SOFTWARE ENGINEERING AND BIG DATA SOFTWARE

Nikolay Sydorov, Nika Sydorova

Software engineering is a mature industry of human activity focused on the creation, deployment, marketing and maintenance of software. The fundamental concepts that engineering explores and uses are life cycle model; three main components of life cycle phases - products, processes and resources; engineering and methodologies for creating, deployment and maintaining software. Software engineering serves the tasks of creating and maintaining software in other domains, which are called application domains. Information technology is one such well-known application domain. The basis of this domain is data. Information systems are being implemented in an organization to improve its effectiveness and efficiency. The functionality of information systems has grown dramatically when big data began to be used. This growth has led to the emergence of a wide variety of software-intensive big data information systems. At the same time, the role and importance of software engineering for solving the problems of this application domain has only intensified. Modern possibilities of software engineering are shown. The aspects of interaction between software engineering and big data systems are analyzed. The topics for the study of big data software ecosystems and big data systems are outlined.

Keywords. Big data, big data software, software engineering, big data software ecosystem, big data system of systems

Introduction

Software engineering is a mature industry of human activity focused on the creation, deployment, marketing and maintenance of software. The fundamental concepts that engineering explores and uses are life cycle model; three main components of life cycle phases - products, processes and resources; engineering and methodologies for creating, deployment and maintaining software. Software is the foundation of technological advances that lead to new high performance products. As the functionality of products grows, so does the need to efficiently and correctly create and maintain the complex software that enables this growth. Therefore, in addition to solving its own problems, software engineering serves the solution of the problems of creating and maintaining software in other domains, which are called application domain. Information technology is one such well-known application domain. The basis of this domain is data. Information systems are being implemented in an organization to improve its effectiveness and efficiency. The functionality of information systems has grown dramatically when big data began to be used. This growth has led to the emergence of a wide variety of software-intensive big data information systems. At the same time, the role and importance of software engineering for solving the problems of this application domain has only intensified. Modern possibilities of software engineering are shown. The aspects of interaction between software engineering and big data systems are analyzed. The topics for the study of big data software ecosystems and big data systems are outlined.

Keywords. Big data, big data software, software engineering, big data software ecosystem, big data system of systems

The software engineering and big data software

The software engineering. Consider software engineering and its relationship with big data software in aspect of their influence on each other. Software engineering, having appeared more than fifty years ago, is now rather mature industry of human activity aimed at creating, deployment, marketing and maintaining software [1]. The credo
of software engineers is expressed in the view of software products as reality objects, which are created, act, changed and destroyed by the implementation of appropriate processes. And, these processes can be controlled and they can be managed. This, successively, led to the concept of modular software design, starting with the method of structured programming, continuing in block-oriented, modular and object-oriented programming languages, and ending with a component approach to software development. This also led to modern software life cycle models, when two phases of the Cod-and-fix model were replaced in the waterfall model on real-life processes: vertical ones - domain analysis, re-

Data ecosystems are considered in [16]. The transition from a data ecosystem to implementing the project and requirements management, de-

Big data can significantly change both approaches to research and the

relationships is growing. In its turn it affects the health of an ecosystem, since the relationship between ecosystem

data; data processing - text analytics, business intelligence, data visualization and statistical analysis, predictive analy-

The challenges for software engineering researchers

For software engineering researchers, new themes from big data application domains are likely to be big data

The big data for challenges of software engineering. The methods and tools used in the big data domain can be used in solving software engineering problems. Using analytics for software engineering, data mining software repositories, and visual analytics for software engineering, are considered in [11]. The application of machine learning algorithms for software analysis [14] and artificial intelligence in the phases of the software life cycle are considered in [15].

The big data software ecosystems. Big data can significantly change both approaches to research and the

architecture of software ecosystems. The big data software ecosystems will be more heterogeneous both in terms of actors, roles and strategies, as well as products and platforms. This is a consequence of the presence of such characteristics of the big data as Variety, Volatility, and Value. In the big data software ecosystem is involved making use of a wide variety of technologies and tools, such as data management - acquisition, storage, management and retrieval of data; data processing - text analytics, business intelligence, data visualization and statistical analysis, predictive analy-

This will entail the presence of different actors, products and platforms in the ecosystem. In software ecosystems, the variety of these entities is usually limited. In the big data software ecosystem, the complexity of the structure and relationships is growing. In its turn it affects the health of an ecosystem, since the relationship between ecosystem health and three factors which are influenced by the health of an ecosystem is known, but not studied: growth, maturity

The big data ecosystems. Data ecosystems are considered in [16]. The transition from a data ecosystem to the big data ecosystems is considered in [17] and the big data architecture framework and big data lifecycle in big data ecosystem are proposed. Big data is growing exponentially and flows are almost endless. It requires distributed, parallel and scalable data storage and processing systems. To cope with such huge constantly changing data using the big data ecosystems will be productiveness. In [18], the actors of such ecosystem are introduced, and in [19], the modules of the big data ecosystems are introduced. In [17], the task of creating a Common Body of Knowledge for big data to provide a basis for a consistent curriculum development is proposed.

The modules that formed the big data ecosystems are software-intensive systems and can be considered as modules of big data software ecosystems.

