemni texnolohiyi. 2019. №1 (41). Kyiv: NAU, pp. 3-9.
11. [Pysarchuk and Pinchuk, 2006] Pysarchuk O., Pinchuk O. Metodyka vyboru zasobiv z vikorystanniam bagatokriterialnyh modeley. Visnik: tehnichni nauky. Vol. 4, Issue 39, 2006, pp. 152-158.
12. [Voronin, 2017] Voronin A. Multi-Criteria Decision Making for the Management of Complex Systems, Multi-Criteria Decision Making for the Management of Complex Systems, March, 2017, 201 p. DOI: 10.4018/978-1-5225-2509-7
13. [Kharchenko and Pysarchuk, 2014] Kharchenko V., Pysarchuk O. Neliniyne ta bagatokriterialne modeliuvannia protcesiv u systemakh keruvania rukhom, NAU, Kyiv. 2014, 372 p.

Authors' Information

Yuliia Bezkorovaina *– Senior Lecturer of Software Engineering Department, National Aviation University, Kyiv, Ukraine; e-mail: yuliia.bezkorovaina@nau.edu.ua*

Major Fields of Scientific Research: *Quality, software verification, software engineering*

SOFTWARE TECHNOLOGY OF FLOOD SIGNALING AND FORECASTING IN THE TRANSCARPATHIAN REGION

Volodymyr Polishchuk

Abstracts: *The research constructs a generalized algorithm for signaling and forecasting floods in the Transcarpathian region, based on a hybrid intelligent model, using the analysis of reasoning and experience of water management experts. Based on the studied concepts, the construction of software technology in the form of web technology for an expert administrator and a cross-platform application for the user is envisaged.*

Keywords: *Forecasting, hybrid data, software, crisis management*.

Introduction

In the life of modern society, the problems associated with overcoming the consequences of emergencies, which lead to huge material losses and sometimes human losses, play a significant role. A special place among many emergencies belongs to floods of man-made and natural origin, catastrophic flooding, and flooding of territories, which are some main socio-environmental problems of our time [Kelemen M., 2019]. Floods have long been a problem for human society, but today their size and frequency are growing rapidly, causing material damage and, worst of all, human casualties. The Transcarpathian region is no exception. The negative effects of flooding cannot be completely avoided, but their impact can be reduced by forecasting and signaling the population while giving time to prepare or mobilize society and relevant bodies.

In the context of crisis management and security of citizens, there is a complex and urgent task of developing a mobile software product that allows, according to the geolocation data of the smartphone, to warn the user about the level of flood risk.

Thus, the aim of the study is to develop a generalized algorithm and software technology for signaling and forecasting floods in the Transcarpathian region, based on a hybrid intelligent model, using the analysis of reasoning and the experience of water experts.

Generalized algorithm of flood signalling and forecasting

Suppose we have n measuring stations, which every period of time t_i send data on the actual water level in n_i. To analyze the measurement data, we use the method of regression analysis, namely the method of pairwise linear regression. We present a general algorithm for signaling and forecasting floods based on a hybrid intelligent model [Polishchuk V., 2019] to determine the level of flood risk. The model is able to derive a normalized assessment of water level rise, uses the analysis of reasoning and experience of water management experts, reveals the vagueness of input estimates, increases the degree of validity of further management decisions based on the results. The output of the model is a normalized assessment of the danger of a flood situation and a linguistic interpretation of the level of danger of a flood situation.

Step 1. Obtaining data and building a time series

In the first step, we construct a time series where y_i is the water level, x_i is the fixation time.

Step 2. Determination of regression coefficients

In the second step, the least-squares method (LSM) is used to determine the regression parameters. Find the values of the regression coefficients. LSM allows obtaining such estimates of parameters in which the sum of the squares of the deviations of the actual values of the resultant feature y, from the theoretical values of y at x is the minimum.

Step 3. Obtaining predicted values

In the third step, we construct a pairwise linear regression equation to predict the water level. To do this, we substitute future periods x into the equation to obtain the predicted values.

Step 4. Fuzzyfication of input data

In the fourth step, we will fuzzification the input data, based on the predicted data, the maximum and minimum allowable water level in the area (stations) $p_{it(max)}$, and $p_{it(min)}$, according to the formula:

$$\mu(p_{it}) = \begin{cases} 0, & p_{it} < a; \\ \frac{1}{2} + \frac{1}{2} cos\left(\frac{p_{it} - b}{b - a} \cdot \pi\right), & a \leq p_{it} \leq b; \\ 1, & p_{it} > b. \end{cases} \quad i = \overline{1,n} \tag{1}$$

Where we set $a = p_{it(min)}$ the safe (lowest) level, $b = p_{it(max)}$ respectively - dangerous (highest) level. The concept of constructing membership functions assumes the following meaning: the higher the water level, the greater the risk of flooding, and the value of the membership function goes to one [Kelemen M. et al., 2020].

Step 5. Estimation of water level projection

In the fifth step, we estimate the water level projection for the relevant area over a period of time:

$$P_{it} = \mu(p_{it}) \cdot 100. \tag{2}$$

The obtained value P_{it} characterizes the estimation of the water level projection for the respective area in a certain period of time [Polishchuk V., 2019].

Step 6. Obtaining a flood risk assessment

The sixth step requires the water expert to assess the risk of flooding and select one of the proposed linguistic variables: H {Low risk of flooding}; C {Medium risk of flooding}; B {High risk of flooding}. After that, we calculate the normalized assessment of the risk of flood situation by the formula:

$$\mu(Z_{it}) = \begin{cases} 0, & P_{it} < 0; \\ \left(\mu(p_{it})\right)^k, & 0 \leq P_{it} \leq 100; \\ 1, & P_{it} > 100. \end{cases} \tag{3}$$

Where k is the degree of risk of flooding. For example, put: for low risk of flooding k = 1/3, for medium risk k = 1, and for high risk of flooding k = 3/5.

According to the received normalized estimation, we will present a linguistic interpretation of the level of danger of a flood situation: $\mu(Z_{it}) \in [0; 0{,}4)$ – LS: low level of flood

danger; $\mu(Z_{it}) \in [0{,}4; 0{,}6)$ – AS: average level of flood danger; $\mu(Z_{it}) \in (0{,}6; 1]$– HS: high level of flood danger.

