ISSN 2587-814X (print), ISSN 2587-8158 (online)
Russian version: ISSN 1998-0663 (print), ISSN 2587-8166 (online)

Vol. 17 No. 2 – 2023

BUSINESS INFORMATICS

HSE Scientific Journal

Publisher:
National Research University
Higher School of Economics

The journal is published quarterly

The journal is included
into the list of peer reviewed
scientific editions established
by the Supreme Certification
Commission of the Russian Federation

CONTENTS

A.S. Akopov

Modeling and optimization of strategies for making individual decisions in multi-agent socio-economic systems with the use of machine learning 7

N.V. Dneprovskaya, I.V. Shevtsova

A knowledge management system in the strategic development of universities 20

M.G. Matveev, N.A. Aleynikova, M.D. Titova

Decision support technology for a seller on a marketplace in a competitive environment 41

A.I. Shaidullin

The problem of interpretation, differentiation and classification of digital products 55

I.E. Nikolaev

An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus 71

F. Wang

The present and future of the digital transformation of real estate: A systematic review of smart real estate 85

Address:
26-28, Shabolovka Street,
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 *28509
http://bijournal.hse.ru
E-mail: bijournal@hse.ru

Circulation:
English version – 100 copies,
Russian version – 100 copies,
online versions in English and Russian –
open access

Printed in HSE Printing House
44, build.2, Izmaylovskoye Shosse,
Moscow, Russia

ABOUT THE JOURNAL

Business Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by National Research University Higher School of Economics (HSE), Moscow, Russian Federation. The journal is administered by HSE Graduate School of Business. The journal is issued quarterly, in English and Russian.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the following areas: modeling of social and economic systems, digital transformation of business, innovation management, information systems and technologies in business, data analysis and business intelligence systems, mathematical methods and algorithms of business informatics, business processes modeling and analysis, decision support in management.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation.

The journal is included into Scopus, Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index on the Web of Science platform (RSCI), EBSCO.

The journal is distributed both in printed and electronic forms.

EDITORIAL BOARD

ABOUT THE HIGHER SCHOOL OF ECONOMICS

Consistently ranked as one of Russia's top universities, the Higher School of Economics (HSE) is a leader in Russian education and one of the preeminent economics and social sciences universities in Eastern Europe and Eurasia.

Having rapidly grown into a well-renowned research university over two decades, HSE sets itself apart with its international presence and cooperation.

Our faculty, researchers, and students represent over 50 countries, and are dedicated to maintaining the highest academic standards. Our newly adopted structural reforms support both HSE's drive to internationalize and the groundbreaking research of our faculty, researchers, and students.

Now a dynamic university with four campuses, HSE is a leader in combining Russian educational traditions with the best international teaching and research practices. HSE offers outstanding educational programs from secondary school to doctoral studies, with top departments and research centers in a number of international fields.

Since 2013, HSE has been a member of the 5-100 Russian Academic Excellence Project, a highly selective government program aimed at boosting the international competitiveness of Russian universities.

ABOUT THE GRADUATE SCHOOL OF BUSINESS

HSE Graduate School of Business was created on September 1, 2020. The School will become a priority partner for leading Russian companies in the development of their personnel and management technologies.

The world-leading model of a 'university business school' has been chosen for the Graduate School of Business. This foresees an integrated portfolio of programmes, ranging from Bachelor's to EMBA programmes, communities of experts and a vast network of research centres and laboratories for advanced management studies. Furthermore, HSE University's integrative approach will allow the Graduate School of Business to develop as an interdisciplinary institution. The advancement of the Graduate School of Business through synergies with other faculties and institutes will serve as a key source of its competitive advantage. Moreover, the evolution and development of the Business School's faculty involves the active engagement of three professional tracks at our University: research, practice-oriented and methodological.

What sets the Graduate School of Business apart is its focus on educating and developing globally competitive and socially responsible business leaders for Russia's emerging digital economy.

The School's educational model will focus on a project approach and other dynamic methods for skills training, integration of online and other digital technologies, as well as systematic internationalization of educational processes.

At its start, the Graduate School of Business will offer 22 Bachelor programmes (three of which will be fully taught in English) and over 200 retraining and continuing professional development programmes, serving over 9,000 students. In future, the integrated portfolio of academic and professional programmes will continue to expand with a particular emphasis on graduate programmes, which is in line with the principles guiding top business schools around the world. In addition, the School's top quality and all-encompassing Bachelor degrees will continue to make valuable contributions to the achievement of the Business School's goals and the development t of its business model.

The School's plans include the establishment of a National Resource Center, which will offer case studies based on the experience of Russian companies. In addition, the Business School will assist in the provision of up-to-date management training at other Russian universities. Furthermore, the Graduate School of Business will become one of the leaders in promoting Russian education.

The Graduate School of Business's unique ecosystem will be created through partnerships with leading global business schools, as well as in-depth cooperation with firms and companies during the entire life cycle of the school's programmes. The success criteria for the Business School include professional recognition thanks to the stellar careers of its graduates, its international programmes and institutional accreditations, as well as its presence on global business school rankings.

DOI: 10.17323/2587-814X.2023.2.7.19

Modeling and optimization of strategies for making individual decisions in multi-agent socio-economic systems with the use of machine learning

Andranik S. Akopov
E-mail: akopovas@umail.ru

Central Economics and Mathematics Institute, Russian Academy of Sciences
Address: 47, Nakhimovsky Prospect, Moscow 117418, Russia

Abstract

This article presents a new approach to modeling and optimizing individual decision-making strategies in multi-agent socio-economic systems (MSES). This approach is based on the synthesis of agent-based modeling methods, machine learning and genetic optimization algorithms. A procedure for the synthesis and training of artificial neural networks (ANNs) that simulate the functionality of MSES and provide an approximation of the values of its objective characteristics has been developed. The feature of the two-step procedure is the combined use of particle swarm optimization methods (to determine the optimal values of hyperparameters) and the Adam machine learning algorithm (to compute weight coefficients of the ANN). The use of such ANN-based surrogate models in parallel multi-agent real-coded genetic algorithms (MA-RCGA) makes it possible to raise substantially the time-efficiency of the evolutionary search for optimal solutions. We have conducted numerical experiments that confirm a significant improvement in the performance of MA-RCGA, which periodically uses the ANN-based surrogate-model to approximate the values of the objective and fitness functions. A software framework has been designed that consists of the original (reference) agent-based model of trade interactions, the ANN-based surrogate model and the MA-RCGA genetic algorithm. At the same time, the software libraries FLAME GPU, OpenNN (Open Neural Networks Library), etc., agent-based modeling and machine learning methods are used. The system we developed can be used by responsible managers.

Keywords: multi-agent socio-economic systems, particle swarm optimization, modeling random sales, machine learning, artificial neural networks, genetic optimization algorithms

Citation: Akopov A.S. (2023) Modeling and optimization of strategies for making individual decisions in multi-agent socio-economic systems with the use of machine learning. *Business Informatics*, vol. 17, no. 2, pp. 7–19. DOI: 10.17323/2587-814X.2023.2.7.19

Introduction

Currently, there is a growing interest in studying the behavior of multi-agent socio-economic systems (MSES) and developing decision support systems (DSS) using agent-based modeling (AOM), machine learning and heuristic (in particular, genetic) optimization algorithms.

Most modern DSS can be divided into two enlarged classes: effective control systems based on simulation, including optimization modeling, and expert decision support systems.

As examples of DSS belonging to the first type, we can highlight: a software package designed to manage the investment activities of a large oil company [1], a decision support system for environmental and economic planning [2], intelligent transport systems [3–5], etc.

The most well-known DSS of the second type include expert systems for making strategic decisions using the hierarchy analysis method [6, 7], systems designed to prioritize decisions in the management of IT projects [8], systems that support the ability to select the best alternatives in case of poorly structured initial data [9], etc.

In that work, the MSES management system of the first type, intended mainly for the formation of optimal strategies for making individual decisions in multiple trade interactions (concluding barter and monetary transactions), is proposed. A software implementation of a modified model of random sales [10] was performed with the use of the agent-based modeling methods [11, 12], machine learning [13, 14], genetic [1, 3, 15], and particle swarm optimization algorithms [16].

The urgency of development of such an intelligent system is mainly due to the high computational complexity of determining the optimal moments for concluding barter and monetary transactions in trading systems with random interactions of economic agents. In particular, in conditions when economic agents maximize the utility of future consumption due to the effective control of their own states, allowing or blocking paired trade interactions. The traditional approach to finding optimal strategies in such multi-agent systems is based on solving optimal control problems using the classical methods of variational calculus and dynamic programming [17]. However, due to the high dimensionality of such MSESs' models (i.e. a large number of interacting agents) the computational complexity of searching for individual solutions increases many times over. Therefore, it is necessary to develop a software package that uses machine learning methods and heuristic algorithms for an approximate solution of the optimal control problems for the strategy of trade interactions in the MSES.

The purpose of this work is to develop a new approach to modeling and forming strategies for making individual decisions in MSES using machine learning methods, particle swarm and genetic optimization algorithms. The general methodology of this approach is to create a simulation model of MSES, perform experiments (such as the Monte Carlo type) with the model to form a training sample, synthesize an artificial neural network (ANN) with the optimal topology and integrate it into a genetic optimization algorithm for use as a surrogate model, significantly accelerating the evolutionary search for solutions within the entire ensemble of interacting economic agents. At the same time, the effectiveness of the approach we developed and the software we designed is investigated with the use of the single-objective optimization problem for the proposed agent-based model of trade interactions, implemented with the agent-based modeling system FLAME GPU [18] and the OpenNN machine learning library [19], as the case study.

1. Agent-based model of trade interactions

The essential feature of the proposed agent-based model of trade interactions that highlights it among the previously known ones is considering the initial spatial arrangement of sellers and buyers, which is set using various configurations, examples of which are shown in *Fig. 1*.

In the model, at each moment, between each arbitrary pair of agents mutually located within the boundaries of the trade interaction zone (*Fig. 1*), a barter or monetary transaction can be completed (i.e., the exchange of goods for goods, or the exchange of goods for its monetary equivalent), if these agents, firstly, are in the state of readiness for such transactions, and secondly, they have the necessary product, or a product that is close to the target product in terms of its consumer characteristics.

Here,

$T = \{t_0, t_1, ..., |T|\}$ is the set of time moments (by days), $|T|$ is the total number of moments;

$t_0 \in T$, $t_{|T|} \in T$ are the initial and final moments of the model;

$I = \{i_1, i_2, ..., i_{|I|}\}$ is the set of agent indices, where $|I|$ is the total number of agents, $\tilde{i} \in I$ are the sellers' indices, $\hat{i} \in I$ are the buyers' indices;

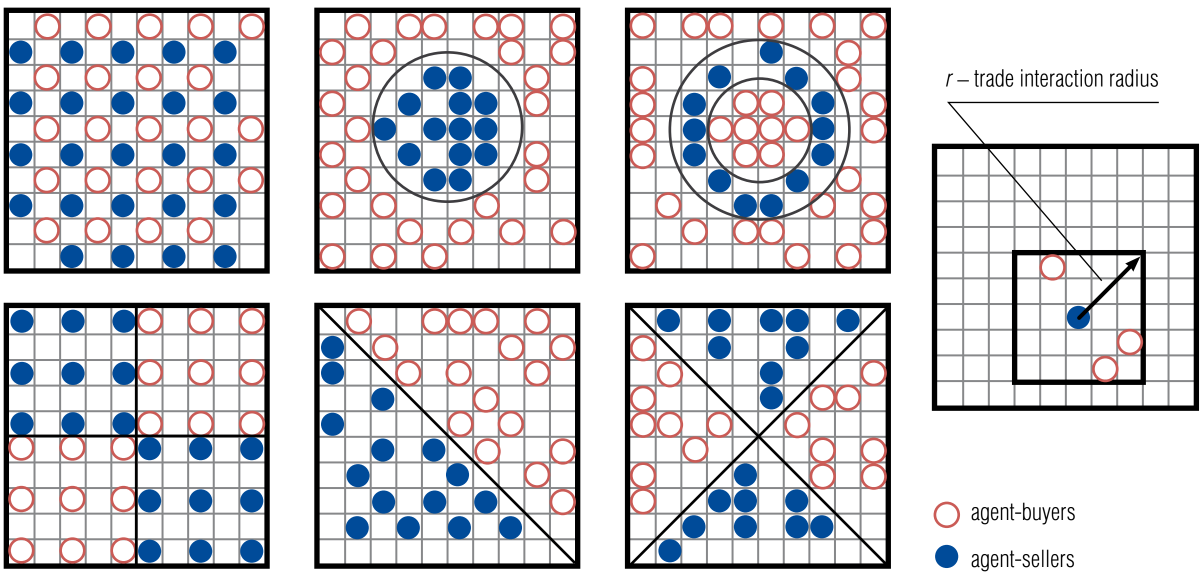

Fig. 1. Possible configurations of the initial distribution of agents in the MSES space.

$P = \{p_1, p_i, \ldots, p_{|P|}\}$ is the set of product indices, $|P|$ is the total number of products, $p_i(t_k) \in P$, $i \in I$, $t_k \in T$ is the product index belonging to the i^{th}-agent, $d_i(t_k) \in P$, $i \in I$, $t_k \in T$ is the product index that is needed for the i^{th}-agent;

$\{b_i(t_k), m_i(t_k)\} \in \{0, 1\}$, $i \in I$ is the state of readiness of the agent to conclude barter and monetary transactions, respectively, at moment t_{k-1} ($t_{k-1} \in T$): 0 — transactions are prohibited, 1 — transactions are allowed.

Then, the distance between the product of $\tilde{i}^{th}$ agent-seller ($\tilde{i} \in I$) and the product of the $\hat{i}^{th}$ agent-buyer ($\hat{i} \in I$), measured along the arc's length of a numerical circle with evenly distributed numbers 1, 2, ..., $|P|$ at moment t_{k-1} ($t_{k-1} \in T$):

$$\delta_{\tilde{i}\hat{i}}(t_k) = \\ = \frac{1}{|P-1|} \min\left\{ \left| p_{\tilde{i}}(t_k) - d_{\hat{i}}(t_k) \right|, |P| - \left| p_{\tilde{i}}(t_k) - d_{\hat{i}}(t_k) \right| \right\}. \quad (1)$$

At the same time, the assessment of the level of compliance of the product of the agent-seller with the interests of the agent-buyer can be given as:

$$\gamma_{\tilde{i}\hat{i}}(t_k) = \begin{cases} 1, \text{ если } \delta_{\tilde{i}\hat{i}}(t_k) \le \varpi, \\ 0, \text{ если } \delta_{\tilde{i}\hat{i}}(t_k) > \varpi, \end{cases} \quad (2)$$

$\varpi \ge 0$ is the coefficient of threshold compliance of the product of the agent-seller with the interests of the agent-buyer (the coefficient of 'contractability').

At the same time, the readiness states of the i^{th}-agent ($i \in I$) to complete barter and monetary transactions can be formed for each moment t_k ($t_k \in T$) using, in particular, the log-normal (*the first method*) or beta distributions (*the second method*) with given characteristics:

$$b_i(t_k) = \begin{cases} \dfrac{\lfloor \ln N(\mu_b, \sigma_b^2) \rfloor}{\lfloor \ln N(\mu_b, \sigma_b^2) \rfloor}, \text{ if I is fulfilled,} \\ 0, \text{ if II is fulfilled,} \\ \lfloor \text{Beta}(\alpha_b, \beta_b) \rfloor, \text{ if III is fulfilled,} \end{cases} \quad (3)$$

$$m_i(t_k) = \begin{cases} \dfrac{\lfloor \ln N(\mu_m, \sigma_m^2) \rfloor}{\lfloor \ln N(\mu_m, \sigma_m^2) \rfloor}, \text{ if IV is fulfilled,} \\ 0, \text{ if V is fulfilled,} \\ \lfloor \text{Beta}(\alpha_m, \beta_m) \rfloor, \text{ if VI is fulfilled,} \end{cases} \quad (4)$$

where

I. if the log-normal distribution is used to form the readiness states of agents for barter transactions and $\ln N(\mu_b, \sigma_b^2) > 0$,
II. if the log-normal distribution is used to form the readiness states of agents for barter transactions and $\ln N(\mu_b, \sigma_b^2) = 0$,
III. if the beta distribution is used to form the readiness states of agents for barter transactions,
IV. if the log-normal distribution is used to form the readiness states of agents for monetary transactions and $\ln N(\mu_m, \sigma_m^2) > 0$,
V. if the log-normal distribution is used to form the readiness states of agents for monetary transactions and $\ln N(\mu_m, \sigma_m^2) = 0$,
VI. if the beta distribution is used to form the readiness states of agents for monetary transactions.

Here,

$\ln N(\mu_b, \sigma_b^2)$, $\ln N(\mu_m, \sigma_m^2)$ are random variables having log-normal distributions with parameters μ_b, σ_b^2 and μ_m, σ_m^2, where $\mu_b, \mu_m \in [-1, 1]$, $\sigma_b^2, \sigma_m^2 \in (0, 1]$;

$\text{Beta}(\alpha_b, \beta_b)$, $\text{Beta}(\alpha_m, \beta_m)$ are random variables having beta distributions with parameters α_b, β_b and α_m, β_m, respectively.

The value of the utility function of the i^{th} agent ($i \in \{\tilde{i} : \hat{i} \in I, \gamma_{\tilde{i}\hat{i}}(t_k) = 1\}$) at moment t_k ($t_k \in T$) calculated as:

$$u_i(t_k) = \gamma_{\tilde{i}\hat{i}}(t_k) \left((\delta_{\tilde{i}\hat{i}}(t_k) + 1)^{-\nu} - \lambda r \right), \quad (5)$$

where

$r \in [1, \bar{r}]$ is the trade interaction radius, i.e. the range of cells of the discrete location space of agents considered to be neighbors, $\bar{r}$ is the maximum allowable distance between interacting agents;

$\{\nu, \lambda\}$ are coefficients that determine the impact of the costs of the distance between the target and the purchased product, as well as between the buyer and the seller, respectively.

The main control parameters of such the MSES are as follows: the configuration of the initial location of agents in space, the radius of trade interaction, the 'contractability' coefficient, the parameters of log-normal and beta distributions used to form the states of readiness of agents to conclude transactions, the probability of moving agents in space, etc.

Each agent-buyer maximizes its own utility function over the set of control parameters under constraints that have a clear physical and economic meaning. At the same time, the average (over the population of agents) utility of future consumption can be considered as the integral objective function of the MSES:

$$U = \frac{1}{|I|}\sum_{i=1}^{|I|}\sum_{t=1}^{|T|} u_i(t_k). \tag{6}$$

The software implementation of model (1)–(6) is made in the FLAME GPU environment using C++ and the graphics processing unit (GPU) architecture, which allows us, in particular, to parallelize the agent behavior logic through special functions of the FLAMEGPU_AGENT_FUNCTION type.

Table 1 presents the main functions developed for the considered model with their input and output parameters.

2. Synthesis procedure for artificial neural network

The MLP (Multilayer perceptron) model was chosen as the main configuration of the ANN for the task of approximating the objective function of the studied MSES (i.e., the agent-based model of trade interactions).

The most important hyperparameters of the designed ANN which significantly affect the approximation quality are as follows:

$\mu > 0$ is the initial learning rate;

$L = \{ l_1, l_2, ..., i_{|L|}\}$ is the number of hidden layers in MLP;

$n_l > 0$, $l \in L$ is the number of neurons in each of the available hidden layers, $n \in N$, where N is the set of all neurons;

$F_l \in \{TANH, ELU, HSig\}$, $l \in L$ is the activation function used for all neurons of the l^{th}-hidden layer (hyperbolic tangent, exponential linear, 'hard' sigmoid);

w_{n_l}, $n_l \in N$, $l \in L$ are weight coefficients of the n_l^{ths}-neurons of the l^{th}-hidden layer.

At the same time, the main criterion for the quality estimation of the ANN in the system being considered is the training error (the loss function), which should be minimized by the set of hyperparameters and weights:

$$\min_{\{\tilde{r}, |L|, n_l, F_l\}} E,\ n_i \in N,\ l \in L, \tag{7}$$

where

$$E = \frac{1}{|M|}\sum_{m=1}^{|M|}\left(\tilde{U}_{qm}(\mu, |L|, n_l, F_l, w_{n_l} X_m) - \hat{U}_m(X_m)\right)^2, \tag{8}$$

Here,

$M = \{ m_1, m_2, ..., |M|\}$ is the set of training sample data, where $|M|$is the size of the training sample;

$Q = \{ q_1, q_2, ..., |Q|\}$ is the set of iterations of the ANN learning algorithm, $|Q|$ is all iterations of the machine learning algorithm;

$\tilde{U}_{qm}$, $q \in Q$, $m \in M$ are the approximated values of the objective function (future consumption utility function) at the ANN output, calculated for the m^{th}-data sample at the q^{th}-training epoch;

$X_m = \{x_{1m}, x_{2m}, ..., x_{|Xm|}\}$, $m \in M$ is the set of values of independent variables of the m^{th} training sample (input layer of the ANN);

$\hat{U}_m(X_m)$, $m \in M$ are the known (factual) values of the objective function calculated using the previously developed agent-based model for the given X_m-set of input parameters values with the use of the Monte Carlo type method [20, 21].

Figure 2 shows a block diagram of the developed two-step ANN synthesis procedure for its further use as a surrogate model in optimization experiments.

In *Fig. 2* the following denotes are used:

Table 1.