The big data software ecosystems. Big data can significantly change both approaches to research and the

architecture of software ecosystems. The big data software ecosystems will be more heterogeneous both in terms of actors, roles and strategies, as well as products and platforms. This is a consequence of the presence of such characteristics of the big data as Variety, Volatility, and Value. In the big data software ecosystem is involved making use of a wide variety of technologies and tools, such as data management - acquisition, storage, management and retrieval of data; data processing - text analytics, business intelligence, data visualization and statistical analysis, predictive analy-

This will entail the presence of different actors, products and platforms in the ecosystem. In software ecosystems, the variety of these entities is usually limited. In the big data software ecosystem, the complexity of the structure and relationships is growing. In its turn it affects the health of an ecosystem, since the relationship between ecosystem health and three factors which are influenced by the health of an ecosystem is known, but not studied: growth, maturity
and popularity. Problems can also arise with ownership of a big data. If it is a keystone player, then it will probably be a platform leader and perform orchestration. Therefore, the study of the big data software ecosystems will require the development of new methods and tools of software engineering.

**The big data system as system of systems.** System of systems is a large-scale integrated system that is heterogeneous and independently operable on its own, but is networked together for a common goal. [20]. Modern system of systems is characterized in terms of five main properties, sometimes referred to as the acronym “OMGEE” [21]: Operational Independence, Managerial Independence, Geographic Distribution, Evolutionary Development, and Emergent Behavior.

One system will definitely be included in the core of big data system of systems - big Data Management System. It provides the implementation of the metadata, data collection, classification, indexing, storage, search, transform functions, representing big data at the physical and application levels [17]. The big Data Management System, creates a common big data space for systems included in the big data system of systems. The composition of other systems may vary and depends on the lead system integrator. For example, it may include an analytics (modeling, prediction) system that performs analytics operations of two categories of analysis - particularly direct analysis and exploratory analysis which requires real time response [19]; presentation system performing data delivery, visualization; situation awareness system, which involves collecting, aggregating, and interpreting information in order to know what is happening in the environment. In addition, the system of systems may include big data systems of application domains, for example, banking and financial services, energy, healthcare, media, and education among others. To create and maintain a system of systems, special engineering is used, which is different from the traditional one [20]. Obviously, when a big data system of systems is created, it is necessary to take into account the characteristics of big data.

**Conclusion**

The big data systems are software-intensive systems. Therefore, to create and maintain software of big data systems, it is advisable to use the methods and tools of software engineering. For software engineering, big data systems are a part of the application domain. The article shows the modern possibilities of software engineering and the interconnection of software engineering and big data software. On the one hand, it is considered which challenges of big data software are already being solved by applying the methods of software engineering, and on the other hand, which challenges of software engineering can be solved by applying the methods of processing of big data. This will be useful for both application domain researchers and software engineers. Finally, we have indicated new topics for software engineering researchers. These are the big data software ecosystems and the big data system as a system of systems. They may be objects of future research. The author did not make a systematic mapping study of the literature on the topic of the article, but took into account the main works in this area.

**References**

1. Report on a conference sponsored by the NATO science committee, Garmisch, Germany, 7th to 11th October 1968, Editors: Peter Naur and Brian Randell.
2. Sidorov M.O. Software engineering lecture course. - Kiev.: National Aviation University, 2007.-140p
4. Sydorov N., Programming Style as an Artefact of a Software Artefacts Ecosystem, Advances in Computer Science for Engineering and Education IV, 2021, pp 244–255
16. Data Ecosystems, ISST-Report, Editors Prof. Dr.-Ing. Boris Otto, Prof. Dr. Jakob Rehof, Germany, October 2019

Received 19.08.2022
Методи і засоби програмної інженерії

Про авторів:

Сидоров Микола Олександрович,
доктор технічних наук,
професор.
Кількість наукових публікацій у
українських виданнях – 140.
Кількість наукових публікацій у
зарубіжних виданнях – 22.
http://orcid.org/0000-0002-3794-780X

Сидорова Ніка Миколаївна,
кандидат технічних наук,
Кількість наукових публікацій у
українських виданнях – 40.
Кількість наукових публікацій у
зарубіжних виданнях – 6
ORCID 0000-0002-2989-3637

Місце роботи авторів:

Сидоров Микола Олександрович,
Національний Технічний Університет України «Київський політехнічний
інститут імені Ігоря Сікорського»,
факультет інформатики та обчислювальної техніки,
кафедра інформатики та програмної інженерії, ІПІ, професор.
02000, Київ, вул. Політехнічна, 41.
Моб. тел.: 067 7980361.
E-mail: nyksydorov@gmail/com

Сидорова Ніка Миколаївна,
Приватне акціонерне товариство «ОТІ»
м. Київ, вул. Феодори Пушкіної 30/32, 03115
Заступник генерального директора з питань зовнішніх відносин
E-mail: nika.sidorova@gmail.com

Прізвища та ім’я авторів і назва доповіді англійською мовою:
Sydorov N.O., Sydorova N.M.
Software engineering and big data software

Прізвища та ім’я авторів і назва доповіді українською мовою:
Сидоров М. О., Сидорова Н. М.
Інженерія програмного забезпечення та програмне забезпечення великих даних