The range of linguistic interpretation, to allow for adjustment, is specified in the settings.

Step 7. Decision-making on flood signaling in Transcarpathian region

In the seventh step, based on the linguistic interpretation, the decision-maker (water management expert) can signal and/or inform users about the dangerous situation, respectively, through the mobile application.

Conclusion

As a result of the implementation of smart technologies of crisis management, strengthening of protection against harmful effects of water of settlements, minimization of the caused losses, and creation of safe living conditions of communities of the region is carried out.

An algorithm based on a hybrid intelligent model has been developed, which uses the analysis of reasoning and experience of water management experts, reveals the vagueness of input estimates, increases the degree of validity of further management decisions based on the results. Web technology and a cross-platform user application will be designed for the expert administrator part. With the help of the application, the user, based on geolocation data, receives notifications about the possible level of danger. In addition, it is assumed that the user can view the state of the flood situation in the location of his stay, the water level, and the predicted danger. Real data will be obtained on the basis of publicly available AIS "Tisa", consisting of 50 stations throughout the Transcarpathian region and neighboring EU countries.

Bibliography

1. [Kelemen M., 2019] Kelemen M. (2019). Safety and Knowledge Alliance of Aviation Education: Human factors in Aviation Safety and Air Law. AMELIA Aneta Siewiorek, Poland, 2019.

2. [Polishchuk V., 2019] Polishchuk V. (2019). Technology to Improve the Safety of Choosing Alternatives by Groups of Goals. *Journal of Automation and Information Sciences*, 51 (9), 66-76. DOI: 10.1615/JAutomatInfScien.v51.i9.60
3. [Kelemen M. et al., 2020] Kelemen, M., Polishchuk, V., Gavurová, B., Andoga, R., & Matisková, D. (2020). The Expert Model for Safety Risks Assessment of Aviation Environmental Projects' Implementation Within the Investment Phase of the Project. *IREASE,* Vol 13, No 6, 198-207. doi: 10.15866/irease.v13i6.18268).

Authors' Information

Volodymyr Polishchuk *– Assoc. Prof. of Department Software Systems, Uzhhorod National University, Uzhhorod, Ukraine. E-mail: volodymyr.polishchuk@uzhnu.edu.ua*

Major Fields of Scientific Research: *Intelligent Information Processing, Data Mining, Expert Systems, Fuzzy set, Knowledge Engineering and management, Intelligent software and Computational Intelligence*

SECTION: SOFTWARE ENGINEERING APPLICATION DOMAINS

AUTOMATIC SEARCH FOR INVOLVANTS OF INFORMATION MESSAGES IN THE VOICE DATABASES

Serhii Zybin, Yana Bielozorova

***Abstract**: Parameters of individual voice characteristics in modern automatic systems of speaker identification was researched. Efficiency of such systems mainly depends on the methods for determination of these parameters and stability of spectral characteristics and the main tone frequency. The structure and concept for an automated search in voice databases was described.*

***Keywords**: Speaker identification system, wavelet, coding and information theory, pattern analysis, phonogram.*

***Conference and topics**: SoftEngine 2021; Software Engineering Application Domains - Image Processing and Computer Vision.*

Objectives

Spectrum analysis of audio data based on Fourier transform mathematical apparatus is the key factor in majority of modern systems for solving the tasks of speaker identification by means of voice characteristics. It is caused by several factors, known neurophysiological rules of audio information processing by primary auditory receptors and absence of more effective analysis methods and, to a certain extent, by historical traditions in this field.

However, in spite of a lot of investigations in this field and creation of rather effective systems for voice characteristics identification and development of systems of identification and text entering of voice information, there is not enough clearness in the principal theoretical and practical issues concerning speech technology till now. Mainly, it is caused by absence of effective mathematical tool for analysis of voice audio information [1,2].

The goal is to study the effectiveness and develop a software system for the formation of voice databases of participants in information messages, as well as automatic search in the voice database.

Model of automated search in voice databases

Concept implemented in the experimental system of instrumental identification of speaker voice characteristics was proposed. This concept is based on presentation of speech fragments as a set of multifractal structures. In order to determine parameters of multifractal structures the wavelet analysis with special basis in form of two-parameter Morlet wavelet is used [3,4].

Till now, actually, the single methodological direction for building the mathematical models for calculation of phoneme parameters and voice characteristics is the application of some of modifications of spectrum analysis based on Fourier transform. However, the application of this spectrum analysis method has number of serious disadvantages (known in the years since Helmholtz). Signal spectrum of “instrumental” transformers of acoustical apparatus – hair cells can be considerably different from Fourier spectrum, including (that is rather important) arrangement of formants in frequency domain. Human acoustical apparatus is able to distinguish frequencies with a step smaller than 1 Hz (within the range up to 500 Hz).

It is clear, that physical and mathematical models of voice information transformation should not necessarily strictly follow neurophysiologic rules. However, analysis of voice information based on orthogonal Fourier transformation for many decades at the phonemic level shows rather high variability of spectra for the same phonemic structures in short time intervals. It is a significant obstacle for effective assessment of phoneme parameters and voice characteristics [5].

It is possible to suppose that alternative models of non-orthogonal frequency transformations of voice acoustic signal in short time intervals allow to obtain more stable assessments of the formant speech structure and parameters of voice characteristics. Experimental model of the program product, created on the basis of investigations, for

realization of these tasks was approved in the expert departments of the Ministry of Justice and the Ministry of Internal Affairs of Ukraine.

Experimental program system for search of involvants of information messages has a form of a search engine. Thus, the search results are not identical to identification of person by voice, because they are just results of ranking by the similarity degree of separate parameters of voice signals.

Using digital records of information messages the EPS performs an automatic calculation of parameters of voice characteristics and further ranking of these characteristics in the voice database.

The EPS uses a ranking method by four different criteria. They include:

- calculation of similarity of two-dimensional probability density functions curves for the main tone frequency and arrangement in the spectrum of seven formants, extracted from the speech recorded in the phonogram;
- calculation of similarity of probability density functions curves for each of these signs separately;
- calculation of similarity degree for the absolute maxima of formats spectra, extracted from the speech recorded in the phonogram.