The main functions and procedures of the stochastic model of the goods exchange

Function name	Description	Input parameters	Output parameters
FLAMEGPU_INIT_FUNCTION (init_function)	Model parameters initialization. Creation of a population of agents and their placement in a discrete space.	No	No
FLAMEGPU_EXIT_CONDITION (exit_condition)	Calculation of the objective function and checking the stop criterion.	No	No
FLAMEGPU_AGENT_FUNCTION (all_agents, flamegpu::MessageNone, flamegpu::MessageArray2D)	Sending data about each agent.	No	Coordinates, agent type, state, etc.
FLAMEGPU_AGENT_FUNCTION (all_products, flamegpu::MessageNone, flamegpu::MessageArray2D)	Sending product data by each agent.	No	Available product index, target product index.
FLAMEGPU_AGENT_FUNCTION (seeking_and_getting_product, flamegpu::MessageArray2D, flamegpu::MessageArray2D)	Search and acquisition of the target product through the goods exchange or for money. Search for the nearest seller with the desired product. Implementation of a monetary or barter transaction. Recalculation of the value of individual utility function. Sending data about the purchased product and money in the case of a monetary transaction.	Available product index, target product index.	The index of the purchased product and the money for the product (if the purchase is for money).
FLAMEGPU_AGENT_FUNCTION (getting_product_or_money, flamegpu::MessageArray2D, flamegpu::MessageNone)	Receiving a product (in barter) or money from a buyer (in a monetary transaction). Recalculation of the value of individual utility function. Completion of a trade transition.	The index of the purchased product and the money for the product (if the purchase is for money).	No
FLAMEGPU_AGENT_FUNCTION (update_agent_state, flamegpu::MessageNone, flamegpu::MessageNone)	Update the state of each agent and produce a new product if there was a trade deal in the previous step.	No	No
FLAMEGPU_AGENT_FUNCTION (update_cell, flamegpu::MessageArray2D, flamegpu::MessageNone)	Update the state of each cell of the discrete space. Checking the availability of a cell for occupation by agents.	Coordinates, agent type, state, etc.	No
FLAMEGPU_AGENT_FUNCTION (looking_for_resource, flamegpu::MessageArray2D, flamegpu::MessageArray2D)	Search for an agent that can be placed in a given cell with a given probability.	Coordinates, agent type, state, etc.	Coordinates of target cell. Information about the agent being moved.
FLAMEGPU_AGENT_FUNCTION (moving_trasaction, flamegpu::MessageArray2D, flamegpu::MessageNone)	Random movement of agents in a discrete space to a target cell.	Coordinates of target cell. Information about the agent being moved.	No

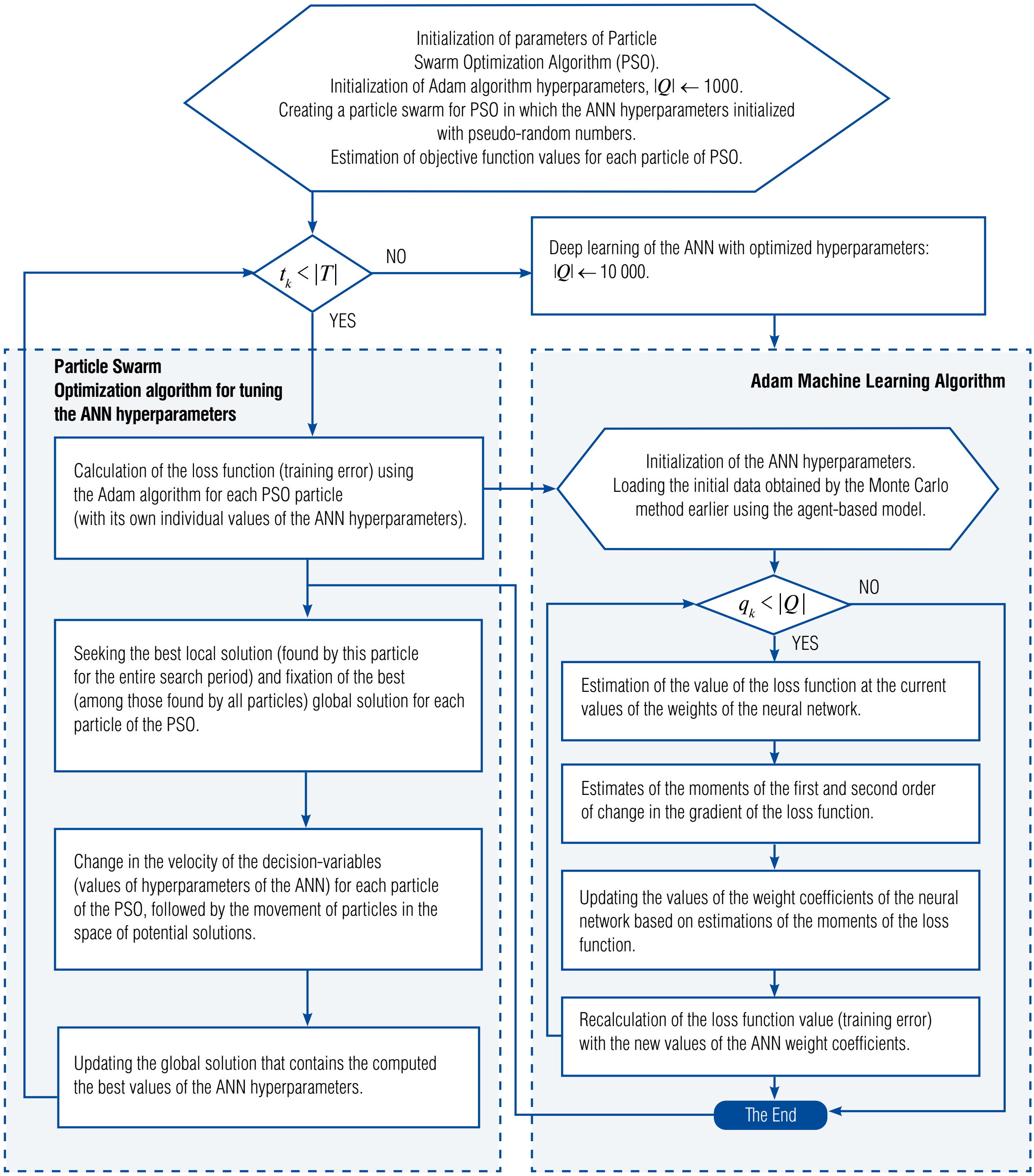

Fig. 2. Procedure for synthesising the artificial neural network using the particle swarm method and the Adam algorithm.

$T = \{t_0, t_1, ..., |T|\}$ is the set of swarm algorithm iterations (PSO), $|T|$ is the total number of iterations of the particle swarm algorithm; $t_0 \in T$, $t_{|T|} \in T$ is the initial and final iterations of the particle swarm algorithm;

$Q = \{q_0, q_1, ..., |Q|\}$ is the set of epochs of the training algorithm (the Adam), $|Q|$ is the total number of epochs of the training algorithm; $q_0 \in Q$, $q_{|Q|} \in Q$ are the initial and final epochs of the training algorithm.

On the first step, with a relatively small total number of epochs of the training algorithm ($|Q| = 1000$), the ANN hyperparameters are optimized with the use of the particle swarm optimization algorithm (PSO) aggregated trough the objective function with the Adam machine learning algorithm.

At the second step, after the best values of ANN hyperparameters are determined, deep learning is carried out using the formed ANN with a significantly larger number of epochs of the training algorithm – $|Q| = 10000$.

The Particle Swarm Optimization Algorithm (PSO) [16] provides the recalculation of ANN's hyperparameters towards improving the value of the objective function, i.e. minimizing the training error of the ANN calculated using a machine learning algorithm (such as the Adam [22]). The advantage of the particle swarm algorithm is a significantly higher time-efficiency, for example, in comparison with classical genetic algorithms, which also can be used to tune ANN hyperparameters.

Within the framework of the procedure we developed, the PSO algorithm is aggregated by the objective function (learning error of the ANN) with the Adam algorithm (*Fig. 2*). When using the particle swarm algorithm (PSO), the velocity vector of change in the desired variables values (i.e., the ANN's hyperparameters) is calculated, which determines the position of i^{ths}-particles ($i \in I$) in the potential solutions space at moment t_k ($t_k \in T$, $k = 1, 2, ..., K$):

$$\mathbf{v}_i(t_k) = \theta \mathbf{v}_i(t_{k-1}) + c_1 h(0,\,1)(\mathbf{x}_i^*(t_{k-1}) - \mathbf{x}_i(t_{k-1})) + \\ + c_2 e(0,\,1)(\mathbf{x}^g(t_{k-1}) - \mathbf{x}_i(t_{k-1})), \qquad (9)$$

$$\mathbf{x}_i(t_k) = \begin{cases} \mathbf{x}_i(t_{k-1}) + \mathbf{v}_i(t_{k-1}), \\ \text{if } \mathbf{x}_i(t_{k-1}) + \mathbf{v}_i(t_{k-1}) \in [\underline{\mathbf{x}}, \bar{\mathbf{x}}], \\ \mathbf{x}_i(t_{k-1}), \\ \text{if } \mathbf{x}_i(t_{k-1}) + \mathbf{v}_i(t_{k-1}) \notin [\underline{\mathbf{x}}, \bar{\mathbf{x}}], \end{cases} \qquad (10)$$

where

$I = \{i_1, i_2, ..., i_{|I|}\}$ is the set of particle indices PSO, where $|I|$ is the total number of particles;

$\mathbf{x}_i^*(t_{k-1})$, $\mathbf{x}_i^*(t_{k-1})$ are the best potential values of ANN's hyperparameters obtained by the i^{th}-particle of the PSO during the search period and all particles at moment t_k ($t_{k-1} \in T$);

$h(0, 1)$, $e(0, 1)$ are random variables uniformly distributed on the range [0, 1];

θ, c_1, c_2 are constants the values of which can be set in the following ranges: $\theta \in [0.4, 1.4]$, $c_1 \in [1.5, 2]$, $c_1 \in [2, 2.5]$.

The Adam machine learning algorithm [22] provides a calculation of individual training rates for different ANN's parameters (*Fig. 2*) using estimations of the first- and second- gradient moments to adapt the training rate for each neural network weight.

In the Adam algorithm the following rule uses for updating neural network weights:

$$w_q = w_{q-1} - \frac{\eta}{\sqrt{\dfrac{\tilde{\beta}_2 \upsilon_{q-1} + (1-\tilde{\beta}_1)\left(\nabla_w E_q\right)^2}{1-\tilde{\beta}_2} + \varepsilon}} \times \\ \times \frac{\tilde{\beta}_1 m_{q-1} + (1-\tilde{\beta}_1)\nabla_w E_q}{1-\tilde{\beta}_1}, \qquad (11)$$

where

$\nabla_w E_q$ is the gradient of the loss function at epoch q ($q \in Q$);

w are weight coefficients of the neural network;

m_{q-1}, υ_{q-1} are estimations of the first- and second-order moments of the change in the gradient of the loss function at epoch q ($q \in Q$);

$\tilde{\beta}_1$, $\tilde{\beta}_2$ are hyperparameters of the Adam algorithm (that usually set up at the level of 0.9 and 0.99, respectively);

ε is a fairly small number.

The proposed synthesis procedure of the ANN using the particle swarm method and the Adam machine learning algorithm is implemented using C++ and OpenNN. The reason for choosing the OpenNN as a framework for machine learning is to provide a sufficiently high performance comparable to such popular software libraries as the TensorFlow and PyTorch with a more convenient and simple development environment [19].

As a result of executing the given procedure, automatic generation of software modules of the synthesized ANN in C++ and Python is provided, which can be used as surrogate models in parallel real-coded genetic algorithms (RCGAs) [2, 3, 23], and this makes it possible to radically increase the time-efficiency of the optimal solution search procedure for multi-agent socio-economic systems.

The architecture of the ANN formed with the use of data obtained through the proposed agent-based model of trade interactions consists of an input layer, one hidden layer and an output layer. The first layer provides distribution of input signals to other neurons. At the same time, the input signals in the ANN are the scaled (normalized) values of the agent model control parameters, in particular, as follows:

- configuration (numerical equivalent) of the initial distribution of agents in the MSES;
- parameters of the log-normal and beta distributions used to form the readiness states of agents for transactions at different moments;
- trade interaction radius;
- probability of moving agents in space, etc.

The intermediate layer consists of ten neurons that use the hyperbolic tangent as the activation function.

The output layer consists of one neuron, which provides the calculation of the approximate value of the objective function – the average utility of future consumption.

Further, the synthesized ANN was integrated through objective function with the previously proposed parallel multi-agent genetic algorithm MA-RCGA [23] implemented with the FLAME GPU 2 framework [18]. This made it possible to significantly increase the time-efficiency of the evolutionary search procedure when solving the important problem of the MSES to maximize the function of the average utility of future consumption (6). The possibilities of using ANN-based surrogate models to improve the performance of evolutionary algorithms are described in [14] in more detail.

3. Results of numerical and optimization experiments

Figure 3 presents the sensitivity analysis for the process time and accuracy of the solution obtained (i.e., the closeness of the solution to the known optimum). Depending on the frequency of using the proposed agent-based MSES model for calculating the objective and fitness functions we present the corresponding ANN-based surrogate model instead. Numerical experiments were carried out using the local supercomputer CEMI RAS – DSWS PRO (2 x Intel Xeon Silver 4114, 1 x NVIDIA QUADRO RTX 600) using 10 parallel interacting agents-processes that implement the evolutionary search procedure. Such a procedure is based on well-known heuristic selection operators, crossovers and mutations. At the same time, both the original (reference) agent-based model of trade interactions and the corresponding ANN-based surrogate model are used to recalculate the values of the objective and fitness functions.

As follows from *Fig. 3*, even under conditions of the prevailing use of the ANN-based surrogate model, a sufficiently high level of accuracy of the solution obtained is ensured (more than 95%). At the same time, a significantly greater time-efficiency of the solution search procedure is achieved in comparison with the approach in which only the original (reference) agent-based model is used to recalculate the objective and fitness functions values in MA-RCGA (*Fig. 3*). Thus, the accuracy of the optimization algorithm with periodic use of the ANN-based surrogate model directly depends on the approximation characteristics of the corresponding ANN.

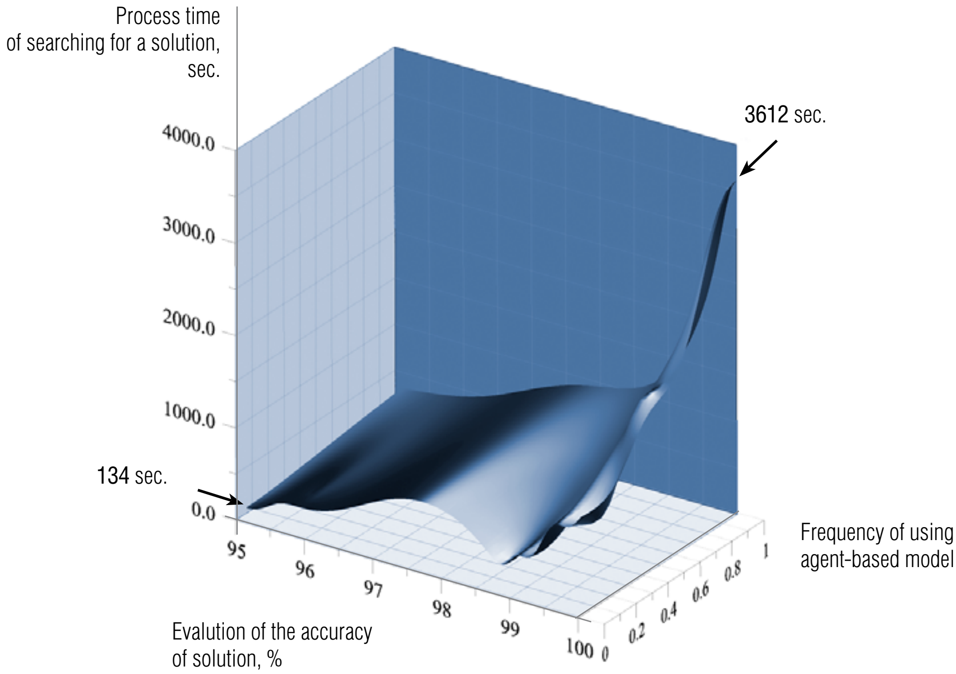

Fig. 3. Analysis of the sensitivity for the process time and accuracy of the solution obtained depending on the frequency of using the reference agent-based model.

In *Fig. 4* we present the results of optimization experiments on maximizing the average (for an ensemble of agents) utility of future consumption performed using the previously developed genetic algorithm MA-RCGA [23] and the generated ANN-based surrogate model. At the same time, various configurations of the initial arrangement of agents in space and various ways of forming the states of readiness of agents for transactions are used.

Two important conclusions follow from *Fig. 4*. Firstly, the choice of the initial configuration for agents' arrangement in space affects the value of the objective function of the MSES studied – the average utility of future consumption, and secondly, it affects the choice of a rational way to form states of readiness for transactions. In particular, for some configurations (*Fig. 4*) the use of the beta distribution is more preferable than the log-normal distribution.

Conclusion

This article presents a novel approach to modeling and optimizing strategies for making individual decisions in large-scale multi-agent socio-economic systems (MSES) using the proposed agent-based model of trade interactions as the example. A new procedure for the synthesis and training of an artificial neural network (ANN) has been developed based on the combined use of particle swarm optimization methods (to determine the optimal values of hyperparameters) and the Adam machine learning algorithm (to calculate the weight coefficients of the ANN). The ANN we designed related to the class of a multilayer perceptron (MLP) and is used as a surrogate model embedded in a previously developed multi-agent genetic algorithm (MA-RCGA) to approximate the values of the objective and fitness functions – the average (over an ensem-

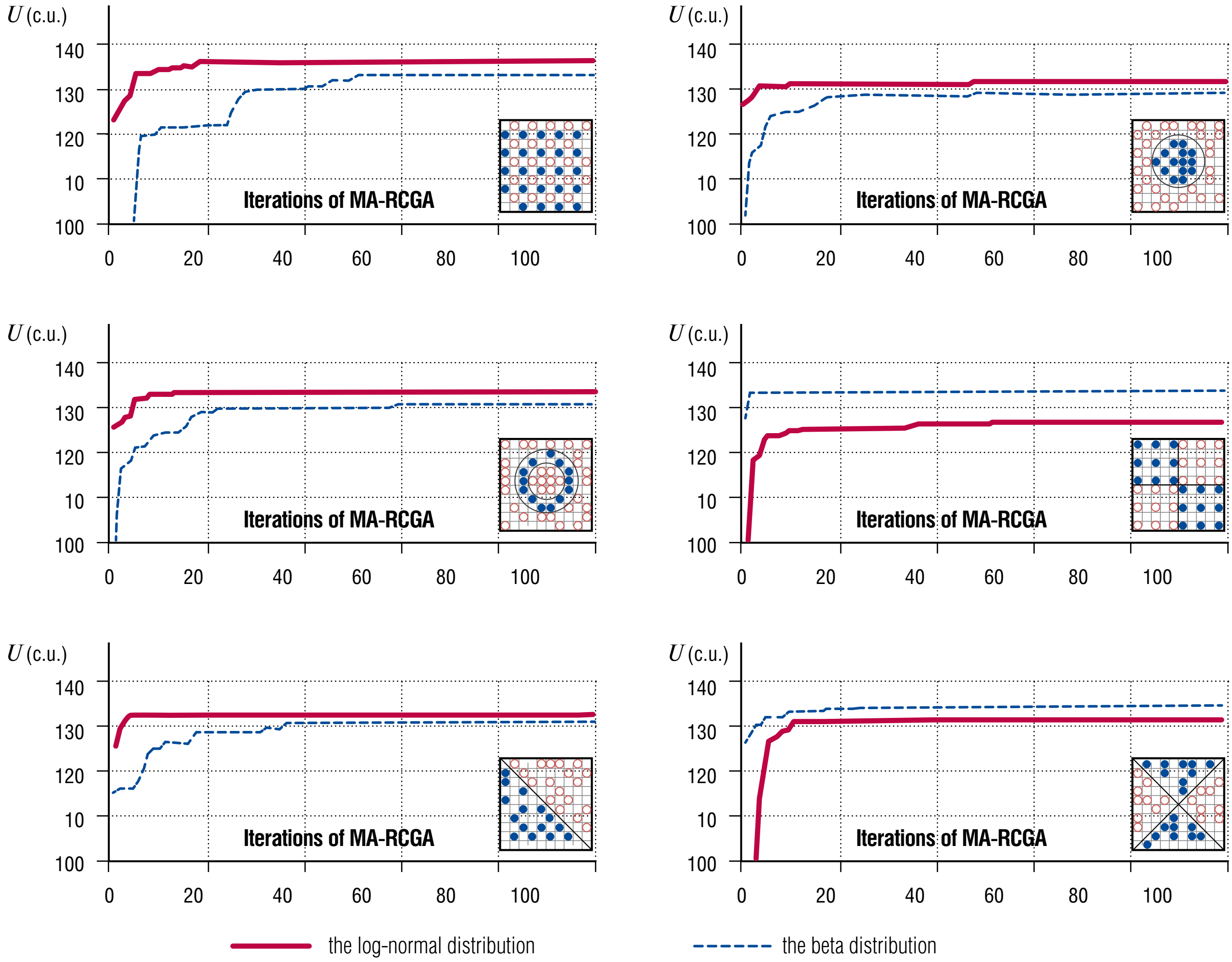

Fig. 4. Results of optimization experiments for maximizing the average utility of future consumption (*U*) when forming the states of readiness for transactions.

ble of agents) utility of future consumption. As a result of the numerical studies, it is shown that even under the conditions of the prevailing use of the ANN-based surrogate model, a sufficiently high level of accuracy for the solution obtained is provided. At the same time, the choice of the initial configuration for agents' arrangement in space affects the value of the objective function and the optimal way to form the states of agent readiness for transactions.

Further research will be aimed at creating simulation models of large-scale multi-agent socio-economic systems using machine learning methods and genetic optimization algorithms for optimal control of the ensemble of interacting economic agents' behavior. ■

Acknowledgments

The research was supported by RSF (project No. 23-21-00012).

References

1. Akopov A.S. (2014) Parallel genetic algorithm with fading selection. *International Journal of Computer Applications in Technology*, vol. 49, nos. 3–4, pp. 325–331. https://doi.org/10.1504/IJCAT.2014.062368

2. Akopov A.S., Beklaryan A.L., Thakur M., Verma B.D. (2019) Developing parallel real-coded genetic algorithms for decision-making systems of socio-ecological and economic planning. *Business Informatics*, vol. 13, no. 1, pp. 33–44. https://doi.org/10.17323/1998-0663.2019.1.33.44

3. Akopov A.S., Beklaryan L.A., Thakur M. (2022) Improvement of maneuverability within a multiagent fuzzy transportation system with the use of parallel biobjective real-coded genetic algorithm. *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12648–12664. https://doi.org/10.1109/TITS.2021.3115827

4. Akopov A.S., Beklaryan L.A. (2022) Simulation of rates of traffic accidents involving unmanned ground vehicles within a transportation system for the 'smart city'. *Business Informatics*, vol. 16, no. 4, pp. 19–35. https://doi.org/10.17323/2587-814X.2022.4.19.35

5. Khachatryan N.K., Akopov A.S. (2017) Model for organizing cargo transportation with an initial station of departure and a final station of cargo distribution. *Business Informatics*, no. 1 (39), pp. 25–35. https://doi.org/10.17323/1998-0663.2017.1.25.35

6. Saaty T. (1993) *Decision making. The Analytic Hierarchy Process.* Translation from English by R.G. Vachnadze. Moscow: Radio and communication (in Russian).

