UDC 004.413 https://doi.org/10.15407/pp2022.03-04.069

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 of engineering 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 domains. Information technology is a 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 system of 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 tasks 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 implemented in an organization in order 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. The big data software is software of big data management systems and software of big data processing and presentation systems. At the same time, the role and importance of software engineering for solving the problems of this application domain has only intensified.

Usually, the interaction of software engineering with the application domain is carried out by a "problem-solution" principle. In the application domain, a problem is formulated to be solved, and in the domain of software engineering, a solution is sought, and if there is none, a new one is created.

If the application domain is software intensive, then postgraduate education programs for application domain specialists can be created to interact with specialists in the software engineering domain. Usually, these are master's programs of universities. If the request for education of specialists in the applied domain is sustainable in demand, then the Common Body of Knowledge can be created as a basis for a consistent curriculum development. Thus, the big data information systems and technologies domain, for software engineering, is similar to other application domains, nothing more.

However, from the point of view of software engineering, it is impossible not to pay attention to the diversity of big data systems and, consequently, the diversity of software of big data systems. This diversity is created due to the presence of characteristics of big data such as Variety, Volatility, and Value. The diversity of big data systems provides the presence of big data software ecosystems. In addition, a variety of big data systems, which are used to achieve an emergent goal, can be a form of a big data system as system of systems. Ecosystems of this type and big data system of systems will be interesting for software engineering.

The article analyzes aspects of the interaction between software engineering and big data software, makes an assumption about the existence, and outlines topics for the study of big data software ecosystems and 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, requirements specification, and design, construction, testing and debugging; horizontal - management, quality assurance, verification and validation, documentation [2]. This contributed to the emergence of new forms of software products, methods, tools and professions. The result of solving a software engineering problem is always a product that has a customer. A software engineering product is created in the context of a life cycle and must meet the requirements specific to any engineering (product construction, quality, standards, documentation, maintenance, environmental impact and economics). In this regard, in software engineering there is a number of engineering types - forward, for creating products, backward (reverse), to extract design and implementation information about legacy products, reengineering for changing products, search-based engineering for finding COTS, empirical, for researching products. The latter uses a product, process, and resource metrics. Software culture and related tools (CMMI, PCMM) are also important [3]. The impact on the environment is studied in the context of the concept of sustainable development in the form of lean and green software development, as well as software reusable development. The evolution of products is explored in the form of a product line. The interaction and mutual influence of products, as well as software assets and artefacts, is explored in the form of software ecosystems and software artefacts ecosystems [4]. In the context of globalization, software engineering products are implemented and researched as a system of systems. Software engineering products and the processes for creating and maintaining them are costly. There is software economics with effective methods, models and tools for the cost estimating of products [5, 6].

SWEBOK (Software Engineering Body of Knowledge) and Curriculum Guidelines have been developed for undergraduate and graduate education of software engineers [7]. Training of software engineers for application domains is carried out at the level of master's postgraduate education.

The software engineering and challenges of big data software. Guided by the "problem-solution" principle, big data application domain researchers formulate challenges of big data software for software engineering. Problems are associated with 10V big data characteristics and are formulated as requirements for requirements specification and architecture design. It is indicated that such processes as architecture design, software quality assurance are less studied [8]. In [9], the problems of implementing the project and requirements management, deployment and operations processes are added to these problems. In [10], the problems of using data-centric software development to implement big data analytics and the problems of testing and debugging for big data analytics are formulated. The paper [11] analyzes the possibilities of software engineering for the implementation of the big data systems. The problems of developing systems in application domains with big data, for example, military, aviation industry, banking and financial industry, are considered in [12, 13].

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 challenges for software engineering researchers

For software engineering researchers, new themes from big data application domains are likely to be big data software ecosystems and software of big data system as system of systems.

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 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 analysis. 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 curse.- Kiev.: National Aviation University, 2007.-140p
3. Wiegers K, Creating a Software Engineering Culture, Dorset House, 2011, 384 p
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
5. Boehm, Barry. Software Cost Estimation with COCOMO II, Prentice Hall PTR, 2000.
6. Royce, Bittner, Perrow, The Economics of Software Development, Addison-Wesley, 2009.
7. SWEBOK Guide V3.0б 2014, IEEE Society
8. Davoudian A Liu M, Big Data Systems: A Software Engineering Perspective, ACM Computing Surveys, Vol. 1, No. 1, 2019, P 1-43
9. Hummel O, Eichelberger H, A Collection of Software Engineering Challenges for Big Data System Development, Euromicro SEAA 2018, Prague
10. Kim, M, Software Engineering for Data Analytics, IEEE Software, july/august 2020, P 36-42
11. Arndt T, Big Data and software engineering: prospects for mutual enrichment, Iran Journal of Computer Science, Springer International Publishing AG, 2017
12. Punia S, Kumar M, Domains and methodologies for big data project in software engineering, International Journal of Scientific & Engineering Research, Volume 8, Issue 8-2017, P 1525-1535
13. Kumar V.D, Alencar P, Software Engineering for Big Data Projects:Domains, Methodologies and Gaps, The proceedings of IEEE International Conference on Big Data, 5-8 Dec. 2016
14. Meinke K, Bennaceur A, Machine Learning for Software Engineering, ICSE '18 Companion, May 27-June 3, 2018, Gothenburg, Sweden
15. Barenkamp M, Rebstadt J, Thomas O, Applications of AI in classical software engineering, https://doi.org/10.1186/s42467-020-00005-4
16. Data Ecosystems, ISST-Report, Editors Prof. Dr.-Ing. Boris Otto, Prof. Dr. Jakob Rehof, Germany, October 2019
17. Demchenko Y, .Membrey P, C. de Laat, Defining Architecture Components of the Big Data Ecosystem, The proceedings of International Symposium on Collaborative Technologies and Systems, 19-23 May 2014, P 1-3
18. Rad B, Ataei P, The big data Ecosystem and its Environs, IJCSNS International Journal of Computer Science and Network Security, VOL.17 No.3, March 2017
19. Hamid A, A Functional View of Big Data Ecosystem, International Journal of Computer Trends and Technology (IJCTT) – Volume 68 Issue 4 – April 2020, P 233-237
20. Systems of systems engineering: innovations for the 21st century / edited by Mo Jamshidi. John Wiley & Sons, 2009, p 588
21. Maier, M. W. Architecting Principles for Systems-of-Systems. In INCOSE 1996 Sixth annual International Symposium of the International Council on Systems Engineering. INCOSE, 1996

Received 19.08.2022

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 Пушиної 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