Conclusion

As result of fulfilled work the specialized program system for search of involvants of anonymous messages in large intergovernmental voice databases should be created. The program system should have several language localizations – English, German, Ukrainian, Russian and other languages. The possibility for quick modification for adding the new EU localization should be provided. The system should ensure quick access (with regulation of access rights) to accumulated distributed databases of voice messages within the frames of unified standard of the EU countries. Structuring of the system of voice messages within the frames of distributed databases will allow to realize access to operative solving of problems, including voice messages analysis systems of other developers.

Investigative part of the work should ensure assessment of efficiency and practical utility of search for involvants of voice messages in the EU countries databases.

Bibliography

1. [Rubalsky et al, 2014] O. Rubalsky, V. Solovyov, A. Shablja, V. Zhuravel. New tools to identify the person for voice of Data (Russian) Protection and Security of Information Systems: Proceedings of the 3rd International Scientific Conference (sity. Lviv, 05 - June 6, 2014). - Lviv: - p.p.. 110 - 112.
2. [Mallat, 1999] Mallat S. A wavelet tour of signal processing, Courant Institute, New York University, 1999, 671 pp.
3. [Solovyov, 2014] V. Solovyov, Y. Byelozorova: Multifractal approach in pattern recognition of an announcer's voice. Polish Academy of Sciences University of Engineering and Economics in Rzeszów, Teka, Vol. 14, no 2, p.p. 164-170, 2014.
4. [Solovyov, 2013] V. Solovyov. Using multifractals to study sound files (Russian). Visnik of the Volodymyr Dahl East Ukrainian national university. Vol 9 (151). – p. p. 281-287, 2013.
5. [Solovyov, 2013] V. Solovyov. Using the fractal dimension of audio files in the problem of segmenting the audio file (Russian). Visnik of the Volodymyr Dahl East Ukrainian national university. Vol 5 (194). – p. p. 165-169, 2013.

Authors' Information

Serhii Zybin *– DSc, Associate Professor, Head of Software Engineering Department,National Aviation University, Kyiv, Ukraine.*

E-mail: *zysv@ukr.net*

Major Fields of Scientific Research: *game theory, cloud computing, cyberspace, distributed nets, software engineering, information security*

Yana Bielozorova *– Senior Lecturer of Software Engineering Department, National Aviation University, Kyiv, Ukraine.*

E-mail: *bryukhanova.ya@gmail.com*

Major Fields of Scientific Research*: Speech Recognition Models, Wavelet analysis, Software Architecture*

BAYESIAN REASONING AND MACHINE LEARNING

Michael Gr. Voskoglou

Abstract*: Although the Bayes' theorem is a straightforward consequence of the equation calculating the value of a conditional probability, Bayesian reasoning has been proved to be very important for the whole science. The present work discusses applications of the Bayes' theorem to everyday life situations and its importance for Machine Learning.*

Keywords*: Conditional Probabilities, Bayes' Theorem, Artificial Intelligence, Machine Learning.*

ITHEA Keywords*: G Mathematics and Computing.*

Conference and topic*: International Software Engineering Conference, National Aviation University, Kyiv, Ukraine, 12.04 – 14.04, 2021.*

Introduction

Artificial Intelligence (AI) is the branch of Computer Science focusing on the theory and practice of creating intelligent machines which work and react like humans, i.e. being able to think, hear, talk, walk and even feel [Mitchell, 2019]. AI, introduced by John McCarthy in 1956 [Moor, 2006], is a synthesis of ideas from mathematics, engineering, technology and science which has the potential to generate enormous benefits to the human society.

Probability and Statistics are among the main mathematical tools used in AI applications in which uncertainty plays an important role. Both of them, however, have been developed on the basis of the principles of the bivalent logic. Consequently, they are able to tackle effectively only the cases of uncertainty which are due to randomness and note those due to imprecision. In cases of imprecision the Zadeh's infinite-valued fuzzy logic and other related to it theories [Voskoglou, 2019] come to bridge the existing gap.

As we shall see here, however, the Bayes' theorem calculating the value of conditional probabilities introduces a kind of multi-valued logic tackling the existing due to imprecision uncertainty in a way analogous to fuzzy logic. The present work discusses

the importance of Bayesian reasoning for Machine Learning, i.e. the branch of AI that refers to any computer program that can "learn" by itself from a training data set.

The Bayes' Theorem

Let A and B be two intersecting events. Then it is straightforward to check [Schuler & Lipschutz, 2010] that the *conditional probability* for the event A to happen when the event B has already happened is calculated by

$$P(A/B)=\frac{P(A\cap B)}{P(B)} \quad (1)$$

In the same way one finds that

$$P(B/A)=\frac{P(A\cap B)}{P(A)}, \text{ or } P(A\cap B) = P(B/A)\,P(A).$$

Replacing the value of $P(A\cap B)$ to (1) one finds that

$$P(A/B)=\frac{P(B/A)P(A)}{P(B)} \quad (2).$$

Equation (2), which calculates the conditional probability P(A/B) with the help of the inverse in time conditional probability P(B/A), the *prior probability* P(A) and the *posterior probability* P(B), is known as the *Bayes' theorem* (or rule, or law). In other words, the Bayes theorem calculates the probability of an event based on prior knowledge of conditions related to that event.

The value of the prior probability P(A) is fixed before the experiment, whereas the value of the posterior probability is calculated with the help of the experiment's data. Usually, however, there exists an uncertainty about the exact value of P(A). In such cases, considering all the possible values of P(A), we obtain through the Bayes' theorem different values for the conditional probability P(A/B). Therefore, the Bayes' theorem introduces a kind of multi-valued logic tackling the existing, due to the imprecision of the value of the prior probability, uncertainty. In other words, one could argue that Bayesian Reasoning constitutes an interface between bivalent and FL.

The Bayes' theorem was first appeared in the work "An Essay towards a Problem in the Doctrine of Chances" of the 18th century British mathematician and theologian Thomas

Bayes. This essay was published by Richard Price in 1763, after the Bayes' death, in the "Philosophical Transactions of the Royal Society of London". The famous French mathematician Laplace (1749-1827), independently from Bayes, pioneered and popularized the Bayesian probabilities.