7. Kravchenko T.K., Isaev D.V. (2017) *Decision support systems*. Moscow: Urait (in Russian).

8. Kravchenko T.K., Bruskin S.N., Isaev D.V., Kuznetsova E.V. (2020) Prioritization of IT product backlog items using decision support systems. *Informacionnye Tehnologii*, vol. 26, no. 11, pp. 631–640 (in Russian). https://doi.org/10.17587/it.26.631-640

9. Kravchenko T., Shevgunov T. (2023) Equivalent exchange method for decision-making in case of alternatives with incomparable attributes. *Inventions*, vol. 8, no. 1, article 12. https://doi.org/10.3390/inventions8010012

10. Pospelov I.G. (2018) A model of random sales. *Mathematical Notes*, vol. 103, pp. 453–465. https://doi.org/10.1134/S0001434618030112

11. Makarov V., Bakhtizin A., Epstein J. (2022) Agent-based modelling for a complex world. Part 1. *Economics and the Mathematical Methods*, vol. 58, no. 1, pp. 5–26 (in Russian). https://doi.org/10.31857/S042473880018970-6

12. Makarov V., Bakhtizin A., Epstein J. (2022) Agent-based modelling for a complex world. Part 2. *Economics and the Mathematical Methods*, vol. 58, no. 2, pp. 7–21 (in Russian). https://doi.org/10.31857/S042473880020009-8

13. Tan R.K., Qian C., Wang M., Ye W. (2022) An efficient data generation method for ANN-based surrogate models. *Structural and Multidisciplinary Optimization*, vol. 65, article 90. https://doi.org/10.1007/s00158-022-03180-6

14. Díaz-Manríquez A., Toscano G., Barron-Zambrano J.H., Tello-Leal E. (2016) A review of surrogate assisted multiobjective evolutionary algorithms. *Computational Intelligence and Neuroscience*, vol. 2016, article 9420460. https://doi.org/10.1155/2016/9420460

15. Herrera F., Lozano M., Verdegay J.L. (1998) Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, vol. 12, no. 4, pp. 265–319. https://doi.org/10.1023/A:1006504901164

16. Bonyadi M.R., Michalewicz Z. (2017) Particle swarm optimization for single objective continuous space problems: a review. *Evolutionary Computation*, vol. 25, no. 1, pp. 1–54.

17. Beklaryan L.A., Flerova A.Y., Zhukova A.A. (2018) *Optimal control methods: textbook*. MIPT (in Russian).

18. Richmond P., Chisholm R., Heywood P., Leach M., Kabiri C.M. (2021) *FLAME GPU Zenodo*. https://doi.org/10.5281/zenodo.5428984

19. Lopez R. (2014) *Open NN: An open source neural networks C++ library [software]*. Available at: http://www.cimne.com/flood (accessed 10 May 2023).

20. Whitelam S., Selin V., Benlolo I., Casert C., Tamblyn I. (2022) Training neural networks using Metropolis Monte Carlo and an adaptive variant. *Machine Learning: Science and Technology,* vol. 3, article 045026. https://doi.org/10.1088/2632-2153/aca6cd

21. de Freitas J.F., Niranjan M., Gee A.H., Doucet A. (2000) Sequential Monte Carlo methods to train neural network models. *Neural Computation*, vol. 12, no. 4, pp. 955–993. https://doi.org/10.1162/089976600300015664

22. Kingma D.P., Ba J. (2014) Adam: A method for stochastic optimization. *arXiv preprint*, arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

23. Akopov A.S., Beklaryan L.A., Thakur M., Verma B.D. (2019) Parallel multi-agent real-coded genetic algorithm for large-scale black-box single-objective optimisation. *Knowledge-Based Systems*, vol. 174, pp. 103–122. https://doi.org/10.1016/j.knosys.2019.03.003

About the author

Andranik S. Akopov

Dr. Sci. (Tech.), Professor, Professor of the Russian Academy of Sciences;

Chief Researcher, Laboratory of Dynamic Models of Economy and Optimisation, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nachimovky Prospect, Moscow 117418, Russia;

E-mail: akopovas@umail.ru

ORCID: 0000-0003-0627-3037

DOI: 10.17323/2587-814X.2023.2.20.40

A knowledge management system in the strategic development of universities

Natalia V. Dneprovskaya [a, c]
E-mail: ndneprovskaya@hse.ru

Inessa V. Shevtsova [b]
E-mail: Shevtsova@spa.msu.ru

[a] HSE University
Address: 20, Myasnitskaya Street, Moscow 101000, Russia

[b] Lomonosov Moscow State University
Address: 27, build 4, Lomonosov Prospect, Moscow 119991, Russia

[c] Financial University under the Government of Russian Federation
Address: 49/2, Leningradskiy Prospect, Moscow 125167, Russia

Abstract

The purpose of this study is a conceptual description of the implementation of knowledge management systems (KMS) as a mechanism for universities' strategic development. Knowledge management (KM) practice from around the world proved the positive influence of KMS on productivity of educational institutions. The theoretical provisions and concept for KMS are determined based on an analysis of international experience of KMS use in higher education (HE). Theoretical provisions consist of 1) the staff activities as an object of KM and knowledge because of these activities, 2) the specificity of HE restrains a transfer of the KM mechanism from business to HE, and 3) the uniqueness of each university determines the structure and content of KMS for strategic development. The KM process in HE is reflected in the Socialization-Externalization-Combination-Internalization (SECI) model, where each stage contains a list of staff activities and a set of digital services. The novelty of the KM process model in HE is that knowledge flows in a wave, not a spiral. In this motion, knowledge passes from uncodified to partly codified and codified form. The study demonstrates that knowledge can go the from stage of partly codified to uncodified for revision, and knowledge flow can stop at any stage. The advantage of the concept we designed is the ability to control the flow of knowledge before it takes the codified form of a document. The digital environment for KM first allows management to control faculty activities at the initial stage of uncodified knowledge through measurement of activities, and then to estimate the knowledge flow itself. The gathered indicators help to make decisions to motivate or restrain faculty. The university management gets a complete picture of faculty activities with knowledge and the intensity of knowledge flow in training courses and educational programs.

Keywords: knowledge flow, knowledge governance, digital environment, knowledge model, metrics

Citation: Dneprovskaya N.V., Shevtsova I.V. (2023) A knowledge management system in the strategic development of universities. *Business Informatics*, vol. 17, no. 2, pp. 20–40. DOI: 10.17323/2587-814X.2023.2.20.40

Introduction

The strategic management of a university has to respond in a timely manner to ongoing changes in society, ensuring its development and the fulfillment by the university of its mission [1]. The source of the current transformation is digitalization, which creates new capabilities for universities and new requirements for structure and content of education [2–4]. Research into a proper way to meet the challenges of digitalization is within the responsibility of the strategic management of a higher education (HE) institution [5]. The purpose of the study is to present a conceptual model of knowledge management (KM) as a mechanism of university development in the context of the digitalization of society.

Business practice gives examples of implementation of KM as a tool for development of large companies. Relying on KM, companies create innovations, develop their personnel, and extend their activities [6, 7]. Corporate management has transmitted modern features to the theory and practice of KM. However, universities were the first to create knowledge management systems (KMS) based on scientific and learning activities. From the beginning, universities provided facilities for working with knowledge for society at large. Traditional methods and technologies in HE should be extended with innovative mechanisms in order to leverage the potential of digitalization for HE evolution and to respond to digital challenges properly.

Digital transformation, compared with the previous stage of technological transformation, is characterized by an increase in the complexity of knowledge [4]; acceleration of scientific and technological progress [8]; expansion of areas of interdisciplinary knowledge; an increase in the intensity of the use of intellectual assets [9]; the emergence of new forms and methods of professional employment of people [10]. These and other characteristics of digitalization give rise to challenges to HE.

Overcoming the challenges depends on how the university organizes in a digital environment working with knowledge, including creation, acquisition, accumulation and use of knowledge, which is a main asset of HE. Meanwhile digitalization opens up new sources, technologies, and methods for KMS, which allow faculty to redistribute routine operations to IT and creative cognitive tasks to humans [11]. The high importance of KMS in HE is due to the great influence of HE on society and its transformation into a knowledge society [12].

1. Statement of research tasks

ISO Knowledge management systems – Requirements to establish a broad framework for KM definition and a variety of different tools and methods for KM [13], because the structure and content of knowledge are very specific for each industry or even activity. There is a generalized definition relying on sources [13–15]: KMS is an organizational and IT environment in which a set of available methods and tools are used by people to create, accumulate, store, search, enrich, exchange and apply knowledge.

The KM mechanism should be adapted to the specificity of HE in the context of digitalization in order to ensure achievement of strategic goals of university development. For conceptual design of KMS for HE, it is necessary to solve the following tasks:

- development of theoretical provisions of KMS in higher education,
- design of a conceptual model of KMS to support the strategic management of the university.

There are many different dimensions of activity and processes in any university. For study, the business process with involvement of most of faculty staff is under the spotlight of consideration. This business process is development of training courses and educational materials. Deja [14] and Yeh [15] highlighted this specific activity as an academic KM.

2. Literature review

Some KM methods penetrate into the operational level of universities' everyday staff activities. Tools embedded in KMS such as network drives, web-services and messengers are actively used by teachers in the educational process for collaboration and sharing [16]. Occasionally implementation of a method or tool from a KM toolbox is not tagged as KMS by researchers. Publications that consider KMS as a field of science can be classified according to KM objectives. KMS can be focused on development of innovation [17], maintaining partnerships with businesses in industries [18], promotion technological transformation in HE [19, 20], facilitating e-learning [21] and teaching the theory and practice of KM [22, 23].This list of KM objectives is not complete due to the huge variety of tasks and dimensions in HE. Meanwhile Quarchioni et al. [24] summarized KM practices in HE as six conceptual approaches to the KM objective: 1) control of intellectual assets, 2) transfer of knowledge and best practices, 3) improvement of KM technologies, 4) KM training, 5) creation and sharing of academic knowledge and 6) implementation of KM.

KM in HE as a scientific field is multidisciplinary, so papers on the topic "knowledge management in higher education" cover 123 research areas in the Web of Science. There are 3102 publications retrieved in the Web of Science for 2000–2020 period: 29% of the papers are in the research area of education, 25.1% – business economics, 9.7% – engineering, 8% –computer science, 7% – information science and library science. Most of the research findings on KM in HE are focused on the practice of solving operational tasks in HE. Only 1% of the papers raise issues of KMS in strategic management in HE.

The practice of KMS in strategic management has already been studied in several universities which are included in the World University Rankings[1]. The experience of King Abdulaziz University in Saudi Arabia [25] demonstrates an organizational culture as a main driver of KM. Italian University of Bari [26] and Ca' Foscari University of Venice [27] use facilities of KMS as an environment for interactions between academic and business communities, also as a mechanism for attracting enrollees. In China's Wuhan University of Technology [28] and in India's University of Delhi [29] KMS is implemented to ensure demand for their graduates and future employment. Another Chinese university, Northwestern Polytechnical University [30], uses KMS to enhance research activities among faculty staff and students. A case study of Moscow State University of Economics, Statistics, and Informatics (Russia) presents KMS as a means for innovative transformation of education into e-learning [31]. Consideration of the above sources shows that the effectiveness of KMS is measured through indicators of university performance. There is a huge experience of applying of the performance indicators for strategic management in HE on large simulation systems, decision support systems and business intelligence [32]. Digital transformation projects, regardless of the field of activity, are always aimed at the strategic development of an organization [33, 34].

Meanwhile, a conclusion about the positive impact of KMS on university performance can hardly be drawn without exclusions due to the so-called survivor bias. In the sources under consideration, the assessment of KMS influence on university performance is based on a survey among students [29, 35, 36] and lecturers [25, 37] to show their satisfaction with KMS and the relevance of KMS to their activities. It is necessary to take into account the limitation of the method of surveys and expertise in evaluating performance when interpreting the results. The conclusions drawn from our review of literature cannot be extrapolated to the entire HE due to differences in the understanding of KMS and its tools on a case-by-case basis.

Several studies have been carried out on a national scale covering the practice of KMS at a few universities: in the United Kingdom [36, 38], Australia [39], Spain [40, 41], Poland [14] and Malaysia [31, 42]. In studies at the national level, the tasks of the KMS are revealed in the context of state regulation and regional specifics. National HE systems differ significantly from each other, but they are united by the dominant influence of public authorities on the operational and strategic activities of universities. The introduction of KMS in Italian [27] and Australian [39] universities may be hindered by regulations. In Poland, the KMS is supported and implemented at the national level as a mechanism to ensure the transparency and manageability of intellectual assets at each university and throughout the country [14].

It is quite difficult to single out a universal structure from the sources of the KMS which could fit universities of at least one type. Moreover, in other industries there is no shared understanding about KMS structure and content. Common to HE and other industries is the awareness that KMS supports and ensures the achievement of strategic goals. A review of the literature shows a gap in the disclosure of the conceptual scheme of the KMS as a mechanism for strategic management of the university in the context of the digitalization of society. The academic community will have to conduct full-scale studies of KMS in the strategic management of the university.

3. Methodology of the study

The empirical data for studying the content of KMS in strategic management in HE were extracted from sources describing KMS practice of universities located in Australia [39], the United Kingdom [36, 38], India [29], Italy [26, 27], Spain [40, 41], China [28, 30], Malaysia [31, 42], Saudi Arabia [25], Poland [14] and Russia [31]. Methods of analysis, comparison and generalization are applied to develop the theoretical provisions of KMS.

[1] World University Rankings https://www.timeshighereducation.com/world-university-rankings

Methods of categorization [43] and semantic modeling [44] are used to design a conceptual model of KMS in strategic management in HE.

Because of the huge number of university activities, the study is limited to considering the processes of developing educational programs and its educational and methodological content, i.e., academic knowledge. An attempt to cover all university fields at once could lead to blurry, non-specific results.

4. Research results and discussion

4.1. Theoretical provisions of KM in HE

4.1.1. KM Object

Universities were the first organizations to hold KM systematically. Their managerial activity focused on knowledge accumulation, storage and dissemination. The relevance of knowledge control appeared in the processes where the value of knowledge is prioritized as assets. The first business cases of KM considered the problem of knowledge retention that arises when an employee generation changes [7]. A well-known and widespread solution to this problem is documenting and storing information about knowledge in an information system, library or knowledge base. Knowledge has been defined for centuries as subjective [45], which does not exist outside the context of human activity. Thus, information systems store information about knowledge and not the knowledge itself. Recent studies support the concept of knowledge as a subjective category [46] and expand the list of knowledge subjects to include an organization and a local community [6, 11, 47]. Therefore, an organization can learn, create, store and use knowledge. Organizational knowledge as a management resource is characterized as intellectual capital and connects human, social and operational assets [48].

The subject property of knowledge determines the priority of the qualitative measurement of its value over quantitative characteristics [14]. The academic community discusses the issue of qualitative measurement of scientific results because the quantitative measurement through the evaluation of bibliometric indicators does not reflect the level and significance of scientific results [49]. The quantitative measurement should be given by an expert in a proper scientific area [50]. Expertise of study is a time-consuming and expansive method, so it can be applied to cases where the main function of KMS is distinguishing the most important knowledge. If the main function is creation, sharing, dissemination and modification of knowledge, the expertise will slow down KM processes. When the scientific and technological progress is accelerating, such a slowdown of KM limits the flexibility and intensity of work with knowledge.

The processes of external and/or internal peer review are used to approve the syllabus of training courses in almost all Russian universities. The process of assessing the quality of knowledge is laborious and cannot cover the entire volume of knowledge circulating at the university.

Early MK theories relied on various surrogates for knowledge to separate knowledge from humans and extract the most valuable information from the available content. Founders of KM theory Nonaka and Takeuchi [6] chose the use of knowledge by people as a sign of knowledge that is of value to the organization. Kurlov and Petrov [51] for the purposes of innovation management introduce a concept of instrumental knowledge on the basis of which an activity is transformed. The ISO [13] deals with the value of knowledge, not knowledge itself. In order to consider KMS as a mechanism for strategic management, it is necessary to put aside the discussion about the structure and content of knowledge.

The value of knowledge is defined by staff activities with knowledge in the performance of their labor functions. Thus, staff activities regarding to knowledge should be considered as an object in KMS. The first theoretical provision is that the object of KMS is the activity of users in the knowledge environment.

ISO [13] defines an environment that provides favorable conditions for people to work with knowledge as a common means of KM. In a broad sense, the environment contains the internal capabilities of the organization and a part of the external sources of knowledge and experts. In a narrow sense, the environment is supported by the KMS, which is a set of organizational and information solutions for performing the functions of the KM. Through KMS employees get access to knowledge, can interact with each other, and use different methods and tools to work with knowledge.

Staff activities drive the knowledge flow in educational and other areas of universities. A study of communication between lecturers shows their high appreciation of the opportunity to interact with each other [52]. A series of conversations conducted with Nobel laureates in economics emphasizes the great role of the communication environment in their scientific progress. World science leaders highlight the importance of informal discussion of hypotheses and theories with colleagues [53]. The stage of informal discussions is included in the cycle of scientific and technical information, including non-published materials; from this stage the life cycle of knowledge begins in the knowledge management system of the state corporation Rosatom [7].

4.1.2. Specificity of KM in HE

The spread of KM technologies and methods among businesses is uneven. Almost every industry has its own KM methodology. The need to adapt and develop a special approach to KMS for a given industry is due to the specific properties of knowledge for each industry and even organization [13, 54]. The dependence of knowledge on subjective interpretation in the context of an industry makes it difficult to directly transfer best KM practices across industries and organizations.

KMS as a mechanism for strategic development came from the business community to HE [55]. In business, various ways of implementing KMS are used which differ depending on the goals of strategic development and the industry or market specificity where the organization operates. Rosatom developed the KMS based on the scientific and technical information system to control codified (documented) intellectual assets [7]. The Japanese companies Honda Motors and Eisai relied on the knowledge environment in which employees have a deal mainly with uncodified (undocumented) knowledge [56].

The specifics of HE institutions influence a methodology of KMS for universities. The main feature lies in setting a goal of strategic development. Kuzminov and Yudkevich [1] point out that goals of strategic development for Russian universities are set by public authorities. There is also a dependence of the national HE system on budget funding, which limits any initiative of universities in choosing their own way of development. A large role of public authorities in the KM practice in HE stands out in Australia, Italy and Poland.

The KM environment is often considered from the perspective of its three enlarged groups of elements: people, processes, technologies [57, 58]. Through human activities, knowledge acquires its value and meaning. Often a department responsible for personnel development also is responsible for KMS. The processes performed in an organization determine the possibility and space to include KM activities into the business. These processes impose requirements on the structure and content of the KMS. Organizational development policy and regulations should rely on KMS and describe the KMS contribution to performance of the organization and productivity of employees. Current digital technologies provide KMS with ingenious tools for creating and sharing knowledge. The emphasis on one of three enlarged groups of KMS elements puts responsibility for KMS on the HR, administration, or IT department of an institution. *Table 1* summarizes the features of KMS at universities by people, processes and technologies.

Table 1.

Features of KMS in HE universities

Group of KMS elements	HE feature
People	Confirmed high intellectual potential of employees (scientists and lecturers) [3, 40]
	The ability to use intellectual potential from the business environment through graduates [59]
	Employees' acceptance of the value of the free exchange of knowledge for the development of education and science [42, 60]
	Academic competition among faculty staff [36]
Processes	Conducting research and educational programs in a large number of fields [61]
	Diverse approaches to forming and supporting creative teams and projects
	Priority for fulfilling the public mission of science and education [35]
	Integration and intensive interaction with external communities [3, 40]
	Strict regulation and control by public authorities of HE [62]
Technologies	External content sources: digital libraries, databases
	Scattered internal sources of content: teaching materials, scientific reports, regulations, etc.
	Strict information security requirements apply to work with personal data, but not to content that is created, used and distributed in education and research
	The concept of "BYOD" according to which the lecturers and students themselves choose the computers, software and web services that are suitable for them in terms of performance and cost [63]

The second theoretical provision is that design, structure and contents of KMS for universities should take into account the features of HE in order to fully realize the high intellectual potential of employees and cover many scientific areas and training courses with the help of the heterogeneity of IT facilities for education and science.

An analysis of the KMS practice in universities shows that each group of elements contributes to success and strategic development. Elements of KMS provide cultural [25], organizational [39] and technological [41] conditions for the success of KMS in HE.

4.1.3. Adaptation of the KM mechanism to strategic goals of university development

KMS as a mechanism for strategic development is based on the mission and values of the university [56]. KM cases in universities differ significantly from each other, but their common features are revealed when they are grouped by mission type. The practices of KM implementation in universities follow a common mission type and also have common features. There are three types of mission in HE: educational, scientific and the so-called third mission. The third mission appeared because of changes in society under the influence of scientific and technological progress, economic globalization, political and economic crises [64]. The third mission of the university directly influences the socio-economic development of a city or local area by facilitating interaction between communities of entrepreneurs and citizens, the dissemination of best practices and new business models, etc. [65]. Meanwhile universities staying on their educational and scientific mission indirectly influence societal development through their graduates and scientific results. Universities have been guided by an educational and scientific mission for centuries. Lomonosov Moscow State University nowadays follows the mission formulated in the 18th century[2].

The productivity of KMS is measured by the performance indicators of a university. Based on analysis of the KM practice in different universities, the characteristics of KMS are extracted in accordance with the type of mission in terms of geographic scape and KM means (*Table 2*).

The university's educational mission focuses on the value of professional evolution and demand for their graduates. Employment of graduates is regarded as one of the main key performance indicators of the university. Therefore, KMS aims to ensure that graduates of educational programs are in demand in the labor market. Universities with an educational mission conduct their activities in selected regions to build relationships with employers and interact with the labor market.