When applied in practice, the Bayes' theorem may have several interpretations. In social sciences, for example, it describes how a degree of belief expressed as a probability P(A) is rationally changed according to the availability of related evidence. In that case the probabilities involved in the Bayes' theorem are frequently referred as *Bayesian probabilities,* although, mathematically speaking, Bayesian and conditional probabilities are actually the same thing. The outcomes of Bayesian reasoning, however, are not always compatible to the common beliefs. This will be illustrated here by the following two timely examples, due to the current COVID-19 pandemic, concerning the creditability of the viruses' diagnostic tests.

Example 1: Statistical data show that 2% of the inhabitants of a country have been infected by a dangerous virus. Mr. X, who has not any symptoms of the corresponding disease, makes a diagnostic test, the statistical accuracy of which is 97%. The test is positive. What is the probability for Mr. X to be a carrier of the virus?

Solution: Consider the events A: The subject is a carrier of the virus, and B: The test is positive. On the basis of the given data it turns out that P(A)=0.02 and P(B/A)=0.97. Further, among 100 inhabitants of the country, 2 on average are carriers and 98 are non-carriers of the virus. Assuming that all those people make the test, we should have on average 2x97%=1.94 positive tests from the carriers and 98x3%=2.94 positive tests from the non-carriers of the virus, i.e.4.88 in total positive tests. Therefore, P(B)=0.488. Replacing the values of P(A), P(B/A) and P(B) in equation (2) one finds that P(A/B)≈0.398. Therefore, the probability for Mr. X to be a carrier of the virus is only 39.8% and not 97%, as it could be thought through a first, rough estimation!

Example 2: Assume now that Mr. X has some suspicious symptoms of the corresponding disease and that 85% of the people presenting such symptoms have been infected by

the virus. Mr. X makes the test, which is positive. What is now the probability for Mr. X to be a carrier of the virus?

Solution: Let A and B be the events defined in Example 1. Here we have that P(A)=0.85 and P(B/A)=0.97. Further, assuming that 100 people who have suspicious symptoms make the test, we should have on average 85x97%=82.45 positive tests from the carriers and 15x0.3% =0.45 from the non-carriers of the virus, i.e. 82.9 in total positive tests. Therefore, P(B)=0.829. Replacing the values of P(A), P(B/A) and P(B) in equation (2) one finds that P(A/B)≈0.995. In this case, therefore, the probability for Mr. X to be a carrier of the virus is 99.5%, i.e. exceeds the statistical accuracy of the test!

In general, the sensitivity of the solution is great, depending on the values of the prior probability P(A). The greater the value of P(A), the higher the creditability of the test.

The Importance of Bayesian Reasoning for Machine Learning

Although the Bayes' theorem is a straightforward consequence of equation (1) calculating the value of a conditional probability, Bayesian reasoning has been proved to be very important for everyday life situations and for the whole science [Athanassopoulos & Voskoglou, 2020]. Recent researches give evidence that even most of the mechanisms under which the human brain works are Bayesian [Bertsch McGrayne, 2012]. Consequently, Bayesian reasoning becomes very useful for AI, which focuses on the design and construction of machines that mimic the human behavior.

Recently researchers have used ML techniques to develop through the Internet a new generation of web-based *smart learning systems (SLS's)* for various educational tasks. A SLS is actually a knowledge - based software used for learning and acting as an intelligent tutor in real teaching and training situations. Such systems have the ability of reasoning and of providing inferences and recommendations by using heuristic, interactive and symbolic processing and by producing results from the big data analytics [Voskoglou & Salem, 2020]. The smart systems of AI are supplied with Bayesian algorithms in order to be able to recognize the corresponding structures and to make autonomous decisions.

The physicist and Nobel Prize winner John Mather has already expressed his uneasiness about the possibility that the Bayesian machines could become too smart in future, so that to make humans to look useless (http://edge.org/response-detail/26871)!.

Conclusion

The discussion performed in this work leads to the conclusion that Sir Harold Jeffreys (1891-1989), a British mathematician who introduced the concept of the Bayesian algorithm and played an important role in the revival of the Bayesian view of probability, has uprightly characterized the Bayes' theorem as the "Pythagorean theorem of Probability theory" (Jeffreys, 1973).

References

1. [Athanassopoulos & Voskoglou, 2020] Athanassopoulos, E. & Voskoglou, M.Gr., A Philosophical Treatise on the Connection of Scientific Reasoning with Fuzzy Logic, Mathematics, 8, article 875, 2020.
2. [Bertsch McGrayne 2012] Bertsch McGrayne, S., The Theory that would not die, Yale University Press, New Haven and London, 2012.
3. [Jeffreys 1973] Jeffreys, H., Scientific Inference, 3d Edition, Cambridge University Press, UK, 1973.
4. [Mitchell, 2019] Mitchell, M., Artificial Intelligence: A Guide for Thinking Humans, Parrar, Straus and Gtraux: NY, USA, 2019.
5. [Moor, 2006] Moor, J., The Dartmouth college artificial intelligence conference: The next fifty years, Al Magazine, 27, 87–91, 2006.
6. [Schuler & Lipschutz, 2010] Schuler, J. & Lipschutz, S., Schaum's Outline of Probability, 2nd Edition, McGraw-Hill, NY, USA, 2010.
7. [Voskoglou, 2019], Voskoglou, M.Gr., Generalizations of Fuzzy Sets and Relative Theories, in Voskoglou, M.Gr. (Ed.), An Essential Guide to Fuzzy Systems, pp. 345-368, Nova Science Publishers, NY, USA, 2019.
8. [Voskoglou & Salem, 2020], Voskoglou, M.Gr & Salem, A.-B.M., Benefits and Limitations of the Artificial with Respect to the Traditional Learning of Mathematics, Mathematics, 8, Article 611, 2020

Author's Information

Michael Gr. Voskoglou *– Graduate Technological Educational Institute (TEI) of Western Greece; Professor Emeritus of Mathematical Sciences; 26334 Patras, Greece e-mail: mvoskoglou@gmail.com*

Major Fields of Scientific Research: *Algebra, Artificial Intelligence, Fuzzy Logic, Markov Chains, Mathematics Education*

UKRVECTŌRĒS: AN NLU-POWERED TOOL FOR KNOWLEDGE DISCOVERY, CLASSIFICATION, DIAGNOSTICS & PREDICTION

Vitalii Velychko, Kyrylo Malakhov, Oleksandr Shchurov

Abstract. *The given paper presents a free and open source toolkit which aim is to quickly deploy web services handling pre-trained distributional semantic models (static word embeddings). It completes the pipeline between training such models and sharing the results to the general public. Our web service, UkrVectōrēs, provides all the necessary routines for online access, sharing and querying to the trained models via modern Web application, built with Angular, Flask and Docker platforms. We also describe a demo setup of the UkrVectōrēs toolkit, featuring several efficient models ("WhiteBook" model, trained using a dataset – the book «The White Book of Physical and Rehabilitation Medicine» Ukrainian version; "Fiction" model, trained using a dataset – Ukrainian fiction literature) for Ukrainian.*

Key words: *Computational Linguistics, Distributional Semantics, Word Embeddings, Natural Language Processing, Machine Learning, Word2Vec, Visualization, Knowledge Discovery.*

ITHEA Keywords*: H.3.3 INFORMATION SEARCH AND RETRIEVAL, H.5.1 MULTIMEDIA INFORMATION SYSTEMS.*

Conference and topic*: Software Engineering Application Domains: Databases and Knowledge Bases.*

Introduction

In this paper we present UkrVectōrēs [docsim, 2021], a free and open-source toolkit to deploy web services implementing distributional semantic models, primarily static word embeddings. UkrVectōrēs is an NLU-Powered (Natural language understanding) tool for knowledge discovery, classification, diagnostics and prediction.

Distributional semantic models are well established in the field of computational linguistics and have been here for decades [Turney, 2010]. However, recently they received substantially growing attention. The main reason for this is a possibility to employ artificial neural networks trained on small/large/huge corpora to learn low-

dimensional distributional vectors for words - word embeddings. The most well-known tools in this field now are word2vec, fastText built on top of the Skip-Gram and CBOW algorithms, which allows fast training on huge amounts of raw linguistic data [Mikolov, 2013]. Distributional static word embeddings are now arguably the most popular way to computationally handle lexical semantics. In such models, every word is represented with a dense float vector: the number of dimensions is usually in the order of hundreds, and may vary depending on the purpose of the model [Kutuzov, 2016]. Static word embeddings represent meaning of words, and can be use in almost any linguistic task: named entity recognition, sentiment analysis, machine translation, corpora comparison, compute semantic similarity between entities (such as words, terms and multi-word expressions) etc. Approaches implemented in word2vec, fastText and other similar tools are being extensively studied and tested in application to the English language [Baroni, 2016]. However, for many other languages the surface is barely scratched. Thus, it is important to facilitate research in this field and to provide access to relevant services for various linguistic communities, in particular Ukrainian linguistic community. UkrVectōrēs service creation were inspired by WebVectors [WebVectors, 2021], [Kutuzov, 2016] and RusVectōrēs [Kutuzov, 2016], [RusVectores, 2021] Web services.

UkrVectōrēs deployment basics with Docker Compose

The UkrVectōrēs toolkit is a web interface between distributional semantic models and a user. Under the hood, we use Gensim open source Python library [Gensim, 2021] which deals with models' operations. The user interface is implemented in Typescript (Angular web framework) and runs on top of the regular Nginx HTTP server or as a standalone service (using uWSGI application server). It communicates with Gensim (functioning as a daemon with our wrapper) via REST API (using particular JSON model), sending front-end queries and receiving back models' answers. Such architecture allows fast simultaneous processing of multiple users querying multiple models over network. Models themselves are permanently stored in memory, eliminating time-consuming stage of loading them from hard drive every time there is a need to process a query.

UkrVectōrēs can be useful in a very common situation when one has trained a distributional semantics model for one's particular corpus or language, but then there is a need to demonstrate one's results to general public.

To simplify the application deployment process as much as possible, it was decided to use the Docker Compose [Docker, 2021] technology (Docker Compose is a tool for defining and running multi-container Docker applications). As UkrVectōrēs application grow more complex, we may find significant benefit in running some services in separate containers. Splitting up application into multiple containers allows us to better isolate and maintain key services, providing a more modular and secure approach to application management. Each service can be packaged with the operating environment and tools it specifically needs to run, and each service can be limited to the minimum system resources necessary to perform its task. The benefits of multi-container applications compound as the complexity of the application grows. Because each service can be updated independently, larger applications can be developed and maintained by separate teams, each free to work in a way that best supports their service. There are two containers: Nginx container for accepting external requests; Flask container for REST API endpoints. The Building and running process under UNIX (Linux or MacOS) with Docker Compose is as follows:

- Clone from UkrVectōrēs GitHub repository:

  $ git clone https://github.com/malakhovks/docsim.git
- Put your model(s) in 'server/models' sub-directory of UkrVectōrēs;
- Change configuration files, stating the paths to your models;
- Optionally change other settings via 'config.models.simple.json' file;
- Build and run a Docker image from Dockerfile-nginx, Dockerfile-docsim with docker-compose:

  $ docker-compose up -d
- Open Web browser and navigate to UkrVectōrēs Web application.

You can preview a completed latest version of the UkrVectōrēs Web service with several efficient models ("WhiteBook" model, trained using a dataset – the book «The White Book of Physical and Rehabilitation Medicine» Ukrainian version; "Fiction" model, trained using a dataset – Ukrainian fiction literature) for Ukrainian here: without ssl certificate – http://ukrvectores.ml/; with self-singed ssl certificate https://test.ulif.org.ua:51043.