The scientific mission of the university sets a task for strategic management to advance in world rankings, promote papers in top scientific resources and obtain world-class scientific results. These tasks determine the global scope of KMS [67]. The activities of faculty staff in dealing with knowledge may be located outside a campus. Case studies of research universities raise the issue of negative impact of some tools or practices of KMS on performance indicators. An analysis of the implementation of KMS by 70 Spanish universities found a relationship between the spread of IT for collaboration and a decrease in the number of publications in top-cited journals [41].

Universities of the third mission focus on the social, cultural, and technological development of a particular region, such as a city [68]. The third mission is most often characteristic of entrepreneurial universities [26], which act as a connector between businesses, citizens and public authorities [65]. In smart cities projects, universities perform functions of generating, collecting and selecting knowledge to fill a lack of scientific and educational expertise in business and society. Rapid changes in technology, the economy and society require HE institutions to diversify sources of knowledge and ensure their transfer to society. Thus, universities link parts of a societal ecosystem: production, education, public administration and research. The considered cases

[2] Mission of Lomonosov Moscow State University: "the education of peoples for the benefit of our common humanity, ... for the well-being of the entire fatherland". Source: Program for the Development of Lomonosov Moscow State University until 2020. Government of Russian Federation. Order of September 27, 2010. No. 1617-r. http://pravo.gov.ru/proxy/ips/?doc_itself=&nd=102141648&page=1&rdk=5&link_id=6#I0 (in Russian).

Table 2.

Characteristics of KMS for mission in HE

	Educational mission [25, 28, 29]	**Scientific mission** [30, 40, 63]	**Third mission** [26, 27]
Key performance indicators	employment, competencies, education, employer, student satisfaction, rating	publications, rating, citations, patents on scientific results, innovations	innovations, competitive advantage, value, strategy, improvement of society
Geographic scope of activity	In selected regions or countries	Global	Regional
The most typical knowledge management tools	Corporate portals, collaboration tools based on cloud computing	Communities of practice [66], knowledge libraries, variety of information sources, collaboration tools	LivingLabs [65], communities of practice

of the use of KMS in universities of the third mission show a local or clearly defined regional scale of their activities.

The third theoretical provision of KMS in HE is to ensure that the university fulfills its mission. At the same time, the productivity of KMS is measured by the key performance indicators of the university, and not by the performance of individual functions of KM.

Following key performance indicators in the strategic management of the university is the basis in BPM (business performance management) systems, which are already used in HE [32]. Thus, KMS can be embedded into an existing IT landscape of strategic management using the available IT infrastructure for data storage and analytical processing.

4.2. Conceptual model of KM

4.2.1. Conceptual scheme of knowledge flow in HE

The activities of faculty staff drive knowledge flow in the university, which goes through stages from the birth of an idea of knowledge (creation of a training course) to its use and distribution in codified form as educational and methodological materials. In HE, knowledge is often understood as scientific and technical information, and a process of creating knowledge goes through a cycle of unpublished knowledge, primary sources of knowledge publication and secondary sources of knowledge publication [7, p. 75]. In business practice, the SECI model by Nonaka et al. [56] is widespread. This model of creation and use of knowledge in organizations consists of the stages: Socialization, Externalization, Combination, Internalization (SECI). The authors of the SECI model distinguish the stages depending on the degree of knowledge codification and the number of participants involved. Based on the SECI model, *Fig. 1* shows the stages in the development of teaching and learning materials. *Figure 1* demonstrates a sequence of stages in a clockwise direction. The inner circle contains a list of staff activities, and the outer circle contains the means of digital environment for performance of these activities. For three stages (E, C, I), types of codified knowledge are given as an example, and the figure does not contain a complete list of possible documents.

Stage S is the initiation or relaunch of a knowledge project. The stage consists of interpersonal interactions of a few lecturers. The results of this stage can be recorded in the form of drafts and a set of ideas, but they are not published as documents. Thus, knowledge is not registered and included in information systems or libraries, because it is uncodified. E-mail or social media can be used at this stage. Participants are a small group of authors.

Teece [69] points out that supporting staff activities with uncodified knowledge ensures intellectual assets as a stable source of competitive advantages for an organization. In Russian universities, this stage is practically not controlled by management since it takes place in the lecturers' environment and is not supported at the university level. Consequently, universities do not receive possible benefits for their development from the stage S of creation of educational materials.

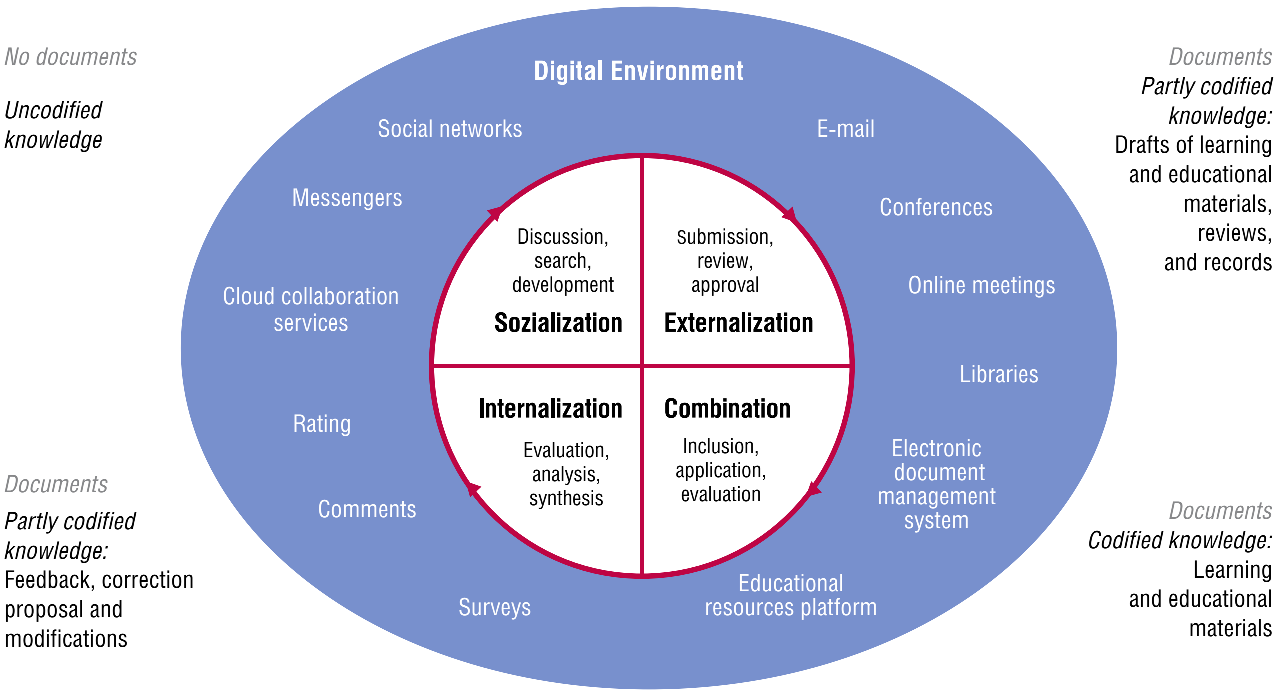

Fig. 1. Process of development of learning and teaching materials at SECI model.

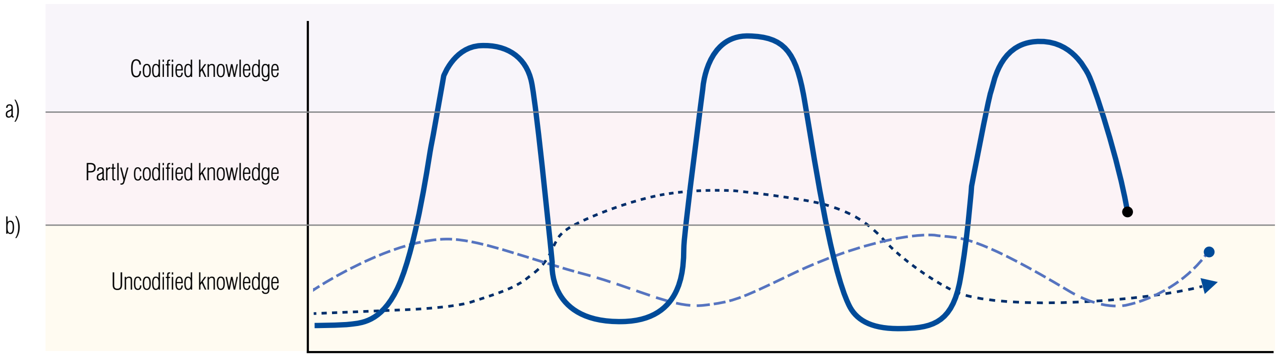

Fig. 2. Academic knowledge flow.

At **stage E**, knowledge is partly codified to involve more people in knowledge processing such as the review, discussion and approval of materials submitted by authors. The approval of educational materials could be done in different ways. At HSE University, an academic council of the educational program reviews and approves a syllabus of training courses. At the Plekhanov Russian University of Economics, this is done by a scientific methodological council of higher schools.

Educational, learning and teaching materials are codified at **stage C**, when materials are approved and accepted. At this stage educational materials become available in libraries and information systems. They are open for lecturers and students to use in training courses of the university.

The final **stage I** in the KM process includes an assessment, feedback, analysis and synthesis based on the experience of using knowledge. At stage I, we find the students' assessment of their learning experience during a training course and the lecturers' assessment of their teaching experience. The knowledge gained at stage I is codified as ratings, proposals, comments and recommendations on the results of the analysis and the synthesis of practice to use the materials.

The SECI model is often presented as a spiral on the timeline, where knowledge sequentially passes through the stages, and the cycle of working with knowledge is repeated on a new round. The development of training and methodological materials in general goes through all the stages of SECI, but trajectories can be different. The different trajectories arise because knowledge can move back and forth. For instance, after discussion on stage E a syllabus returns to the previous stage S for a revision. Thus, on the timeline, the knowledge flow looks like a wave. *Figure 2* is a schematic presentation of academic knowledge flow, where the x-axis is a time scale, and the y-axis is a categorical scale reflecting the levels of knowledge codification.

In *Fig. 2* the wave shown by the solid line crosses level a of codified knowledge three times. That means that the training materials went through three full cycles and were used in the training course. On the peak of the wave, the educational materials are being approved and accepted. Waves shown in dashed lines do not reach the stage C and cross level a; they do not enter a library or repository; and they are not introduced into training courses. Meanwhile the work on this material is ongoing. The full stop at the end of the wave means the end of work on materials. Some flows of knowledge are stopped after a week, while others can run on for years. Knowledge flows differ significantly in the duration and intensity of the waves, depending on a training course, scientific area, motivation and competence of the author team. If in some scientific areas a life cycle of knowledge can be more than five years, then in others it will not exceed a year [70]. Knowledge flows in various areas of training and scientific areas can take various periods of time from several weeks to years.

The number of knowledge flows in a university can be indirectly measured through the number of educational programs and training courses. Knowledge flows can be grouped in an educational program or a scientific area based on departments.

The model of academic knowledge flow offered here does not change the usual course of its development but formalizes it for control and management. The traditional approach to KM through a codification and storage of knowledge in libraries allows universities to control knowledge that entered a library in codified forms as syllabus, curricula, textbooks, etc. The importance of libraries as knowledge repositories is not subject to revision, but they should be complemented by digital means that support knowledge operations and interaction between employees. The staff activities use partly codified knowledge and are partly controlled by the administration. All activities below level *b* are out of sight for the university administration. The flow of knowledge in the digital environment allows the administration to bring all its stages out of the blind zone and ensure control over them.

4.2.2. System of indicators for knowledge flow measurement

The main function of KMS is to support the knowledge flow which is provided through measurement and control. The control of knowledge flow requires a system of indicators to assess the state, intensity and volume of knowledge flows.

The digitalization of society enhances the transfer of many activities and processes to the digital environment. One of the advantages of the digital environment is the ability to automatically gather data on selected metrics. The modern knowledge environment is a digital environment. A significant part of activities with knowledge is carried out using digital services, such as e-mail, messengers, online conferences, collaboration through cloud services and network storage disks. Thus, the digital environment of KM meets the necessary condition for the automatic measurement of the staff activities driving the knowledge flow in motion.

The SECI model shows that knowledge codification is preceded by the stage of knowledge emergence, which assumes operations with implicit knowledge. It is impossible to measure uncodified or implicit knowledge, but it is known that it appears in staff interaction. This stage is usually not considered and controlled by the university administration. The existence of implicit knowledge in KMS can be compared to the phenomenon of a black box in cybernetics, in which input and output can be under control, but not inside of the black box [71]. Precisely at stage S (socialization) there is the occurrence of new knowledge or adaptation of already known knowledge to changes and new requirements.

The digital environment allows for capture of the state of each stage of the knowledge flow and control of its progress. The object of control in KMS is a staff activity; therefore, the system of indicators of the knowledge flow quantitatively measures the staff activities in the knowledge flow. In accordance with the SECI model for the stages of developing training materials, the indicators can be grouped as follows:

1) interaction and communication between employees characterize stage S, which does not contain codified knowledge;
2) contribution of employees to the knowledge library — stage C;
3) knowledge sharing at stages E and I, where knowledge is partially codified. In the knowledge flow scorecard, these two stages are combined because they both involve discussion and interaction involving a group of stakeholders (a supervisor of the educational program, students, lecturers, etc.).

The knowledge flow indicators are presented in *Table 3*, which contains a short description of source, data type and method for measurement.

A comparison of values of the indicators of knowledge flows for different training courses and educational programs is a function of KMS specific to HE. Similar values of staff activities at stage S for most of knowledge flows point to a homogeneity of the organizational culture at a university. A means for managerial impact on promotion of the organizational culture is justified by measuring the indicators of stage S. If the values of staff activities in one knowledge flow are lower than in other flows, then this indicates the disunity of the lecturer team in the area. In business, the phenomenon of sabotage is known [72]: this is when employees deliberately exclude themselves from the flow of knowledge.

At stage C, the contribution of a lecturer to the accumulation and keeping of knowledge is assessed. Meanwhile, the value of knowledge is not assessed. The indirect assessment of the value of knowledge through its relevance supposes a risk that some knowledge may be underestimated and lost. This risk was first described in the middle of 20th century, when it was discovered that society does not have enough capacity to store and process the entire information flow which is permanently growing and varying [73]. Despite the breakthrough development and spread of digital technologies, this risk remains relevant [74].

Knowledge sharing indicators characterize the stages of work with partially codified knowledge when other persons in addition to the authors join the knowledge flow to discuss and improve materials. Values of these indicators point to the intensity and volume of the flow of knowledge, and help determine the need for support for staff activities in the stages E and I.

The knowledge flows of a contemporary university are growing and changing all the time. The digital environment is suitable for measuring and considering the processes of working with knowledge. The stages of creation and use of academic knowledge become transparent for control and, therefore, manageable.

KMS should be considered as one of the application layer elements of the IT architecture of a university shown in *Fig. 3*. Using the service approach, KMS is integrated into the IT landscape of the university in such a way as to use the capabilities of the multidimensional warehouse for storing and processing the indicators of the knowledge flow, and BPM systems for measuring performance indicators and evaluating the performance of KMS.

On the one hand, KMS uses the possibilities of digitalization in terms of simulation modeling and predictive analytics of knowledge flows. On the other hand, KMS complements the strategic management systems of HE with data on the flows of knowledge, all of which have a decisive impact on the university performance.

Conclusion

In the context of high technological and economic dynamics, the university, along with business, needs a favorable environment for creating innovations that ensure its development. In business practice, an approach using methods and technologies of knowledge management has become widespread. These means, combined in KMS, can complement traditional higher education approaches based on scientific research and systematic university staff training.

The specificity of KMS in higher education lies in the fact that the object of control is the activities of faculty staff for the development, modification, discussion and use of educational materials. The flow of academic knowledge is set in motion by lecturers from the birth of an idea to its implementation in the educational process and subsequent refinement. KMS introduction in the university requires taking into account the specifics of higher education, such as a large number of training courses and scientific areas, the proven high intellectual potential of staff, and

Table 3.

Indicators of knowledge flow in KMS

Groupe of indicators	SECI stage	Indicator	Source	Type of data	Type of measurement
1. Interaction and communication between employees	S	1.1. Communication intensity, number and frequency of messages sent and received	Digital interaction services: e-mail, messengers, forums	Numerical	Frequency, quantity
	S	1.2. The content of the interaction, messages sent and received	Digital interaction services: e-mail, messengers, survey	Categorial	Content analysis
	S	1.3. Coverage of interactions, number of lecturers involved in interactions	Digital interaction services: e-mail, messengers, forums	Categorial	Frequency, quantity
2. Contribution of employees to the knowledge base	C	2.1. Download of materials	Knowledge library	Numerical	Frequency, quantity
	C	2.2. Content of uploaded content	Knowledge library	Categorial	Content analysis
3. Knowledge sharing	E, I	3.1. The intensity of reviewing, commenting, feedback on colleagues' materials	Reviewing and commenting services	Numerical and categorial	Number, volume of reviews (comments)
	I	3.2. Rating, feedback on practice of knowledge use	Digital interaction services: e-mail, messengers, survey	Numerical and categorial	Number of ratings
	E, I	3.3. Commenting	Digital interaction services: e-mail, messengers, survey	Numerical and categorial	Number, volume of comments
	E	3.4. Update intensity, number of versions and frequency of changes	Knowledge library	Numerical	Frequency, quantity

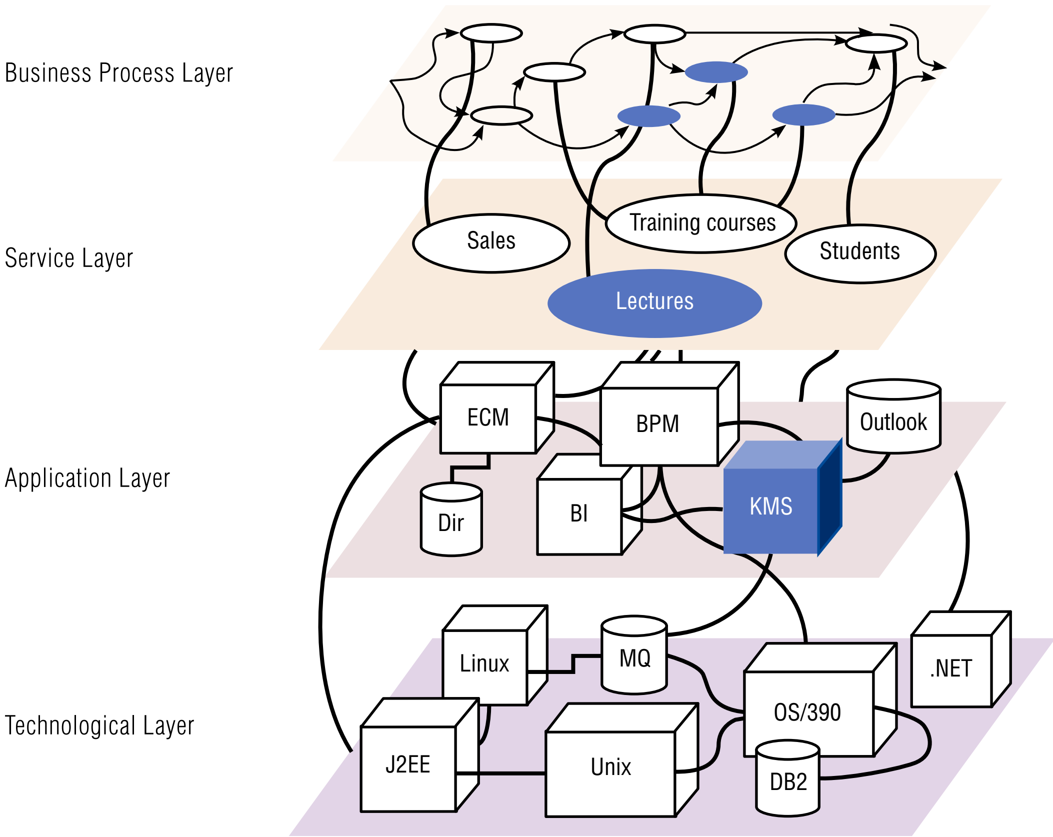

Fig. 3. KMS in IT architecture of university. Source: the figure is adapted from [34, p. 229].

the disparate IT infrastructure of the university with many involved technologies and knowledge sources. Also, the methods and technologies in KMS should be adapted to the individual needs and capabilities of each university which are determined by the mission, region, scale and other parameters. The specifics of each university make it difficult to develop a standard of KMS suitable for all institutions of higher education but do not prevent knowledge flow modeling.

The flow of academic knowledge at the university is presented based on the SECI model of the process of creating and using knowledge in organizations. Our modified SECI model, adapted to higher education, contains a list of activities and digital services that ensure the motion of the knowledge flows. The flow moves in waves through the stages of uncodified knowledge (S), partially codified (E, I) and fully codified knowledge (C). Currently almost all knowledge management functions are carried out using IT, which allows us to control the indicators of the intensity of the knowledge flows.

A knowledge flow in the digital environment become a transparent to measure its scope, intensity and volume. Timely and informed decision making relies on the measurement of knowledge flows. The proposed system of indicators measures the interaction and communication between faculty staff, their contribution to the creation of educational materials, their participation in collaboration. Since many knowledge flows exist simultaneously at the university, by comparing the flows with each other it is possible to identify the most and least successful practices and have an appropriate impact on staff.

The modern methodology of the KMS makes it possible to form a set of events to involve almost all university staff in the development and dissemination of knowledge. A university that does not fully control the knowledge flows does not have a complete understanding of the innovative potential of its strategic development. Further research in the field of KM in higher education is aimed at developing the principles of KMS at universities, structuring the methods and technologies of KMS by levels of management and areas. The authors of this study are working on testing the theoretical and methodological provisions of KMS proposed in the article at a team level in Russian universities. ■

Acknowledgments

This paper was prepared on the basis of a study funded from the budget by the State Assignment of the Financial University.

References

1. Kuzminov Y.I., Yudkevich M.M., Altbach P.G. (2022) *Higher education in Russia*. Johns Hopkins University Press.
2. Rogozin D., Solodovnikova O., Ipatova A. (2022) How university teachers view the digital transformation of higher education. *Voprosy Obrazovaniya (Educational Studies Moscow)*, no. 1, pp. 271–300. https://doi.org/10.17323/1814-9545-2022-1-271-300
3. Frunzaru V., Vătămănescu E.-M., Gazzola P., Bolisani E. (2018) Challenges to higher education in the knowledge economy: anti-intellectualism, materialism and employability. *Knowledge Management Research & Practice*, vol. 16, pp. 388–401. https://doi.org/10.1080/14778238.2018.1493368
4. Ilvonen I., Thalmann S., Manhart M., Sillaber C. (2018) Reconciling digital transformation and knowledge protection: a research agenda. *Knowledge Management Research & Practice*, vol. 16, no. 2, pp. 235–244.
5. Salmi J. (2016) Excellence strategies and the creation of world-class universities. *Matching Visibility and Performance. Global Perspectives on Higher Education* (eds. N.C. Liu, Y. Cheng, Q. Wang). Sense Publishers, Rotterdam. https://doi.org/10.1007/978-94-6300-773-3_2
6. Nonaka I. (2007) The knowledge-creating company. *Harvard business review*, vol. 85, pp. 162–171.
7. Pershukov V.A., Medovnikov D.S., et al. (2012) *Rosatom shares knowledge*. Moscow: HSE (in Russian).
8. Akaev A., Rudskoy A., Khusainov B., Zeman Z. (2022) Information model for calculating the rate of technical progress. *Digital Transformation and the World Economy. Studies on Entrepreneurship, Structural Change and Industrial Dynamics* (eds. A. Rudskoi, A. Akaev, T. Devezas). Springer, Cham. https://doi.org/10.1007/978-3-030-89832-8_2
9. Angelidou M. (2015) Smart cities: A conjuncture of four forces. *Cities*, vol. 47, pp. 95–106. https://doi.org/10.1016/j.cities.2015.05.004
10. Kuzminov Ya., Sorokin P., Froumin I. (2019) Generic and specific skills as components of human capital: New challenges for education theory and practice. *Foresight and STI Governance*, vol. 13, no. 2, pp. 19–41. https://doi.org/10.17323/2500-2597.2019.2.19.41

11. Slavin B.B. (2022) Technologies of collective intelligence in the management of business processes of the organization. *Business Informatics*, vol. 16, no. 2, pp. 36–48. https://doi.org/10.17323/2587-814X.2022.2.36.48

12. Independent Evaluation Group (2017) *Higher education for development: an evaluation of the World Bank Group's support (English)*. Sector or thematic evaluation Washington, D.C.: World Bank Group. Available at: http://documents.worldbank.org/curated/en/729101493052924041/Higher-education-for-development-an-evaluation-of-the-World-Bank-Group-s-support (accessed 10 May 2023).

13. ISO (2018) *Knowledge management systems – Requirements*. ISO 30401:2018. Available at: https://www.iso.org/ru/standard/68683.html (accessed 10 May 2023).

14. Deja M. (2019) Information and knowledge management in higher education institutions: the Polish 5 case. *Online Information Review*, vol. 43, no. 7, pp. 1209–1227. https://doi.org/10.1108/OIR-03-6 2018-0085

15. Yeh Y.M.C. (2005) The implementation of knowledge management system in Taiwans. *Higher Education. Journal of College Teaching & Learning (TLC*), vol. 2, no. 9. https://doi.org/10.19030/tlc.v2i9.1861

16. Maloshonok N. (2016) How using the Internet and multimedia technology in the learning process correlates with student engagement. *Voprosy Obrazovaniya (Educational Studies Moscow)*, no. 4, pp. 59–83. https://doi.org/10.17323/1814-9545-2016-4-59-83

17. Vick T.E., Nagano M.S., Popadiuk S. (2015) Information culture and its influences in knowledge creation: Evidence from university teams engaged in collaborative innovation projects. *International Journal of Information Management*, vol. 35, no. 3, pp. 292–298.

18. Huang M.-H., Chen D.-Z. (2017) How can academic innovation performance in university-industry collaboration be improved? *Technological Forecasting and Social Change*, vol. 123, pp. 210–215. https://doi.org/10.1016/j.techfore.2016.03.024

19. Al-Jedaiah M. (2020) Knowledge management and e-learning effectiveness: Empirical evidence from Jordanian higher education institutions. *International Journal of Emerging Technologies in Learning*, vol. 15, no. 5, pp. 50–62.

20. Altınay F., Altınay M., Dagli G., Altınay Z. (2019) A study of knowledge management systems processes and technology in open and distance education institutions in higher education. *International Journal of Information and Learning Technology*, vol. 36, no. 4, pp. 314–321. https://doi.org/10.1108/IJILT-02-2019-0020

21. Dneprovskaya N., Shevtsova I., Bayaskalanova T., Lutoev I. (2016) Knowledge management methods in online course development. Proceedings of the *15th European Conference on e-Learning (ECEL 2016)*, Charles University, Prague, Czech Republic, 27–28 October 2016, pp. 159–165.

22. Stankosky M. (2005) Advances in knowledge management: University research toward an academic discipline. *Creating the Discipline of Knowledge Management* (ed. M. Stankosky). Elsevier Butterworth-Heinemann, Washington.

23. Vold T., Have H., Kaloudis A. (2020) On work relevance of adult education: A case study narrative. *The Electronic Journal of Knowledge Management*, vol. 18, no. 2, pp. 105–120.

24. Quarchioni S., Paternostro S., Trovarelli F. (2020) Knowledge management in higher education: a literature review and further research avenues. *Knowledge Management Research & Practice*, vol. 20, no. 2, pp. 304–319. https://doi.org/10.1080/14778238.2020.1730717

25. Adeinat I.M., Abdulfatah F.H. (2019) Organizational culture and knowledge management processes: case study in a public university. *VINE Journal of Information and Knowledge Management Systems*, vol. 49, no. 1, pp. 35–53. https://doi.org/10.1108/VJIKMS-05-2018-0041

26. Lombardi R., Massaro M., Dumay J., Nappo F. (2019) Entrepreneurial universities and strategy: the case of the University of Bari. *Management Decision*, vol. 57, no. 12, pp. 3387–3405. https://doi.org/10.1108/MD-06-2018-0690

27. Zanin F., Massaro M., Bagnoli C. (2014) Towards a competitive knowledge management strategy approach in the university setting: The case of ca'Foscari university of Venice. In: *Building a Competitive Public Sector with Knowledge Management Strategy*, pp. 362–382. IGI Global.

28. Wu J., Tian J., Lu L., Weng C., Xiang J. (2016) Surveys on knowledge management and knowledge creation in a Chinese research university. Proceedings of the *2016 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 186–193.

29. Jain N., Gupta V. (2019) The impact of knowledge management system on student performance. *VINE Journal of Information and Knowledge Management Systems*, vol. 49, no. 1, pp. 115–135. https://doi.org/10.1108/VJIKMS-07-2018-0065

30. Zhou Y. (2019) The knowledge management status and improvement strategy of continuing education in university. *International Journal of Intelligent Information and Management Science*, vol. 8, no. 5, pp. 111–115.

31. Tikhomirov V., Dneprovskaya N. (2022) Knowledge management system as a university digital transformation environment on the case of MESI. *Informatsionnoye obshchestvo (Information Society)*, no. 5, p. 44–57 (in Russian). https://doi.org/10.52605/16059921_2022_05_44

32. Isaev D. (2009) Informatization of strategic management in higher education area. *Terra Economicus*, vol. 2, no. 2, pp. 147–152 (in Russian).

33. Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S. (2021) Digital plant: methods of discrete-event modeling and optimization of production characteristics. *Business Informatics*, vol. 15, no. 2, pp. 7–20. https://doi.org/10.17323/2587-814X.2021.2.7.20

34. Zaramenskikh Ye., Kudryavtsev D., Arzumanyan M. (2023) *Enterprise architecture: A textbook for universities*. Moscow: Urait (in Russian).

35. Mohayidin M.G., Azirawani N., Kamaruddin N.M., Margono M.I. (2007) The application of knowledge management in enhancing the performance of Malaysian universities. *The Electronic Journal of Knowledge Management*, vol. 5, no. 3, pp. 301–312.

36. Cranfield D.J., Taylor J. (2008) Knowledge management and higher education: A UK Case study. *The Electronic Journal of Knowledge Management*, vol. 6, no. 2, pp. 85–100.

37. Megnounif A., Kherbouche A. (2020) Knowledge management promising contribution to university performance: Empirical study based on teachers' opinions. *Journal of Information & Knowledge Management*, vol. 19, no. 3, article 2050022. https://doi.org/10.1142/S0219649220500227

38. Chaston I. (2012) Knowledge management systems and open innovation in second tier UK universities. *Australian Journal of Adult Learning*, vol. 52, no. 1, pp. 153–172.

39. Blackman D., Kennedy M. (2009) Knowledge management and effective university governance. *Journal of Knowledge Management*, vol. 13, no. 6, pp. 547–563. https://doi.org/10.1108/13673270910997187

40. Ramírez-Córcoles Y., Manzaneque-Lizano M. (2015) The relevance of intellectual capital disclosure: empirical evidence from Spanish universities. *Knowledge Management Research & Practice*, vol. 13, pp. 31–44. https://doi.org/10.1057/kmrp.2013.27

41. Fernández-López S., Rodeiro-Pazos D., Calvo N., Rodríguez-Gulías M.J. (2018) The effect of strategic knowledge management on the universities' performance: an empirical approach. *Journal of Knowledge Management*, vol. 22, no. 3, pp. 567–586. https://doi.org/10.1108/JKM-08-2017-0376

42. Tan C.N. (2016) Enhancing knowledge sharing and research collaboration among academics: the role of knowledge management. *Higher Education*, vol. 71, pp. 525–556. https://doi.org/10.1007/s10734-015-9922-6

43. Kudryavtsev D., Gavrilova T. (2017) From anarchy to system: A novel classification of visual knowledge codification techniques. *Knowledge and Process Management*, vol. 24, pp. 3–13. https://doi.org/10.1002/kpm.1509

44. Gavrilova T.A., Alsufyev A.I., Grinberg E.Y. (2017) Knowledge visualization: Critique of the St. Gallen School and an analysis of contemporary trends. *Business Informatics*, no. 3 (41), pp. 7–19. https://doi.org/10.17323/1998-0663.2017.3.7.19

45. Filatov V.P. (2016) Knowledge. *Great Russian Encyclopedia. Electronic version*. Available at: https://old.bigenc.ru/philosophy/text/1994642 (accessed 10 May 2023) (in Russian).

46. Bibi G., Padhi M., Dash S.S. (2021) Theoretical necessity for rethinking knowledge in knowledge management literature. *Knowledge Management Research & Practice*, vol. 19, no. 3, pp. 396–407. https://doi.org/10.1080/14778238.2020.1774433

47. Davenport T., Prusak L. (1998) *Working knowledge: How organizations manage what they know*. Harvard Business School Press.

48. Bratianu C. (2018) Intellectual capital research and practice: 7 myths and one golden rule. *Management and Marketing*, vol. 13, no. 2, pp. 859–879. https://doi.org/10.2478/mmcks-2018-0010

49. Kuleshova A.V., Podvoyskiy D.G. (2018) Paradoxes of publication activity in the field of contemporary Russian science: Genesis, diagnosis, trends. *Monitoring of Public Opinion: Economic and Social Changes*, no. 4, pp. 169–210 (in Russian). https://doi.org/10.14515/monitoring.2018.4.10

50. Mindeli L.E., Ostapyuk S.F., Fetisov V. (2016) The role of monitoring in the strategic planning of scientific and technological development. *Inovatsii* (*Innovations)*, no. 3(245), pp. 25–32 (in Russian).

51. Kurlov A.B., Petrov V.K. (2014) *Methodology of information analytics*. Moscow: Prospekt (in Russian).

52. Dneprovskaya N.V., Shevtsova I.V. (2019) Prospects for university's open educational resources. *Vysshee obrazovanie v Rossii (Higher Education in Russia)*, vol. 28, nos. 8–9, pp. 110–118 (in Russian). https://doi.org/10.31992/0869-3617-2019-28-8-9-110-118

53. Samulson P. (2020) *What economists think about: Conversations with Nobel laureates*. Moscow: Alpina Publisher (in Russian).

54. Telnov Yu.F., Kazakov V.A., Trembach V.M. (2020) Developing a knowledge-based system for the design of innovative product creation processes for network enterprises. *Business Informatics*, vol. 14, no. 3, pp. 35–53. https://doi.org/10.17323/2587-814X.2020.3.35.53

55. Kidwell J.J., Vander Linde K.M., Johnson S.L. (2000) Applying corporate knowledge management practices in higher education. *Educause Quartery*, no. 4, pp. 28–33.

56. Nonaka I., Toyama R., Hirata T. (2008) *Managing flow. A process theory of the knowledge-based firm*. New York: Palgrave Macmillan. https://doi.org/10.1057/9780230583702

57. Chan I. (2017) Knowledge management hybrid strategy with people, technology and process pillars. In: *Knowledge Management Strategies and Applications*. https://doi.org/10.5772/intechopen.70072

58. Maier R. (2007) *Knowledge management systems. Information and communication technologies for knowledge management*. 3d edition. Springer.

59. Yang J., Roy S., Goel V. (2022) Who engages in explicit knowledge creation after graduation? Evidence from the alumni impact survey of a large Canadian public university. *Studies in Higher Education*, vol. 47, no. 1, pp. 116–130. https://doi.org/10.1080/03075079.2020.1732907

60. Dneprovskaya N.V., Shevtsova I.V. (2020) Open educational resources in the development of digital learning environment. *Vysshee obrazovanie v Rossii (Higher Education in Russia)*, vol. 29, no. 12, pp. 144–155 (in Russian). https://doi.org/10.31992/0869-3617-2020-29-12-144-155

61. Titova L.N. (2008) *The path of success and failure: the strategic development of Russian universities*. Moscow: MAKS Press (in Russian).

62. Dang Q.T., Jasovska P., Rammal H.G., Schlenker K. (2019) Formal-informal channels of university-industry knowledge transfer: the case of Australian business schools. *Knowledge Management Research & Practice*, vol. 17, pp. 384–395. https://doi.org/10.1080/14778238.2019.1589395

63. Petrova G.I., Smokotin V.M., Kornienko A.A., Ershova I.A., Kachalov N.A. (2015) Knowledge management as a strategy for the administration of education in the research university. *Procedia – Social and Behavioral Sciences*, vol. 166, pp. 451–455. https://doi.org/10.1016/j.sbspro.2014.12.552

64. Compagnucci L., Spigarelli F. (2020) The third mission of the university: A systematic literature review on potentials and constraints. *Technological Forecasting and Social Change*, vol. 161, article 120284. https://doi.org/10.1016/j.techfore.2020.120284

65. Ardito L., Ferraris A., Petruzzelli A.M., Bresciani S., Del Giudice M. (2019) The role of universities in the knowledge management of smart city projects. *Technological Forecasting and Social Change*, vol. 142, pp. 312–321. https://doi.org/10.1016/j.techfore.2018.07.030

66. Raman L. (2017) Application of knowledge management in university research and higher education: An experiment with communities of practice (COP). *Enhancing Academic Research with Knowledge Management Principles*, IGI Global, pp. 92–114.

67. Chirikov I. (2013) Research universities as knowledge networks: The role of institutional research. *Studies in Higher Education*, vol. 38, pp. 456–469.

68. Grimaldi D., Fernandez V. (2017) The alignment of university curricula with the building of a Smart City: A case study from Barcelona. *Technological Forecasting and Social Change*, vol. 123, pp. 298–306. https://doi.org/10.1016/j.techfore.2016.03.011

69. Teece D.J. (2009) Obtaining economic benefits from knowledge and competencies. *Innovative development: economics, intellectual resources, knowledge management* (ed. Z. Milner), pp. 244–262. Moscow: Infra M (in Russian).

70. Arbersman S. (2012) *The Half-Life of facts: Why everything we know has an expiration date*. Penguin Group.

71. Novikov D.A. (2016) *Cybernetics: navigator. History of cybernetics, current state, development prospects*. Moscow: Lenand (in Russian).

72. Serenko A., Abubakar A.M. (2022) Antecedents and consequences of knowledge sabotage in the Turkish telecommunication and retail sectors. *Journal of Knowledge Management*, vol. 27, no. 5, pp. 1409–1435. https://doi.org/10.1108/JKM-01-2022-0029

73. Price D.J. (1963) *Little Science, Big Science*. N.Y.: Columbia University Press.

74. Reinsel D., Gantz J., Rydning J. (2018) *The digital of the world: From edge to core*. Available at: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (accessed 10 May 2023).

About the authors

Natalia V. Dneprovskaya

Dr. Sci. (Econ);

Associate Professor, Department of Business Informatics, Graduate School of Business, HSE University, 26–28, Shabolovka Street, Moscow 119049, Russia;

Professor, Department of Business Informatics, Financial University under the Government of Russian Federation, 49/2, Leningradskiy Prospect, Moscow 125167, Russia;

E-mail: ndneprovskaya@hse.ru

ORCID: 0000-0002-9600-8474

Inessa V. Shevtsova

Cand. Sci. (Econ);

Associate Professor, Department of Mathematical Methods and Information Technologies in Management, School of Public Administration, Lomonosov Moscow State University, 27, build 4, Lomonosov Prospect, Moscow 119991, Russia;

E-mail: Shevtsova@spa.msu.ru

ORCID: 0000-0003-1518-6277

DOI: 10.17323/2587-814X.2023.2.41.54

Decision support technology for a seller on a marketplace in a competitive environment*

Mikhail G. Matveev
E-mail: mgmatveev@yandex.ru

Natalya A. Aleynikova
E-mail: balbashovan@mail.ru

Maria D. Titova
E-mail: 29_06_titova@mail.ru

Voronezh State University
Address: 1, Universitetskaya pl., Voronezh 394018, Russia

Abstract

This article deals with the problem of improving the effectiveness of a marketplace. The stakeholders of a marketplace are buyers and sellers. The objects are the aggregate of homogeneous products. The effectiveness of the trading platform, which can be characterized by the number of transactions made, will depend on how sufficiently the sellers put up offers. The paper looks at mathematical models to support the decision-making of the seller in making such offers. Focusing not only on the buyer demand but also on the presence of competitors on the site is a distinguishing feature of the models. To describe the competition, the apparatus of game theory is offered, namely the normal form of the game with a bimatrix model with two players: the seller – customer of service and the coalition of other sellers. To match offer and demand, as well as to find the probability of a transaction, fuzzy set theory and aggregation using the Choquet integral are used.

* The article is published with the support of the HSE University Partnership Programme

Keywords: electronic trading platform, marketplace, homogeneous product, linguistic variable, aggregation operator, Choquet integral, bimatrix game, solution in mixed strategies

Citation: Matveev M.G., Aleynikova N.A. Titova M.D. (2023) Decision support technology for a seller on a marketplace in a competitive environment. *Business Informatics*, vol. 17, no. 2, pp. 41–54. DOI: 10.17323/2587-814X.2023.2.41.54

Introduction

The development of e-commerce is naturally accompanied by an increase in the level of automation of business processes. At the same time, the emphasis is shifting from automating routine document management processes towards automating more complex decision support processes by e-commerce entities [1–3]. An important aspect of the seller's work in the formation of a product offer is the indicator of the liquidity of the product, which measures the possibility of making a deal with the buyer for a given product offer [4]. Support for the buyer in finding suitable products and sellers is already developing. Examples of the relevant tools can be found in [5]. Automation of this type of business processes for the seller is only in its infancy and there are not enough tools to help the seller [5]. Our work is aimed at increasing the liquidity of the seller's product offer on the marketplace. We propose a numerical measure of the liquidity of a product offer and a new mathematical model for the formation of such an offer (seller's strategy), which in a competitive environment will have high liquidity.

In [6–9], a mathematical model was proposed for formalizing the activities of subjects of an electronic trading platform (ETP) of the marketplace type based on matching demand and acceptable supply [10]. Consumer demand and acceptable seller offers are set parametrically within a single commodity classifier in the form of fuzzy characteristic properties of a homogeneous group of goods. The vector of values of the characteristic properties of a product defines its differentiation in a group of homogeneous products in terms of a set of characteristics (price, quality, and other characteristic properties) specified both in numerical and categorical formats. The main result of this research is a numerical measure of liquidity, which is formalized as a correspondence between the demand and the allowable supply of the seller for a given set of their characteristic properties (not only for the price, as it takes place in most cases). This makes it possible to compare (on a numerical scale) the correspondences of various product offers to market conditions and select the offer with the best match.

However, [6–9] did not take into account the fact that the seller, putting forward his proposals, is in a competitive environment and the proposals of other sellers can significantly change liquidity. The competition of sellers in the online trading market is appropriately described by game-theoretic models, examples of which are shown in [11–18]. When developing a game-theoretic model, the key tasks are setting the equilibrium conditions, interpreting and calculating the payoff matrix elements, and interpreting the players' strategies. In these papers, game-theoretic modeling of a duopoly is considered with the choice of an equilibrium solution (in the Nash sense) and the interpretation of the outcome of the game in the form of profit. A characteristic feature of these works is the choice of strategies not only in the form of the price of the goods, but also taking into account their quality. Other characteristic properties of the goods are not taken into account.

We believe that the calculated maximum profit is not always achievable in real market conditions. Such a transaction may not take place due to the discrepancy between some properties of the optimal product offer in terms of price and the demand or acceptable offer of the seller. A more realistic goal is a trade-off between maximizing potential profits and being able to conclude a deal. The search for such a compromise consists in maximizing the correspondence between demand and the acceptable product offer along the vector of the characteristic properties of the product, including the price characteristic.

The game situation arises in the form of a seller's competition with a set of other sellers operating in the ETP market and considered as a generalized competitor. The game strategies are set by the acceptable variants of the seller's product offer represented by the corresponding vectors of characteristic properties. The resulting correspondence is interpreted as the subjective probability of making a deal and is proposed as the outcome of the game in game-theoretic modeling of competition in the duopoly market.

The purpose of is paper is to obtain game-theoretic models that make it possible for the seller to form rational offers from an acceptable set of interchangeable types of homogeneous goods.

1. Formalization of the activities of the subjects in ETP

Let the objects of trade on the marketplace be sets of homogeneous goods. A homogeneous product is a set of its interchangeable types, for example, a set of cars of various brands. The types differ in the values of the characteristic properties of a given product, given by the vector of values of the corresponding parameters. Such parameters may be commercial, technical and other possible properties or characteristics of the goods. Let j be the index of the type of a homogeneous product ($j=\overline{1,J}$), its vector of characteristic properties will be denoted as

$$q_j = (q_j^1; \dots; q_j^n; \dots; q_j^N), \tag{1}$$

Here each n coordinate can take values both on a quantitative and on a qualitative scale.