UkrVectōrēs features

Immediately after that you can interact with the loaded model via UkrVectōrēs Web application. From a user's point of view, UkrVectōrēs is a kind of "cognitive-semantic calculator" which operates on relations between entities (words and terms) in distributional models. The online toolkit UkrVectōrēs covers the following elements of distributional analysis:

- ✓ calculate semantic similarity between pairs of words;
- ✓ find words semantically closest to the query word;
- ✓ apply simple algebraic operations to word vectors (addition, subtraction, finding average vector for a group of words and distances to this average value);
- ✓ draw semantic maps (using the TensorFlow' TensorBoard Embedding Projector) of relations between input words (it is useful to explore clusters and oppositions, or to test your hypotheses about them);
- ✓ get the raw vectors (arrays of real values) and their visualizations for words in the chosen model;
- ✓ download default models;
- ✓ use other prognostic models distributive semantics freely distributed, by adjusting the configuration file.

Another feature of the service is the possibility to use several models simultaneously. If several models are enumerated in the configuration file, the UkrVectōrēs daemon loads all of them. UkrVectōrēs graphical user interface is multi-lingual (English and Ukrainian). Extending the interface with another language is implemented with Angular internationalization (i18n) feature.

Results and Discussion

Natural Language Processing is the automatic processing of text written in natural (human) languages (English, French, Ukrainian, etc.), as opposed to artificial languages such as programming languages, to try to “understand” it. It is also known as Computational Linguistics or Natural Language Engineering. NLP encompasses a wide range of tasks, from low-level tasks, such as segmenting text into sentences and words, to high-level complex applications such as semantic annotation (semantic similarity between words and multi-word expressions) and data/knowledge mining (automatic creation of thesauri and bilingual dictionaries, creating semantic maps of different subject domains). Distributional word embeddings are now arguably the most popular way to computationally handle lexical semantics. The UkrVectōrēs – a free and open source toolkit which aim is to quickly deploy web services handling pre-trained distributional semantic models (static word embeddings). It completes the pipeline between training such models and sharing the results to the general public. The UkrVectōrēs provides all the necessary routines for online access, sharing and querying to the trained models via modern Web application, built with Angular, Flask and Docker platforms.

Conclusion

The main aim of UkrVectōrēs is to quickly deploy web services processing queries to distributional semantic models, independently of a particular language. It allows to make complex linguistic resources available to wide audience in almost no time. Additionally, another major cases for using UkrVectōrēs Web service is to help in the study and discovery the of arbitrary domains. In particular, using the “WhiteBook” distributional semantic model (trained using a dataset – the book «The White Book of Physical and Rehabilitation Medicine» Ukrainian version) is intended for training specialists in physical and rehabilitation medicine. The authors plan to continue adding new features aiming at better understanding of embedding models, including sentence similarities, documents similarities and analysis of correlations between different kind of models (to changeover

from distributed word representations to distributed term representations or term embeddings).

Bibliography

1. [docsim, 2021] UkrVectōrēs/docsim – an NLU-powered tool for knowledge discovery, classification, diagnostics and prediction. Entities similarity tool. https://github.com/malakhovks/docsim.
2. [Turney, 2010] Turney P.D., Pantel P., et al. From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res., 2010 37(1), pp. 141–188.
3. [Mikolov, 2013] Mikolov T., Sutskever I., Chen K., Corrado G.S., Dean J. Distributed representations of words and phrases and their compositionality. Adv. Neural Inf. Process. Syst., 2013 26, pp. 3111–3119.
4. [Kutuzov, 2016] Kutuzov A., Kuzmenko E. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, vol 661. Springer, Cham.
5. [Baroni, 2016] Baroni M., Dinu G., Kruszewski G. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, vol. 1, 2014.
6. [WebVectors, 2021] WebVectors: word embeddings online. http://vectors.nlpl.eu/explore/embeddings/en/.
7. [RusVectores, 2021] RusVectōrēs: word embeddings for Russian online https://rusvectores.org/en/.
8. [Gensim, 2021] Gensim – Topic modelling for humans. https://radimrehurek.com/gensim.
9. [Docker, 2021] Overview of Docker Compose tool. https://docs.docker.com/compose/.

Authors' Information

Vitalii Velychko *– PhD, Senior researcher. V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, e-mail: aduisukr@gmail.com*

Major Fields of Scientific Research*: Artificial Intelligence, NLP*

Kyrylo Malakhov *– Research fellow in IT. V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, e-mail: malakhovks@nas.gov.ua*

Major Fields of Scientific Research: *Artificial Intelligence, Computational linguistics, Distributional semantics*

Oleksandr Shchurov – *Junior researcher. V.M. Glushkov Institute of Cybernetics of NAS of Ukraine, e-mail: alexlug89@gmail.com*

Major Fields of Scientific Research: *Artificial Intelligence, NLP*

MULTILINGUAL TEXT CLASSIFICATION USING TRANSFORMER APPROACH

Larysa Chala, Mykola Dudnik, Mariia Sokolovskaya

***Abstract**: Recently, experts have been developing methods for processing multilingual texts using machine learning methods. This article proposes a method that allows one model to support 100 languages for processing and analyzing text streams. The results of an experiment on text classification using this model are presented, which confirm its effectiveness.*

***Keywords:** Deep Learning, Artificial Intelligence, Natural Language Processing.*

***ITHEA Keywords:** D.2.0 Genera*

Introduction

As you may know there are two problems in the classification of multilingual texts: the first one is a linear growth of computational costs with an increasing of number of languages; the second one is the lack of ability in the model to transfer knowledge between different languages (as a rule, they are formed only for one language and cannot be directly applied to another language). At the same time, many different languages are used on the network (for example, Facebook users publish content in more than 160 languages).

Existing solutions to these problems are based on building the model, as well as the entire architecture associated with it, for each language that needs to be worked with. The approach proposed in this article assumes that one model can be used for 100 languages, which simplifies the server architecture of the word processing system and provides further support for this solution.

Basic definitions

As you may know there are two problems in the classification of multilingual texts: the first one is a linear growth of computational costs with an increasing of number of languages; the second one is the lack of ability in the model to transfer knowledge between different languages (as a rule, they are formed only for one language and

cannot be directly applied to another language). At the same time, many different languages are used on the network (for example, Facebook users publish content in more than 160 languages).

Existing solutions to these problems are based on building the model, as well as the entire architecture associated with it, for each language that needs to be worked with. The approach proposed in this article assumes that one model can be used for 100 languages, which simplifies the server architecture of the word processing system and provides further support for this solution.