Consumer demand for some kind of homogeneous product can also be formalized in the form of vector g (product characteristics) which is structurally identical to vector q. As a rule, the desires of buyers are vague and approximate. For example, a buyer needs a car with an engine power of 100 to 150 hp. with a cost from the price interval specified by the buyer. Moreover, some values from these intervals are more desirable, some less. Such demand of the buyer can be provided due to the variety of homogeneous goods with different characteristics. It is convenient to represent the coordinates of the consumer demand vector as linguistic variables [19]

$$\tilde{g}_k = (\tilde{g}_k^1; \dots; \tilde{g}_k^n; \dots; \tilde{g}_k^N), \tag{2}$$

where $k=\overline{1,K}$ is the buyer index. The names of linguistic variables coincide with the names of the corresponding characteristic parameters of the description of the type of a homogeneous product. Each variable has piecewise linear membership functions $f_g(x) \in [0;1]$ whose carriers $x_{\min} \le x \le x_{\max}$ reflect the choice of the buyer, and the function values determine the level of his preference [8].

The seller's offer (strategy) represents a specific type of product and is given by a vector similar to (1). It is required to choose the values of the coordinates of q_j in such a way as to ensure high liquidity of the transaction. At the same time, it is necessary to pay attention not only to consumer demand, but also to the permissible possibilities of the seller himself. The values of q_j are limited by the financial and commodity stocks of the seller as well as its functionality. It is assumed that the seller is able to evaluate their functional cost constraints (FCC) and seeks to formulate their offer in such a way as to obtain the maximum compliance with these limitations. FCC are set as admissible intervals of parameter values with the construction of a membership function over each interval reflecting the preferences of its values. Then the seller's restrictions can be represented by the following vector of linguistic variables

$$\tilde{q} = (\tilde{q}^1; \ldots; \tilde{q}^n; \ldots; \tilde{q}^N) \qquad (3)$$

with the same names of fuzzy characteristics as the demand vector, but with their own membership functions $f_{\tilde{q}^n}(x) \in [0;1]$. Note, that by forming the price membership function in this way, one can set the seller's desire to sell the goods at a high price, i.e., get maximum profit.

Since it is a cumbersome task to track the interaction between the seller and each buyer, we proposed to pass to the generalized consumer demand which can be presented as (2). The membership function for each generalized vector coordinate (2) can be found as a weighted sum:

$$f_{\tilde{g}^n}(x) = \sum_{k=1}^{K} f_{\tilde{g}_k^n}(x) \cdot w_k \in [0;1]; \quad n = \overline{1,N}, \qquad (4)$$

where $w_k = \dfrac{v_k}{v}$ is the corresponding weight the buyer k, calculated as the ratio of the volume of goods requested by the buyer k to the total volume of commodity demand $v = \sum_{k=1}^{K} v_k$. Further, we will consider the interaction of the seller with a generalized consumer demand but not with a specific buyer. The validity of this approach is shown in [8].

The range of values of the components of the seller's proposals vector that simultaneously satisfy the generalized demand and the FCC is defined as the intersection of the graphs of the corresponding membership functions for each component pair of the demand vector and the FCC vector [6]. Denote this vector as $\tilde{s}$. The membership function of the intersection with respect to the component n is determined as

$$f_{\tilde{s}}^n(x) = \min(f_{\tilde{g}}^n(x); f_{\tilde{q}}^n(x)). \qquad (5)$$

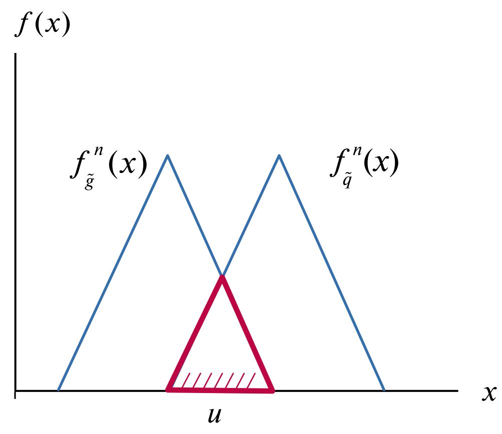

Fig. 1. Graphic illustration of determining the membership function by the component n of the vector $\tilde{s}$.

Graphical illustration of a possible intersection is shown in *Fig. 1.*

Carrier $u \subseteq x$ of function $f_{\tilde{s}}^n(x)$ determines the allowable values of the property n of the product; the values of the function determine the degree of compliance of the allowable values of the properties with the generalized demand and the FCC of the seller. The boundaries of the carrier, within which the seller chooses the allowable values of the property n of the product, are calculated as

$$L(f_{\tilde{s}}^n) = \max(L(f_{\tilde{g}}^n); L(f_{\tilde{q}}^n)), \qquad (6)$$

$$R(f_{\tilde{s}}^n) = \min(R(f_{\tilde{g}}^n); R(f_{\tilde{q}}^n)), \qquad (7)$$

where L and R are left and right boundaries of the carriers, respectively.

Substituting value x^* of the property n of a specific type of the product into function (5), we obtain the degree of local correspondence $f_s^n(x^*)$ for the property n. To obtain the conformity of the product to the allowable values for the entire set of properties, it is necessary to aggregate local matches:

$$f_s = \underset{n}{\text{agr}}[f_s^n(x^*)] \in [0;1], \qquad (8)$$

where agr is the aggregation operator for the local matches.

The choice of the aggregation operator is carried out taking into account the specifics of the subject area and the corresponding properties of various operators. An overview of aggregation operators, their properties, and recommendations for use can be found in [20–23]. Particularly, the required property of the aggregation operator is the following expression: agr: $[0, 1]^n \to [0, 1]$.

As an aggregation operator, we suggest the discrete Choquet integral with respect to a fuzzy measure [24] which is used when the aggregation result is impacted by the value of each of the properties of the product, as well as if it is necessary to take into account the interaction of properties with each other. An example is the interaction of price and quality of goods. Let us designate the set of product properties as the set of their indices $M = \{n_i\}$, $i = 1, 2, \ldots, N$, and let m be an arbitrary subset of M. The interaction of properties can be taken into account due to the fact that when calculating the Choquet integral, the λ-fuzzy Sugeno measure is used, which is specified on set M and expresses the subjective weight or significance of each subset of properties. It is defined as follows [25]:

$$\varphi(m) = \frac{\prod_{n \in m}\left(1+\lambda\varphi_n\right)-1}{\lambda} \in [0;1],\ m \subseteq M, \tag{9}$$

where φ_n are the coefficients of importance (weights) of individual properties, which can be determined either using special methods or set by an expert [26–29]. The value of λ can be found by solving the following equation

$$\lambda + 1 - \prod_{n=1}^{N}\left(1+\lambda\varphi_n\right) = 0, \tag{10}$$

with the following condition $\lambda > -1$, $\lambda \neq 0$.

Then the Choquet integral determining the aggregated correspondence is calculated as follows:

$$f_s = \underset{n}{\mathrm{agr}}\,[f_s^n(x^*)] = \\ = \sum_{n=1}^{N}\left(f_s^{(n)}(x^*) - f_s^{(n-1)}(x^*)\right) \times \\ \times \varphi\left(m \big/ f_s^i(x^*) \geq f_s^{(n)}(x^*),\ i \in m\right), \tag{11}$$

where $f_s^{(1)}(x^*), f_s^{(2)}(x^*), \ldots, f_s^{(N)}(x^*)$ is a permutation of elements $f_s^1(x^*), f_s^2(x^*), \ldots, f_s^N(x^*)$ such that

$$f_s^{(1)}(x^*) \leq f_s^{(2)}(x^*) \leq \ldots \leq f_s^{(N)}(x^*),\ f_s^{(n)}(x^*) = 0.$$

We assume, that at $f_s = 0$ the transaction will not take place, but at $f_s = 1$, the transaction will definitely take place. Then the correspondence $f_s \in [0; 1]$ can be interpreted as the subjective probability of making a deal, that is, a numerical measure of liquidity. The concept of subjective probability (hereinafter referred to as probability), based on expert judgment and the use of mathematical methods for processing this judgment are widely used in economic applications, for example in [30, 31].

In the following, the probability of a transaction for an arbitrary seller for the specific product j calculated by formula (8) will be denoted by p_j taking into account that p_j is a function of the strategy q_j of the seller (his specific offer) and is considered as a measure of the liquidity of the product offer.

2. The game model for selecting the seller's offer on the marketplace under competition

So far, we have determined the probability of a trade for a seller provided that there is only one seller on the market. The presence of other competing sellers in the market can significantly change this probability.

Consider the interaction of a seller, a service applicant, with a set of other sellers of some homogeneous product. Let the alternative variants of the seller's product offer (strategy) be represented by the corresponding vectors with different values of the characteristic properties of a homogeneous product. The choice on a subset of alternative strategies should be made taking into account the competitive offers of the set of other sellers. If the entire set of sellers on the ETP is large enough, then the chosen strategy of one seller that does not dominate in terms of volume will have practically no effect on the choice of sellers from the set. On the contrary, the generalized offer of the set of sellers will significantly change the probability of a transaction of one seller. The seller chooses his strategy for a certain long time period, during which the generalized offer of the set of sellers changes randomly, which leads to a game situation corresponding to the conditions of the bimatrix game.

Two players are considered: a seller (a service applicant) with his own set of strategies $(q_j, j=\overline{1,J})$ and a certain generalized seller, composed of a set of sellers, with his own strategies represented by variants of a generalized offer $(\hat{q}_t, t=\overline{1,T})$.

The payment matrix of the seller is represented by the probabilities of transactions p_{jt}. The payment matrix of the generalized seller is represented by his transaction probabilities $\hat{p}_{jt}$. Probabilities $\hat{p}_{jt}$ can be defined as follows. A component-by-component generalization of sellers' proposals from the set is performed. Assuming no dominance of individual sellers on the ETP, the generalization for each component of the supply vectors is determined as the average value. The matching of the generalized offer and the generalized demand for each component is performed in the same way as it was done for the individual seller and the generalized demand. The results of the obtained correspondences are aggregated following (8) using the Choquet integral (11). Aggregate matches are considered as probabilities $\hat{p}_{jt}$ of the generalized offer that do not depend on the offers of the seller, i.e., Those elements $\hat{p}_{jt} = \hat{p}_t$.

The formation of generalized strategies is carried out randomly under the assumption of a random nature of the values of the characteristic properties. The average and standard deviation of each generalized property is determined from a sample of values of the characteristic properties of the offers in a set of sellers. Then a set of strategies, for example, under the assumption that the distributions of random variables are normal, can be obtained using a standard random number generator.

Probability p_{jt} of a transaction for a seller in a competitive environment is obviously a function that depends both on probability of making a transaction p_j with no competition, and on probabilities $\hat{p}_{jt}$, that is, $p_{jt} = p_{jt}(p_j;\hat{p}_{jt})$. At the same time, we suggest that if probability $\hat{p}_{jt}$ of selling the goods for the generalized seller is less than the similar probability p_j for the seller, then the buyer will buy it from the seller with probability $p_{jt}(p_j;\hat{p}_{jt}) = p_j$. If the probability of selling the product by the generalized seller exceeds p_j for the seller, then the value should decrease, since, most likely, buyers will prefer the goods of the set of sellers with more attractive characteristic values. The probability of sale in this case for the seller can be determined using the following reasoning. Consider a complete group of incompatible events, which includes three situations. The first, when the buyer buys the product from the generalized seller with probability $\hat{p}_{jt}$; the second, when the product is bought from the seller (we have to find this probability $p_{jt}(p_j;\hat{p}_{jt})$); and the third, when the product is not bought from either the generalized seller or the regular seller. The probability of the last situation can be defined as $(1-p_j)(1-\hat{p}_{jt})$. From the normalization condition

$$p_{jt} + \hat{p}_{jt} + (1-p_j)(1-\hat{p}_{jt}) = 1 \tag{13}$$

we have

$$\begin{aligned} p_{jt} &= 1-\hat{p}_{jt} - (1-p_j)(1-\hat{p}_{jt}) = \\ &= p_j - p_j\hat{p}_{jt} = p_j(1-\hat{p}_{jt}). \end{aligned} \tag{14}$$

Then the probability of a transaction for the seller in a competitive environment is:

$$p_{jt}(p_j;\hat{p}_{jt}) = \begin{cases} p_j - p_j\hat{p}_{jt}, & p_j \le \hat{p}_{jt}, \\ p_j, & p_j > \hat{p}_{jt}. \end{cases} \quad (15)$$

For example, the seller has a probability of selling 0.21, and the generalized seller has 0.65. Then the probability of selling under competition from the seller is

$$p_{jt}(0.21; 0.65) = 0.21 \cdot (1 - 0.65) = 0.0735.$$

Thus, it is assumed that finite sets of the seller q_j and generalized seller $\hat{q}_t$ strategies are given. The values of the payoff function are given as a bimatrix with elements $A_{jt} = (p_{jt}, \hat{p}_{jt})$. The solution to the problem consists in the rational choice of the strategy (offer) of the seller with a random change in the strategies of the generalized seller.

As a criterion of rationality, we consider the Nash concept of equilibrium [32, 33]. A set of mixed strategies $q^* = (q_j^*; \hat{q}_t^*)$ is called as the Nash equilibrium situation in mixed strategies if the choice of any side of the mixed strategy other than one which leads to one of the inequalities

$$V_1(p_{jt};\hat{p}_{jt}^*) \le V_1(p_{jt}^*;\hat{p}_{jt}^*) \quad (16)$$

or

$$V_2(p_{jt}^*;\hat{p}_{jt}) \le V_2(p_{jt}^*;\hat{p}_{jt}^*). \quad (17)$$

where V_1, V_2 are the mathematical expectations of the winnings of the seller and the generalized seller, respectively.

The above inequalities indicate that a deviation from the equilibrium situation by one side cannot increase its payoff.

3. Numerical example

Using the example of one seller and three buyers, we calculate the equilibrium mixed strategy of the seller. Let a homogeneous product be characterized by three parameters (properties). The values of the parameters characterizing the FCC of the seller are given in the form of triangular fuzzy numbers in *Table 1* indicating the left boundary of the carrier, the mode and the right boundary of the carrier, respectively. The carrier is normalized in the range from 0 to 1.

Table 1.

FCC of the seller given as triangular fuzzy numbers

1-st parameter	2-nd parameter	3-rd parameter
(0.4; 1; 1)	(0.2; 0.4; 1)	(0.5; 1; 1)

For buyers, the initial data is given in *Table 2.*

To determine the values of the membership function of the generalized consumer demand for each parameter separately, we use (4). To do this, we divide interval [0, 1] into 10 parts and at each point determine the value of the membership function of the generalized consumer demand $f_{\tilde{g}^n}(x), n = \overline{1,3}$.

For example, at x = 0.3 the value of the membership function by the first parameter for the first buyer is 0.5, for the second 0.25, for the third 0.833. Weights are w_1 = 10 : (10 + 4 + 6) = 0.5, w_2 = 0.2, w_2 = 0.3. Then, using (4) we have

$$f_{\tilde{g}^1}(0.3) = 0.5 \cdot 0.5 + 0.25 \cdot 0.2 + 0.833 \cdot 0.3 = 0.55.$$

In *Fig. 2* we show the membership function for the generalized demand by each parameter.

Next, we determine the boundaries and membership functions of the components of the fuzzy vector of the seller's proposals that simultaneously satisfy the generalized demand and FCC following to (5)–(7). In this case, membership functions $f_{\tilde{s}}^n(x), n = \overline{1,3}$ by each parameter are shown in *Fig. 3.*

Let the seller offer three goods (strategies) with parameters written as vectors, the coordinates of which are normalized q_1 = (0.5; 0.4; 0.8), q_2 = (0.4; 0.6; 0.9),

Table 2.

Characteristics of consumer demand given by triangular fuzzy numbers

Buyer	The volume of goods requested by the buyer	1-st parameter	2-nd parameter	3-rd parameter
1	10	(0.2; 0.4; 1)	(0.2; 0.5; 1)	(0; 0.4; 0.6)
2	4	(0.2; 0.6; 1)	(0.5; 1; 1)	(0.4; 0.4; 1)
3	6	(0.2; 0.2; 0.8)	(0.2; 0.6; 1)	(0; 0; 1)

q_3 = (0.8; 0.5; 0.5). Then, substituting the corresponding coordinates in functions $f_{\bar{s}}^n(x), n=\overline{1,3}$, we obtain the local correspondences $f_s^n(x^*), n=\overline{1,3}$. For the first strategy, we get the vector of local correspondences (0.167; 0.483; 0.1267); for the second one we have (0; 0.267; 0.725); for the third one we have (0.267; 0.725; 0).

Now, we aggregate the local correspondences using the Choquet integral (9)–(11). To determine λ from equation (10), we set the coefficients of importance of the product parameters equal to $\varphi_1 = 0.3$, $\varphi_2 = 0.6$, $\varphi_3 = 0.2$. Then, equation (10) takes the form

$$\lambda + 1 - (1 + \lambda \cdot 0.3) \cdot (1 + \lambda \cdot 0.6) \cdot (1 + \lambda \cdot 0.2) = 0,$$

at $\lambda > -1, \lambda \neq 0$. Root of the equation is $\lambda = -0.286$.

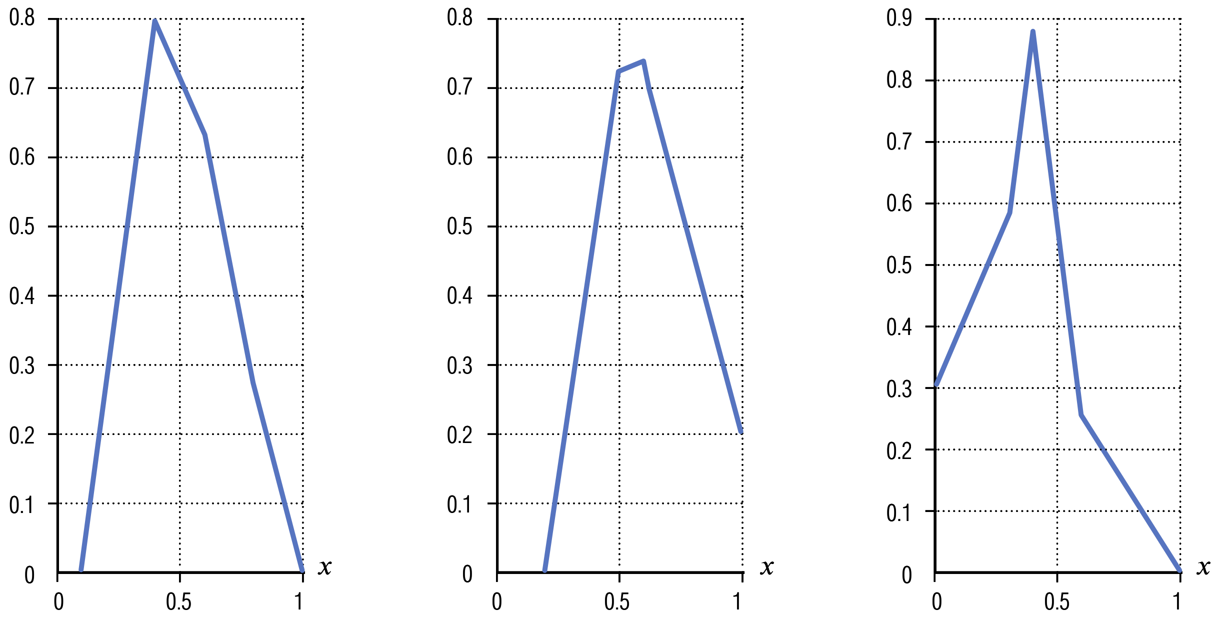

Fig. 2. The membership function of the generalized demand by each parameter.

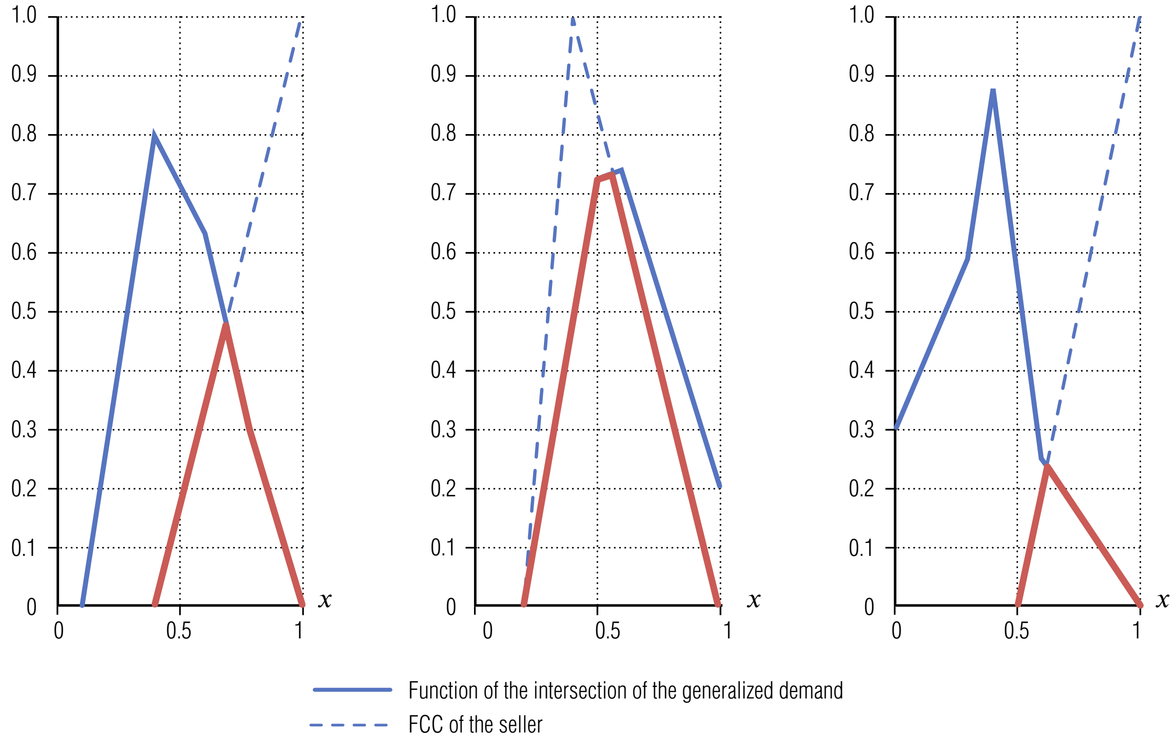

Fig. 3. Membership functions of the intersection of the generalized demand and the FCC of the seller by each parameter.

Determine the Choquet integral for the first strategy of the seller. Arrange the coordinates of the vector of local correspondences in ascending order $f_s^{(1)} = f_s^3 = 0.1267,\ f_s^{(2)} = f_s^1 = 0.167,\ f_s^{(3)} = f_s^2 = 0.438.$ Then,

$$f_s = (f_s^{(1)} - f_s^{(0)}) \cdot \varphi(\{1,2,3\}) + $$
$$+ (f_s^{(2)} - f_s^{(1)}) \cdot \varphi(\{1,2\}) + (f_s^{(3)} - f_s^{(2)}) \cdot \varphi(\{2\}),$$

where

$$\varphi(\{1, 2, 3\}) =$$
$$= \frac{(1-0.286\cdot 0.3)\cdot(1-0.286\cdot 0.6)\cdot(1-0.286\cdot 0.2)-1}{-0.286} = 1$$

(equation (9)),

$$\varphi(\{1,2\}) = \frac{(1-0.286\cdot 0.3)\cdot(1-0.286\cdot 0.6)-1}{-0.286} = 0.8485,$$

and $\varphi(\{2\}) = 0.6.$

As a result,

$$f_s = \left(f_s^{(1)} - f_s^{(0)}\right)\cdot\varphi(\{1,2,3\}) +$$
$$+\left(f_s^{(2)} - f_s^{(1)}\right)\cdot\varphi(\{1,2\}) + \left(f_s^{(3)} - f_s^{(2)}\right)\cdot\varphi(\{2\}) =$$
$$= (0.1267-0)\cdot 1 + (0.167-0.1267)\cdot 0.8485 +$$
$$+ (0.483-0.167)\cdot 0.6 = 0.35.$$

Similarly, for the second strategy, the Choquet integral is $f_s = 0.41$, for the third strategy is $f_s = 0.501$.

We have the probabilities of the transaction for the seller by each product:

$$p_1 = 0.35,\ p_2 = 0.41,\ p_3 = 0.501. \qquad (18)$$

Now consider a generalized seller which is competitor for the seller offering a homogeneous product.

Suppose that a generalized seller has three strategies $\hat{q}_t, t=\overline{1,3}$, and the probabilities of a deal for each of them are $\hat{p}_1 = 0.414$, $\hat{p}_2 = 0.374$, $\hat{p}_3 = 0.264$.

Probabilities (18) for the seller has been obtained in the case when there is no competition. Under competition for seller, it is necessary to obtain the values of (15) that are substituted into the bimatrix as his payoffs.

Because $p_1 = 0.35 < \hat{p}_1 = 0.414$, then, using (15) $p_{11} = 0.35 - 0.35 \cdot 0.414 = 0.205$. At the same time $p_{13} = 0.35$ because $p_1 = 0.35 > \hat{p}_3 = 0.246$. Other winnings of the seller are calculated similarly.

As a result, the bimatrix game takes the form:

$$\begin{pmatrix} & \hat{q}_1 & \hat{q}_2 & \hat{q}_3 \\ q_1 & (0.205; 0.414) & (0.219; 0.374) & (0.35; 0.264) \\ q_2 & (0.240; 0.414) & (0.41; 0.374) & (0.41; 0.264) \\ q_3 & (0.501; 0.414) & (0.501; 0.374) & (0.501; 0.264) \end{pmatrix}. \quad (19)$$

Note, that we are only interested in the choice of the seller strategy.

The solution to the bimatrix game using the Nash equilibrium strategy in the situation of mixed strategies [32, 33] gives the equilibrium in pure strategies for the seller with the price of the game 0.501. That is, the seller should put up for sale the first product with characteristics (0.5; 0.4; 0.8); the probability of sale under competition, will be equal to 0.501. For the generalized seller, equilibrium is reached in pure strategies with the price of the game 0.414.

Note that the method for calculating the Nash equilibrium is quite cumbersome and its computational complexity increases with the increase in the dimension of the problems being solved.

The above result can be obtained using a simpler technique. It is shown in [34] that in the game 2 × 2 the same result can be obtained by each side only based on their own payoff matrices. To do this, it is necessary to split the bimatrix game into two ordinary zero-sum matrix games. Each player can calculate from the matrix of his own payoffs the optimal average payoff, which coincides with the payoff in the equilibrium situation; using his own matrix, the player can find the optimal strategy of the other player, but not his own. In our case, consider the matrices 3 × 3:

$$A = \begin{pmatrix} 0.205 & 0.219 & 0.35 \\ 0.240 & 0.41 & 0.41 \\ 0.501 & 0.501 & 0.501 \end{pmatrix}$$

$$\text{and } B = \begin{pmatrix} 0.414 & 0.414 & 0.414 \\ 0.374 & 0.374 & 0.374 \\ 0.264 & 0.264 & 0.264 \end{pmatrix}. \quad (20)$$

Find the solution to the matrix game in mixed strategies for the generalized seller using matrix A. To do this, we denote by $(\alpha_1, \alpha_2, \alpha_3)$ the vector of probabilities of applying the corresponding strategies by the generalized seller, and by v – the price of the game. Substituting

$$x_1 = \frac{\alpha_1}{v},\ x_2 = \frac{\alpha_2}{v},\ x_3 = \frac{\alpha_3}{v}.$$

we compose a linear programming problem:

$$F = x_1 + x_2 + x_3 \to \max,$$

under restrictions

$$\begin{cases} 0.205x_1 + 0.219x_2 + 0.35x_3 \le 1, \\ 0.240x_1 + 0.41x_2 + 0.41x_3 \le 1, \\ 0.501x_1 + 0.501x_2 + 0.501x_3 \le 1. \end{cases}$$

Solutions to this problem are $x_1 = 1.996$, $x_2 = 0$ and $x_3 = 0$. The game price is

$$v = \frac{1}{x_1 + x_2 + x_3} = \frac{1}{1.996} = 0.501.$$

When passing to probabilities, we get $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_3 = 0$. Therefore, the solution in pure strategies for the generalized seller is (1; 0; 0). The game price for the seller is 0.501. The pure strategies obtained for the generalized seller and the game price for the seller coincide with the strategies and the game price that were found when solving the bimatrix model using the Nash equilibrium technique.

Obtain the solution to the matrix game in mixed strategies for the seller, using matrix *B*. To do this, we denote by $(\beta_1, \beta_2, \beta_3)$ the vector of probabilities of applying the corresponding strategies by the seller, and by v – the price of the game. Then, the linear programming problem for solving the game in mixed strategies, taking into account the change of variables, as in the previous model, has the form:

$$F = x_1 + x_2 + x_3 \rightarrow \max,$$

under restrictions

$$\begin{cases} 0.414x_1 + 0.414x_2 + 0.414x_3 \le 1, \\ 0.374x_1 + 0.374x_2 + 0.374x_3 \le 1, \\ 0.264x_1 + 0.264x_2 + 0.264x_3 \le 1. \end{cases}$$

Solutions to this problem are $x_1 = 2.4$, $x_2 = 0$, $x_3 = 0$. The game price is

$$v = \frac{1}{x_1 + x_2 + x_3} = \frac{1}{2.4} = 0.414.$$

Therefore, the solution in pure strategies for the seller is (1; 0; 0). That is, in a competitive environment, the optimal strategy for him is the first one. The price of the game for the generalized seller is 0.414. The resulting solution also coincides with the solution which has been obtained for a bimatrix game using the Nash equilibrium technique.

Thus, we can simplify the procedure for finding a Nash equilibrium in mixed strategies by reducing the solution of a bimatrix game to solving two zero-sum games with payoff matrices (20).

Conclusion

The research we conducted made it possible to obtain a set of decision support models for the seller in the formation of a product offer for a homogeneous product on the marketplace in a competitive environment. The formation of the proposal is carried out in two stages. First, the seller, receiving information about the generalized demand, and knowing his functional and cost limitations, using the proposed models, can determine the permissible ranges of values for the characteristics of a homogeneous product which provide non-zero liquidity. Based on them, he can form alternative versions of his product offerings (product strategies). The choice of a product strategy in a competitive environment is carried out within the game-theoretic duopoly model using the Nash criterion. The example shows that it is possible to simplify the procedure for finding a Nash equilibrium in mixed strategies by reducing the solution of a bimatrix game to the solution of two zero-sum games. ■

References

1. *Intelligent marketplace assistants: how to make decisions safely in an era of change.* Available at: https://www.tadviser.ru/a/683081 (accessed 14 September 2022).

2. *Conducting parametric positional procurement (b2b-center.ru)* Available at: https://www.b2b-center.ru/help/ (accessed 15 October 2022).

3. *Automation of the procurement process at the enterprise.* Available at: https://www.agora.ru/avtomatizaciya-zakupok/?ysclid=l3uqm2ggxd (accessed 22 September 2022).

4. *The Marketplace Glossary (a16z.com)*. Available at: https://a16z.com/2020/02/18/marketplace-glossary/ (accessed 16 October 2022).

5. *Analytics services for sellers on the Marketplace*. Available at: https://vc.ru/services/106305-servisy-analitiki-dlya-prodavcov-na-marketpleysah (accessed 10 October 2022).

6. Matveev M., Podvalny S., Yadgarova Y. (2020) Automated service for product offer creation on the e-trading platform with marketplace technology. Proceedings of the *2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency, Summa 2020 Virtual, Lipetsk, 10–13 November 2020*, pp. 672–676.

7. Matveev M., Shmelev M., Aleynikova N. (2021) Information technologies for formation of services on e-traiding platform. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, no. 1, pp. 63–73 (in Russian).

8. Matveev M. (2021) Information technologies for supply creation on e-trading platform with marketplace technology. *Economics and Mathematical Methods*, vol. 57, no. 1, pp. 105–112 (in Russian).

9. Aleynikova N., Matveev M. (2022) Digital technology of centralized procurement organization. *Economics and Mathematical Methods*, vol. 58, no. 1, pp. 70–79 (in Russian).

10. *Matching of goods (matching)*. Available at: https://marketparser.ru/sopostavlenie-tovarov-(matching).html (accessed 17 October 2022).

11. Tirole J. (1988) *The theory of industrial economics*. Mass.: MIT Press.

12. Choi J.C., Shin H.S. (1992) A comment on a model of vertical product differentiation. *The Journal of Industrial Economics*, vol. 60, pp. 229–231.

13. Wauthy X. (1996) Quality choice in models of vertical differentiation. *The Journal of Industrial Economics*, vol. 3, pp. 345–355.

14. Dmitrienko K.Yu. (2009) Modeling of optimal firm's behavior on oligopolistic market under non-price product differentiation. *World of Economics and Management*, vol. 9, pp. 42–53 (in Russian).

15. Zenkevich N.A., Gladkova M.A. (2007) Game-theoretical model «Quality – Price» under competition on the industry market. *Vestnik of Saint Petersburg University*, vol. 8, pp. 3–31 (in Russian).

16. Gladkova M.A., Zenkevich N.A., Berezinets I.V. (2010) Model of product quality choice and internet-trading case study. *Vestnik of Saint Petersburg University*, vol. 8, pp. 55–77 (in Russian).

17. Malyutina T.D. (2013) Equilibrium investment strategies of firms in the vertically differentiated duopoly of Stackelberg. *Management of economic systems: electronic scientific journal*, vol. 12(60), pp. 1–21 (in Russian).

18. Aganin Yu. (2017) Optimal control of investments in a dynamic models of duopoly. *Vestnik Universiteta (Gosudarstvennyi universitet upravleniya)*, nos. 7–8, pp. 146–152 (in Russian).

19. Zadeh L. (1976) *The concept of a linguistic variable and its application to making approximate decisions.* Moscow: Mir (in Russian).

20. Sakulin S. (2007) Visualization of the aggregation operator based on the Shoquet integral by a fuzzy measure of the 2nd order. *Bulletin of Irkutsk State Technical University*, no. 2, pp. 45–51 (in Russian).

21. Sakulin S., Alfimcev A. (2012) On the practical application of fuzzy measures and the Shoke integral. *Bulletin of the Bauman Moscow State Technical University*, vol. 4: Computer systems and technologies, pp. 55–63 (in Russian).

22. Grabisch M., Marichal JL., Mesiar R., Pap E. (2009) *Aggregation functions*. Cambridge University Press.

23. Ledeneva T., Podvalniy S. (2016) The aggregation of information in evaluation systems. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, no. 4, pp. 155–164 (in Russian).

24. Detyniecki M. (2000) *Mathematical aggregation operators and their application to video querying.* Thesis for the degree Docteur de l'Universite Paris VI. Pierre and Marie Curie University.

25. Yager R. (1986) *Fuzzy sets and possibility theory. Resent developments.* Moscow: Radio and Communications (in Russian).

26. Sicilia M., Garsia E., Calvo T. (2003) An inquiry-based method for choquet integral-based aggregation of interface usability parameters. *República Checa Kybernetica*, vol. 39(5), pp. 601–614.

27. Grabisch M., Orlovski S., Yager R. (1998) Fuzzy aggregation of numerical preferences. H*andbook of Fuzzy Sets Series (ed. R. Slowinski), Dordrecht: Kluwer Academic*. Vol. 4: Fuzzy Sets in Decision Analysis, Operations Research and Statistics, pp. 31–68.

28. Grabisch M., Nguyen H. T., Walker E. A. (1995) *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Dordrecht: Kluwer Academic Publishers.

29. Grabisch M. (1996) The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, vol. 89, no. 3, pp. 445–456.

30. Pindyck R., Daniel L. (2013) *Microeconomics*. Pearson Education.

31. Shammazova A.M., Frolova Yu.A. (2000) *Oil space and market: terms and definitions. Problem-oriented terminology dictionary*. Ufa: USNTU (in Russian).

32. Petrosyan L., Zenkevich N., Semina E. (1998) *Game theory*. Moscow: Higher school (in Russian).

33. Goubko M., Novilov D. (2002) *Game theory for control mechanisms in organizations*. Moscow: Institute of Control Sciences of the Russian Academy of Sciences (in Russian).

34. Shikin E., Chkhartishvili A (2002) *Mathematical methods and models in management*. Moscow: Delo (in Russian).

About the authors

Mikhail G. Matveev

Dr. Sci. (Tech.), Professor;

Head of Department of Information Technologies in Management, Voronezh State University, 1, Universitetskaya pl., Voronezh 394018, Russia;

E-mail: mgmatveev@yandex.ru

ORCID: 0000-0002-6528-6420

Natalya A. Aleynikova

Cand. Sci. (Physics and Mathematics);

Associate Professor, Information Technologies in Management, Voronezh State University, 1, Universitetskaya pl., Voronezh 394018, Russia;

E-mail: balbashovan@mail.ru

ORCID: 0000-0001-5967-8260

Maria D. Titova

Master student, Information Technologies in Management, Voronezh State University, 1, Universitetskaya pl., Voronezh 394018, Russia;

E-mail: 29_06_titova@mail.ru

ORCID: 0000-0002-4647-7803

DOI: 10.17323/2587-814X.2023.2.55.70

The problem of interpretation, differentiation and classification of digital products

Ansel I. Shaidullin
E-mail: aishajdullin@hse.ru

HSE University
Address: 20, Myasnitskaya Street, Moscow 101000, Russia

Abstract

Digital innovative products often become a significant factor in the revision of companies' business strategies and influence consumer preferences. A key component in the process of formulating such strategies is understanding the implications underlying the attributes of digital products. This requires a good understanding of their nature and characteristics. To date, there is no solid basis for classifying various digital products according to their inherent characteristics. This paper presents a new interpretation of "digital products" based on the analysis of 2954 scientific articles from the Scopus database. It discusses the problems of differentiation of digital products from other types of products (such as "cyber-physical products," "digitized products," "smart products," etc.). We also developed a new classification of digital products by the method of highlighting their key attributes. The purpose of the study is to develop an advanced classification of digital products based on their differentiation from other types of products. The classification we constructed based on the principles of differentiation will allow innovators and businessmen to create more profound and more advanced business models.

Keywords: digital products, digitalization, physical product, classification, cyber-physical products, bibliometric analysis

Citation: Shaidullin A.I. (2023) The problem of interpretation, differentiation and classification of digital products. *Business Informatics*, vol. 17, no. 2, pp. 55–70. DOI: 10.17323/2587-814X.2023.2.55.70

Introduction

The popularity of the Internet has given companies around the world many tasks related to the promotion of their products via e-commerce. In particular, an increasing number of companies, including publishers, banks, news and insurance agents, are revising the concepts of their products in order to create and sell digital versions of traditional goods and services [1]. The growing popularity of selling digital products as the main way to make a profit has prompted business leaders and research scientists to study optimal competitive strategies associated with the sale of these products [2]. Interest in digital products is also noted in the number of published scientific articles on this topic. *Figure 1* shows the trend in the number of publications over the past 10 years for the keyword "digital product" (materials from the Scopus article database were used). It can be remarked that the greatest "surge of interest" occurred in the period 2019–2020. This can be attributed to the COVID-19 pandemic, when the demand for digital products increased significantly [3–6].

Different digital products demonstrate different growth rates [7], which largely depend on the main characteristics of the product [8–10] and the market environment [11–14]. Often, even minor changes in the structure of a digital product can seriously affect demand and change the existing market [15]. According to Christensen, innovations that significantly affect the market and break technological cycles are called "disruptive" [16].

Thus, different types of digital products require different approaches in modeling and in ways of implementation in the business process. Despite this, there is no solid basis for classifying various digital

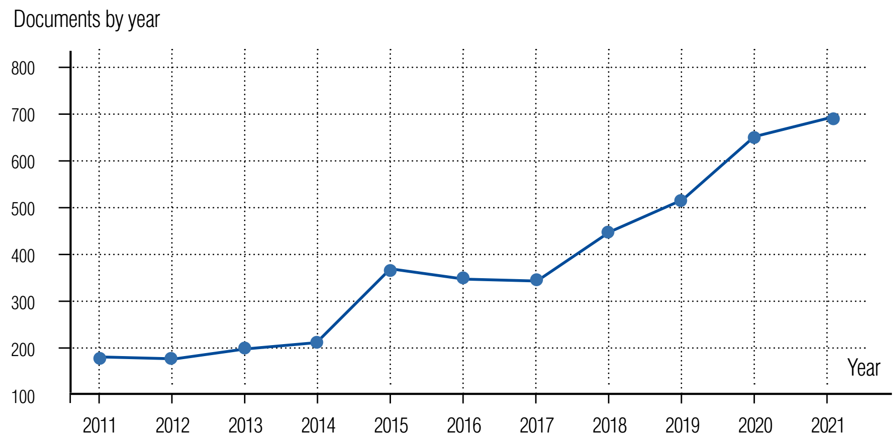

Fig. 1. The number of published articles on the topic "digital products" (based on analysis of articles from the Scopus database).

products according to their inherent characteristics [17]. The impact of digitalization on business and technology has several aspects that directly affect the digital architectures of products and services. Unfortunately, the current modeling approach for developing proper models of digital services and products suffers from the presence of many uncontrolled diverse approaches and modeling structures. High-quality digital models should follow a clear concept of value and service. Next, an attempt will be made to systematize and classify digital products based on their main types and characteristics. The purpose of the study is to develop an advanced classification of digital products based on their differentiation from other types of products.

There is also currently no reliable relationship between digital strategies and business modeling. Value is usually associated with utility and combines such categories as importance and desirability [18]. The concept of value is important in the development of appropriate digital services and related digital products.

1. Interpretation of digital products

What is a digital product? From the point of view of economic theory, most digital products are public goods delivered privately with all the consequences that follow from this: "the problem of the stowaway" and "tragedy of the commons" [19]. With the transition to the digital format, these problems are only getting worse, and the problem of combating media piracy on the Internet is becoming more complex than in the former analog world. Another definition is given in study [20], where the representation of a digital product as visual and verbal elements from the point of view of mental images was additionally studied. In article [21], the concept of a digital product is a complex scientific category that is subject to change.

Initially, when digital technologies were developed, they themselves were digital products [22]. This logic implies that digital products include digital devices (for example, mobile devices) and related (complementary) goods and services (for example, software). In the course of the spread of digital technologies, the typologization of digital products has also become more complicated. At the moment, these include not only digital devices, but also digital services, as well as manufactured and sold goods. However, such a classification of digital products closely intersects with the definitions of "intelligent products" and "cyber-physical products," which does not allow us to fully disclose the meaning of "digital products".

The interpretation of digital products also depends on who is the beneficiary of the introduction of a digital product to the market. The attitude of stakeholders to the digital product is contradictory [23]. For the state, this product is a means of developing the digital economy, stimulating an increase in the global competitiveness of the economic system and accelerating its economic growth. An example of the macroeconomic advantages obtained by replacing traditional (pre-digital) products with digital ones is to increase the transparency of economic activity and prevent tax evasion [24]. Another example is the reduction of government spending on the money supply during the transition to electronic money [25].

In turn, it is also beneficial for entrepreneurs to support the popularization of digital products, since they create business benefits. One of these advantages is the reduction of business risks and costs in the long term [26]. For example, online trading helps us to minimize reserves (logistics optimization) and more accurately predict demand (marketing optimization). Another advantage is associated with the expansion of activities: diversification of sales markets and obtaining "economies of scale." For example, e-commerce companies can conduct business cooperation and sell their products in remote markets, which is very difficult in the case of conventional retail. As a result, the importance of network effects is growing.

Modern consumers are showing increased interest in a digital product due to its greater availability and lower price compared to a pre-digital product. Thus, the popularity of online commerce, online finance and online public services is growing. However, consumers prefer a digital product only if it is of high quality [27]. Although consumers do not always take advantage of the lower price of a digital product, in most cases they face such disadvantages of a digital product as a high risk of its purchase and use (due to novelty, ambiguity of the legal field and other reasons).

This contradiction – high demand with high uncertainty indicators – constrains the production and sale of digital products and slows down the development of the digital economy. Attempts to overcome this by improving the quality of a digital product in the conditions of the modern digital economy are ineffective due to the weak development and underdevelopment of the scientific vision of the quality of a digital product as an economic category [28]. Therefore, an important scientific and practical task is to overcome the existing contradiction with the most complete, accurate and correct definition of the quality of a digital product as an economic category. To do this, it is important to identify the factors that can be used to distinguish between "digital" and "physical" products.