Proposed experiment and its results

In the course of research, a news headline classifier was created based on the XLM-R. This classifier is built using LSTM layers and is a deep neural network. The public kaggle dataset [4] was used as the training dataset.

The textual dataset on which the XLM-R is trained is not marked up, it is a set of texts and nothing more. The use of languages with limited resources means that there are practically no texts for the preliminary training of language models. In our experiment, we used a tagged dataset consisting of text pairs “news headlines - news categories”.

The results of the experiment showed that when training a classifier in one language (using XLM-R for vectorization), it will also work well with other languages. In our case, the model trained in English showed an accuracy of 0.87 on a test set with texts in French (Fig. 1).

```
[ ]  model.evaluate(eval_embeddings, y_test)

     16/16 [==============================] - 1s 54ms/step - loss: 0.3830 - accuracy: 0.8680
     [0.38295432925224304, 0.8679999709129333]
```

Figure 1. Results of model evaluation

Conclusion

Using the XLM-R model for text vectorization is an important step towards building multilingual text classifiers.

Possible applications using XLM-R technology include serving highly accurate rules-breaking inappropriate content detection models across a wide range of languages. The proposed approach makes it possible to implement the principle: "one model for many languages", in contrast to the existing one, which creates one model for each language [5].

"One model for many languages" is a new approach that became available after the emergence of the transformer architecture and the mechanism of self-attention embedded in them. This approach allows using one model for text vectorization for all languages that participated in its training.

To sum up, this approach will simplify the further launch of high-performance products in several languages at the same time. In addition, the approach discussed will improve the performance of multilingual models and systems that use self-taught methods to better understand languages with limited resources.

Bibliography

1. [Alammar, 2018] Alammar, Jay. The Illustrated Transformer [Blog post]. Visualizing machine learning one concept at a time. June 27, 2018. https://jalammar.github.io/illustrated-transformer/ [Accessed 3 March 2020].
2. [Branden, 2020] Branden, Chan. XLM-RoBERTa: The alternative for non-english NLP [Blog post]. deepset-ai, Bridging the gap between NLP research & industry. February 17, 2020. https://medium.com/deepset-ai/xlm-roberta-the-multilingual-alternative-for-non-english-nlp-cf0b889ccbbf [Accessed 5 March 2020].
3. [Rohan, 2020] Rohan, Jagtab. XLM-RoBERTa: Unsupervised Cross-lingual Representation Learning at Scale [Blog post]. The Startup, Build Something Awesome. October 18, 2020. https://medium.com/swlh/xlm-roberta-unsupervised-cross-lingual-representation-learning-at-scale-f9b144d452cb [Accessed 28 November 2020].
4. [Rishabh, 2019] Rishabh, Misra. News Category Dataset. Kaggle. https://www.kaggle.com/rmisra/news-category-dataset [Accessed 3 April 2020].
5. [Paul, 2020] Paul, Barba. Challenges in Developing Multilingual Language Models in Natural Language Processing (NLP) [Blog post]. Towards Data Science, A Medium publication sharing concepts, ideas and codes. October 23, 2020. https://towardsdatascience.com/challenges-in-developing-multilingual-language-models-in-natural-language-processing-nlp-f3b2bed64739 [Accessed 3 April 2020].

6. [Martin and Ondřej, 2018] Martin, P., Ondřej, B., Training Tips for the Transformer Model. Journal The Prague Bulletin of Mathematical Linguistics 110, April 2018, pp. 43-70. doi: 10.2478/pralin-2018-0002 https://arxiv.org/abs/1804.00247.

Authors' Information

Larysa Chala *– Kharkiv National University of Radio Electronics, Ukraine. Department of Artificial Intelligence. Associate Professor, PhD; e-mail: larysa.chala@nure.ua*

Major Fields of Scientific Research*: Natural Language Processing, Software Programming, Machine Learning*

Mykola Dudnik *– Kharkiv National University of Radio Electronics, Ukraine. Department of Artificial Intelligence. Undergraduate Student; e-mail: mykola.dudnyk@nure.ua*

Major Fields of Scientific Research: *Natural Language Processing, Software Programming, Machine Learning*

Mariia Sokolovskaya *– Kharkiv National University of Radio Electronics, Ukraine. Department of Artificial Intelligence. Undergraduate Student; e-mail: mariia.sokolovska@nure.ua*

Major Fields of Scientific Research: *Natural Language Processing, Software Programming, Machine Learning*

EPIDEMIC EFFECTS IN NETWORK INDUSTRIES

Gorbachuk Vasyl, Dunaievskyi Maksym, Syrku Andrii

Abstract*: Modern research in the field of global cellular telephony says that technology reveals the effects of an established base in acceptance (of a new technology) and that cellular diffusion varies between countries and groups of countries due to technological, socioeconomic and regulatory factors that influence the diffusion process. The effects of an established base may come from the extra effects of social dissemination, such as social learning through uncertainty and regulation pressures. With a severe decrease in the intensity of use, the epidemic effect is largely absent or at least overshadowed by other factors, notably consumer heterogeneity. Hence it follows that the driving forces behind diffusion are constantly falling prices and / or increasing quality of the network services.*

Keywords: *network products, network externalities, monopoly, social welfare.*

ITHEA Keywords*: computer-communication networks, network architecture and design, network operations, distributed systems, performance of systems.*

Introduction

As a rule, existing models cannot empirically differentiate periods of rapid diffusion in network industry, which are common in most industrialized countries, from critical mass that does not rely on price reductions or exogenously changing technologies [Gorbachuk, 2013]. Attempts have been made to exclude another potential cause of endogenous diffusion − epidemic effects, in which the rate of penetration increases with constant intensity of use.

The problem statement

The problem statement is to investigate the possibility of an endogenous and almost discontinuous diffusion path (new network technology) based on the effect of the established base (available number of technology users) [Chikrii, Gorbachuk et al., 2012].

Unanswered question and purpose

An unanswered question is the properties of models which allow to find the critical mass of technology's spread and propose appropriate regulation. The purpose of the work is to strictly justify the regulation in the field of telecommunications [Gorbachuk et al., 2013].

Main results

The main results come from an economy model consisting of two identical groups of consumers who want to join a certain telecommunications service (for example get a phone call): η consumers of type H highly appreciate joining this service, a $\eta > 0$ consumers of type L – low [Shy, 2001].

Denote $p \geq 0$ as fee for connecting to this service. If the actual number of connected (to the service) consumers is q, then the utility function of the consumer of type L equals

$$U_L = \begin{cases} q - p, \text{ consumer connected} \\ 0, \text{ consumer not connected} \end{cases} \quad (1)$$

and the utility function of the consumer of type H –

$$U_H = \begin{cases} \alpha q - p, \text{ consumer connected} \\ 0, \text{ consumer not connected} \end{cases} \quad (2)$$

where $1 < \alpha$ measures the importance of service to the consumer of type H : $U_H > U_L$.While building demand function for telecommunication services we will use the assumption about the absence of coordination failure: if every consumer of group (H or L) wins from subscription to the service, then all customers of the group will subscribe to the service.

Critical mass is the minimum number of users required to ensure that at least that number of consumers will have not negative subscription benefit. In order to organize a party or weekend trip the organizer must convince potential participants that it will be for sure attended by a certain minimum number of people, which will in fact mean even more entrants due to the growing network effects. In telecommunications, critical mass is

always a function of market price: an increase in price will mean an increase in critical mass, and a decrease in market price will reduce critical mass, since under a lower price consumers will be satisfied with a smaller network size. With this concept in mind, a former student of one of the authors developed the Djuice project, which earned Kyivstar more than two billion U.S. dollars over five years.

By the 1980s, most countries had a monopoly market structure in the field of telecommunications. For example, in Ukraine such a monopoly firm was called Ukrtelecom, in Estonia – Estonia Telecom. In Israel, until the 1960s, a similar firm also provided postal services.

During the 1980s, governments began to realize that monopoly telecommunications market structures, which were considered natural monopolies, distorted industry markets. The major event leading to competition in this area was the division of AT&T USA into 7 regional telephone companies in 1982, as well as the creation of MCI and SPRINT as major competitors in the intercity and international markets. In the 1980s, regulators were discussing three basic issues: knowing that many users (of type H) are already connected to an existing monopoly telecommunications provider, can social well-being be improved by allowing the new operator to connect other consumers (of type L) to the network; whether the new operator entering the market will have profit; when the entry of new providers is socially desirable, how to avoid dragging of the existing monopoly into predatory pricing to attract more consumers and thus narrow the potential market for a new firm [Gorbachuk, 2014].

Proposition 1. Entry into the telecommunications industry increases the utility of already connected customers and leaves unchanged the utility of newly connected consumers, increases the profits of a new firm and leaves the profits of an existing firm unchanged.

When, in addition to the connection service market, the market for the flow of services (telephone calls) after the connection of consumers of type L at the entry of a new firm is also taken into account, the existing firm will suffer a decrease in profits, but social welfare will increase as a result of the price decrease. The analysis of the

telecommunications industry is based on the fact that the utility of the consumer from the communication service is enhanced when others join the service. Let us denote $x \in [0,1]$ the reluctance of the consumer to pay for the service: the greater the value of x, the less willing to pay the consumer is. The utility of consumer of type x will define as

$$U_x = \begin{cases} (1-x)q_e - p, \text{consumer connected} \\ 0, \text{consumer not connected} \end{cases} \quad (3)$$

where q_e – expected number of consumers who join this network. As this consumer utility increases with increasing q_e, have network externalities.

Proposition 2. A monopoly telephone company maximizes its profits by setting its cost per connection so that the number of users exceeds half of all existing customers and that unconnected consumers exist.

Proposition 3. With increasing of the total consumers number η the monopoly price and the utility of the connected users increase proportionally, monopoly profit increases quadratically.

Conclusion

The main conclusion is that the entry of a new firm into the telecommunications industry increases the utility of already connected consumers and leaves unchanged the utility of newly connected consumers. A monopoly network company maximizes its profits by setting its price per connection so that the number of users exceeds half of all customers available, but not all customers subscribe to the network service. Questions that need further investigation are: whether exogenous changes are responsible for total diffusion (without critical mass); whether there is an element of endogenous diffusion.

Bibliography

1. [Shy, 2002] Shy O. *The economics of network industries*. Cambridge: Cambridge University Press, 2001. 315 p.
2. [Chikrii, Gorbachuk et al., 2012] Chikrii A., Denisova N., Gorbachuk V., Gromaszek K., Krivonos Y., Lytvynenko V., Matychyn I., Osypenko V., Smailova S., Wojcik W.

Current problems in information and computational technologies. V. 2. W. Wojcik, J. Sikora (eds.) Lublin, Poland: Politechnika Lubelska, 2012. 196 p.

3. [Gorbachuk, 2013] Gorbachuk V. The ways of coping with market imperfections in the telecommunications industry. *Odesa National University Herald. Economy*. 2013. V. 18. № 3 (1). P. 84-87. (In Ukrainian).
4. [Gorbachuk et al., 2013] Gorbachuk V., Chumakov B. A computable critical mass on market of new network product. *Transport systems and logistics*. Chisinau, Moldova: Academia de Transporturi, Informatică şi Comunicaţii, 2013. P. 272–281.
5. [Gorbachuk, 2014] Gorbachuk V.M. Equilibria of international public goods. *Education and science and their role in social and industrial progress of society.* Kyiv: Alexander von Humboldt Foundation, 2014. P. 13–14.

Authors' Information

Gorbachuk Vasyl *− V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, DSc (Physics and Mathematics), Head of Department, e-mail: GorbachukVasyl@netscape.net*

Major Fields of Scientific Research: *Equilibria of Decentralized Systems*

Dunaievskyi Maksym *− V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, MSc (Finance), PhD Student, e-mail: MaxDunaievskyi@gmail.com*

Major Fields of Scientific Research*: Data Science Techniques*

Syrku Andrii *− V.M.Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, MSc (Engineering), PhD Student, e-mail: saan@ukr.net*

Major Fields of Scientific Research: *Smart Sensors, Signal Processing*

Тези доповідей

Міжнародної науково-практичної конференції

Інженерія програмного забезпечення

Технічні редактори – Markov K., Чебанюк О.В.

Матеріали друкуються за редакцією авторів.

Підписано до друку.