Digital products can be distributed without loss in a purely digital form (for example, using computer networks.). A digital product serves a specific purpose, is intended for sale or exchange, and can satisfy the user's desire or need. Other criteria that help distinguish digital products from physical ones can be found in *Table 1.*

Industrial standard items are static. Only a small amount of alteration is possible with them. Digital products, on the other hand, are dynamic. They include both cloud services and software. Through network connections, they can be updated. As a result, the functionality of the products can be modified to meet the evolving demands and wants of clients. Digital goods and services might be produced gradually or offered momentarily. Digital products can be copied almost free of charge and are subject to non-commercial copying by end consumers. Since the quality of the copy usually does not deteriorate, copies can become available on a large scale. At the same time, the problem of online piracy is becoming more acute. Article [34] analyzes the basic models of piracy, models with indirect assignment, models with network effects and models with asymmetric information.

Digital products are able to capture their own state and present this information in related contexts [35]. The so-called "servitization of products" is based on this. The buyer is not being sold a physical product, but a service. The supplier can remotely determine if the product is working and initiate maintenance and repair if necessary. Evaluation of the status information and analysis of the product usage history allow you to predict when a malfunction is likely. Maintenance or replacement of the product is performed before the predicted failure. The collected data also provides information for on-site repairs, so that a high speed of problem solving can be achieved the first time. Thus, it is possible to significantly reduce unplanned shutdowns of products.

Digital products also allow network effects [36], which grow exponentially with the number of participating devices [37, 38]. Increasing the number of digitized products increases incentives for additional service providers. At the same time, it makes further product digitization more appealing. Network effects arise not only to enhance functionality, but also for the analytical use of data collected by digitized products (network intelligence). It is feasible to spot patterns considerably earlier and more accurately by merging data from numerous devices.

Digital products and services become part of an information system that accelerates learning and cognition processes in all products [39]. In parallel, a number of other useful effects can be achieved, such as network optimization, maintenance optimization and improved recovery capabilities when considering individual systems [40]. The consumer turns into a "co-producer" [41]. Platforms complement products that interact through standardized interfaces.

Table 1.

The difference between a "physical" and a "digital" product

Criteria	Digital product	Physical product
Product properties		
Value after use	After the first use, they are identical to new ones, and in some cases even better (for example, for digital games, the achieved levels add value). Only "moral" wear and tear is relevant (for example, obsolescence, going out of fashion, etc.) [29].	Usually depreciated after purchase and use ("used product"). For these products, the concepts of "depreciation" and "physical and moral" wear are relevant.
Product flexibility and service delivery speed	Flexible products. Changes can be easily and quickly implemented in the product. However, this may cause certain difficulties in the context of intellectual copyrights. The possibility of instant "delivery" of the order (or access).	Static products: the composition, idea, appearance, design of the product are usually clearly defined, and the introduction of any changes is accompanied by a change in the product itself. There are delays in the delivery of products: additional difficulties are created in logistics issues.
Costs		
Fixed and variable product costs	High fixed costs for R&D. A small or practically zero cost of delivery per unit of product. Low overhead.	There are certain fixed costs. Non-trivial unit delivery cost.
The costs of "audience building," the problem of network effects	Audience growth depends on the influence of the network effect and "accumulates" faster than for a physical product. This reduces the cost of attracting an additional audience.	High costs. The effect of the "network effect" depends on the type of product.
Transaction costs	Low, completion of purchase and sale agreements "in a few clicks".	High.
The costs of product search, "menu", switching and copying	Low. Piracy and copyright issues arise when copying.	High. Copying requires copying directly the physical object itself.
Risks		
Risk for the developer	The risk can be high for products such as digital games, because market demand and reaction to it are very volatile. To mitigate risks, development managers usually use non-cascading business process methodologies: Agile [30] or Scrum [31] project management methodologies.	Depends on the nature of the product. For seasonal products, market demand is very unstable, and the risk is high. Cascading project management methodologies are mainly used for product output, plus there is a need to create "roadmaps" for product development.
Risk to consumers	It may be high, since consumers may have to learn how to use the product, and they may not know about it long before buying.	May be available for touch and detailed visual examination before purchase.
Information asymmetry	Low information asymmetry. The occurrence of the principal-agent problem is less likely [32].	High information asymmetry. The high significance of the principal-agent problem.

Criteria	Digital product	Physical product
Market factors		
Price discrimination and market segmentation	Price discrimination is possible, but unlikely due to the lack of information asymmetry. It is preferable to use Big Data analysis for audience and market segmentation. Moderate accuracy, significant role of quantitative marketing research [33].	Price discrimination of all three types is likely. Audience analysis is carried out mainly with the help of various surveys, focus groups and other methods of marketing analysis. Low accuracy, high error, high influence of subjective factors (for example, cognitive distortion such as "observer error/bias").
Profitability	Higher profitability compared to "physical products": there are no recurring costs for goods, hence saving most of the profits.	Profitability is usually lower than that of "digital products": usually due to high fixed production costs.
Disintermediation	Intermediaries are often excluded from the service provision process.	Often, the active participation of 1–2 intermediaries is necessary.

Producers will not simply rely on supply and demand according to marginal revenue and marginal cost pricing. Based on the characteristics of the digital products themselves, the cost price, the network market environment, the characteristics of consumer behavior and network expansion, the theory of group pricing follows [42], on the basis of which a business strategy is put forward and a business model is built.

2. Differentiation of digital products from other types of products

Based on the identification of the definition of digital products and their differentiation from physical products (an attempt to solve the "interpretation problem"), it is possible to build models of digital business strategies. However, there are currently no articles in the scientific community that would clearly distinguish between such concepts as intelligent products, digitalized (digitized) products, cyber-physical products, digital products, etc. *Table 2* presents definitions of these concepts, and *Fig. 2* shows a comparison of terms in Euler circles (an attempt to solve the "differentiation problem").

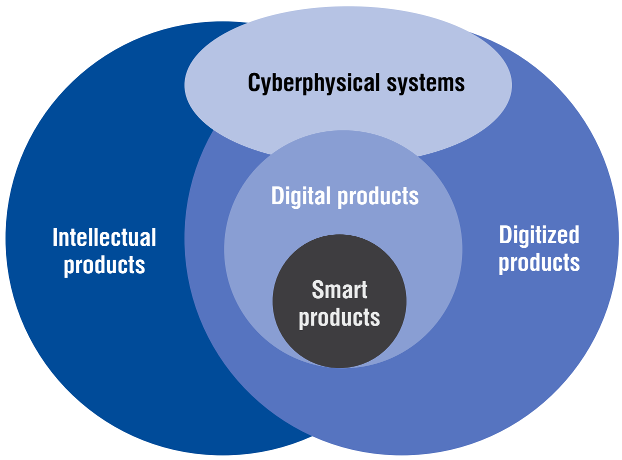

Fig. 2. Differentiation of terms related to different types of digitalized products.

Table 2.

Separate concepts and concepts related to “digitized products”

Concept	Description
Digitized products	“(...) digitization makes physical products programmable, addressable, intelligent, sociable, memorable, traceable and associated (...)” [43]. Such products combine physical and digital attributes. The inclusion of the physical shell in the definition is an important factor. When building various models, it is necessary to take this into account.
Cyberphysical systems	“(...) represent the integration of computing with physical processes. Embedded computers and networks monitor and control physical processes, usually with feedback loops when physical processes affect calculations and vice versa (...)” [44]. Cyber-physical products, in addition to the physical shell, take into account the internal “physical processes” of the product.
Intelligent products	“(...) contain the possibilities of perception, memory, data processing, reasoning and communication (...)” [45]. Intelligent products are separated from being classified as physical matters; here the “content” of the product comes to the fore, namely the ability of the product to store, process and transmit information.
Smart objects	“(...) have a unique identity, are able to communicate effectively with the environment, can store data about themselves, use language and are able to make decisions (...)” [46]. The definition is very similar in meaning to the definition of “intelligent products.” Smart objects are part of the “smart products” system. The key point here is the ability to make decisions and communicate to the external environment [47]. They know not only about the steps of the process that have already been completed, but are also able to determine future steps [48]. Sensors allow you to record physical measurements, cameras – to receive visual information about the product and its surroundings in real time.
Smart, connected products	“(...) consist of physical components, intelligent components (sensors, microprocessors, data storage, controls, software, operating system) and connection components (ports, antenna, protocols) (...)” [49]. The definition is close in meaning to the definition of “digitalized products.” However, the definition is narrower: these products refer specifically to “smart objects.”
Internet of things	“(...) everyday objects can be equipped with identification, recognition, networking and processing capabilities that will allow them to communicate with each other and with other devices and services via the Internet (...)” [50]. The definition emphasizes the systemic nature of such products. Objects can make decisions and interact with both humans and other robotic objects.

3. Classification of digital products

The problem of “digital products” is being dealt with by scientists from different fields of life. One of the ways to classify objects is to classify them by scope of application. Data from the Scopus website was used to construct the following tables and figures. From *Fig. 3* and *Table 3* it can be noted that the most popular areas where the theoretical foundations and practical methods of using digital products are studied are computer science, engineering, social sciences, management and business, mathematics, etc.

The complexity of identifying clusters for classification is observed when clustering terms based on the

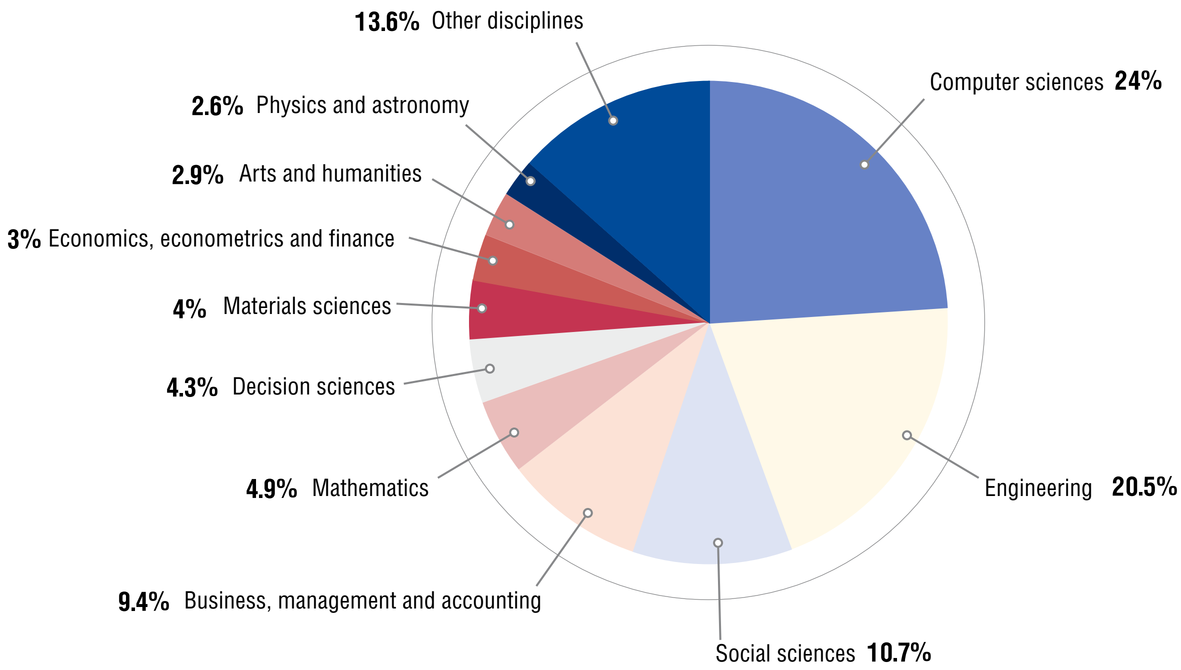

Fig. 3. Distribution of publications on digital products by research categories based on bibliometric analysis of the Scopus database.

processing of 2954 articles from the Scopus database (*Fig. 4*). The methodology proposed in the article [51] was used to build a term map. To cluster terms, the VOSviewer program was used, which identified five large clusters.

The first cluster includes terms from the field of digital technologies. Digital technologies, such as additive manufacturing, artificial intelligence, cloud computing, data analysis, social networks and wireless sensor networks [52, 53], open up unprecedented opportunities for the development and release of new products [54]. Rather, this cluster reflects the applied nature of the use of digital products in the context of digital production. Digital production is a digital representation of the entire production process. It includes three main components: a digital factory, a virtual factory and the corresponding data management. The second cluster includes areas of application of digital products (for example, in the field of sales). The third cluster highlights the spheres of interaction between a machine and a person. The fourth cluster reflects measures to protect digital products. The fifth cluster emphasizes the importance of digital innovation.

Figure 4 also shows a heat map of keywords by year. Such a map allows you to highlight the basic (fundamental) concepts within the digitalization process, as well as new elements that relate to the topic under study. New directions in this area are digital twins, digital transformation, added reality, digital innovations within the concept of "Industry 4.0".

Some authors, among them [55–57], distinguish a separate niche in the classification of digital products in the form of "digital data". In 2018, a new measure emerged based on the foundation of data citation: data

Table 3.

Number of articles on the term "digital products" by research disciplines

Sphere	Number of articles	Sphere	Number of articles
Computer science (Informatics)	3812	Energy and energy systems sciences	251
Engineering	3259	Chemical engineering	179
Social sciencies	1706	Psychology	159
Business, management and accounting	1492	Agricultural and biological sciences	155
Mathematics	786	Chemistry	119
Decision sciences	679	Biochemistry, genetics and molecular biology	110
Materials science	636	Multidisciplinary directions	70
Economics, econometrics and finance	483	Health sciences	68
Arts and humanities	464	Neurology	40
Physics and astronomy	418	Dentistry	21
Environmental science	373	Nursing	19
Earth and planetary sciences	316	Pharmacology, toxicology and pharmaceuticals	16
Medicine	253	Immunology and microbiology	12

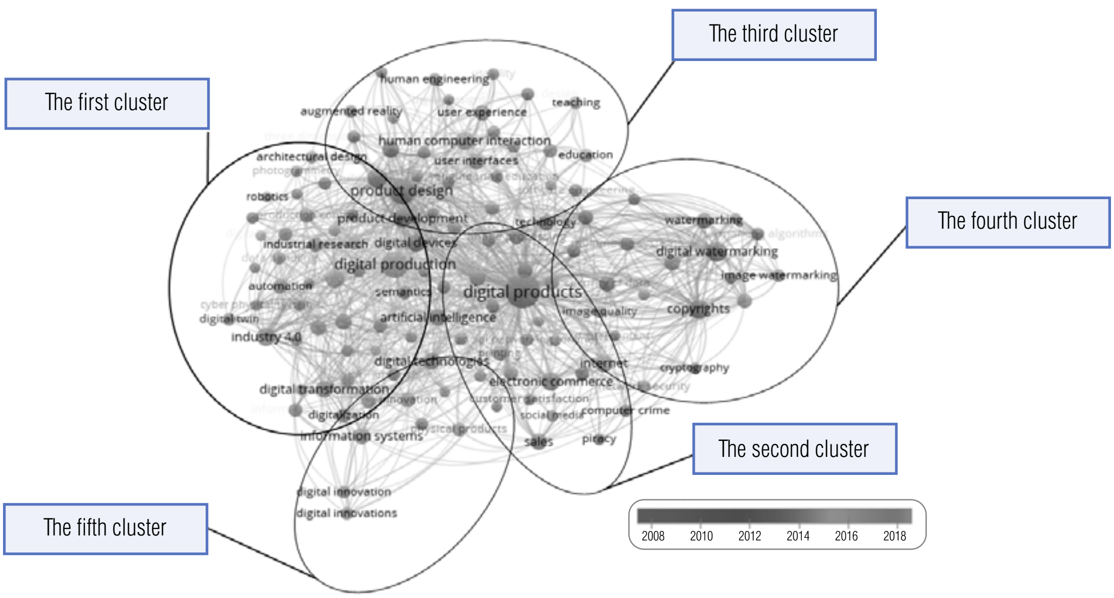

Fig. 4. Cluster map and keyword heat map.

reliability is a real value reflecting the importance of data cited by a research organization [58].

There is such a phenomenon as digital information products (DIP), which are a subset of digital products. DIP is a special type of digital product, the main advantage of which is the provision of information [59]. DIPs often consist of a mixture of information and software. The difference between DIP and pure software is that DIP is focused on delivering information. In this respect, only a limited set of software systems can qualify as DIP [60]. DIP is widely distributed, for example, electronic magazines, films, electronic weather reports, digitized educational programs, textbooks and lectures.

The main limitation of all existing classifications is a vague idea of the object under study: there is no clear opinion on how digital products differ from other types of products. In this article, about 2 954 articles were studied to highlight this problem. Thanks to the differentiation of products, it is possible to build a better classification. *Figure 5* shows an approximate division of digital products into categories. The constructed classification is based on "differentiated criteria": only those types of digital products that differ markedly from other categories of digitalized products are included in the classification. In future works, it is planned to expand the existing classifications. To enhance the depth of the construction of classifiers of digital products, it is necessary to identify additional criteria that determine the differentiation of one category by a product from others. For example, article [62] suggests several classifications of digital products based on the allocation of various criteria: 1) digital products based on content; utilities and tools; online services; 2) categories based on the concepts of 4P, 4S and 4S; 3) based on the possibility of litigation and the degree of detail. Despite the fact that the authors create a systematic view of the problem and strive to

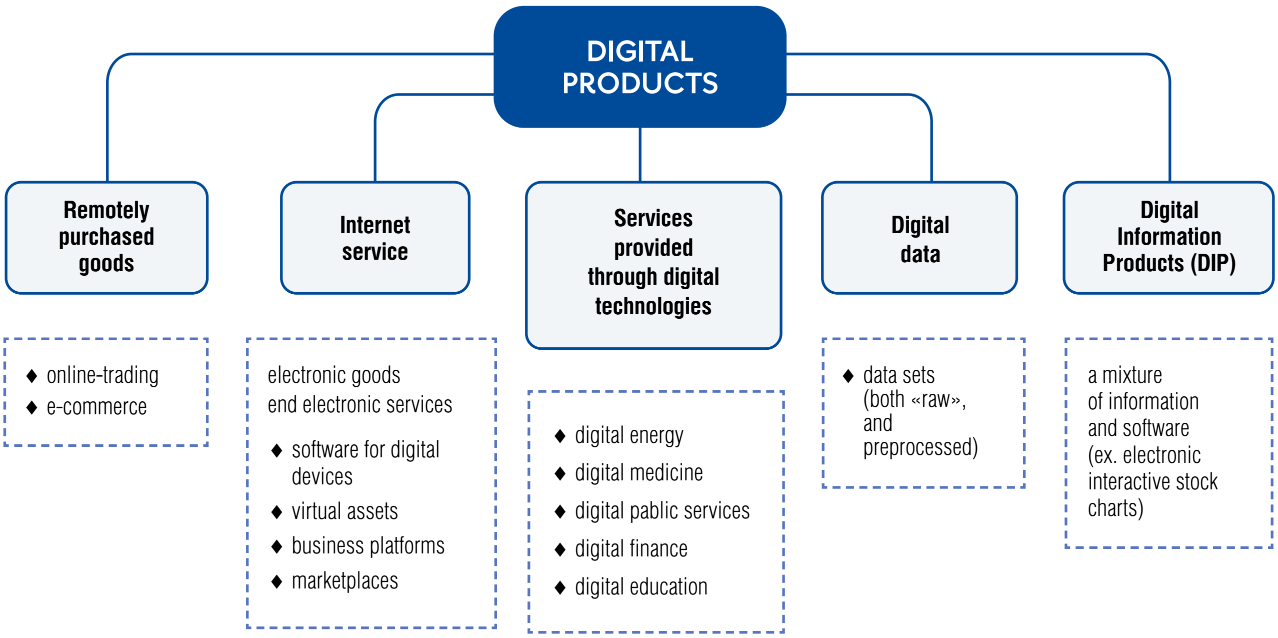

Fig. 5. Classification of digital products based on their differentiation from other types of digitalized products.

combine all the criteria into a single structure, the proposed classification does not take into account the nature of digital products, as well as differences between digitalized products. To further develop the concept of "digital products," it is necessary to study the methods of modeling classifiers.

Among the methods of classifier modeling, we can distinguish: sentimental methods, the Rocchio method, the probabilistic classification method (Bayes method), clustering methods, etc. Future articles may also be devoted to the issues of comparing the effectiveness of using the above modeling methods.

The next stage in the development of this topic is the construction of various business models where digital products are used. Understanding the specifics of digital products for modeling is extremely important, since business models often include a description of product characteristics. Digital products do not have a physical form as such. As we can see, there are certain factors that can significantly affect the quality of models. Those rules that are optimal, for example, for physical products, may not be relevant for digital products.

Article [61] uses a hybrid system based on fuzzy modeling to identify dependencies between user characteristics and the evaluation of digital products in order to develop a dynamic pricing system. Currently, industrial companies are gradually moving from a product-oriented business model to a service-dominant logic. Such logic offers personalized products and services in the form of a set of solutions to meet individual customer needs.

Conclusion

Digital product development has been booming in recent years due to the maturity of the entire environment. However, most e-commerce research still focuses on physical products and misses the value of the digital wave. In this article, criteria were proposed by which it is possible to distinguish between physical and digital products. To further build a product development strategy, it is critically important to understand the main characteristics by which one type of product differs from another. Among the criteria that make it possible to distinguish between physical and digital products, the following can be distinguished: the properties of the product itself, the costs of production, distribution, support, etc. of products, risks and market factors. Understanding the structure and properties of the product, as well as key attributes, will make it possible to commercialize them more efficiently and fit them more harmoniously into the country's economic system.

The development of digital product platforms is a prevailing trend in many industries. As firms introduce digital technologies into established product categories, they need to cope with tensions at several organizational levels, including strategy, technology and structure. A new fundamental paradigm shift in industrial production is brought about by the integration of Internet technologies and cutting-edge technology in the area of "smart" items, which is based on the digitization of factories. The future of production envisions modular and effective production systems as well as scenarios where goods manage their own manufacturing processes.

There is an evolution of Internet systems combining features of both technical and economic aspects. In this regard, there is a problem with solutions related to modeling and managing various aspects of the organization of the system. This article presents options for interpreting digital products, as well as their differentiation and classification. The differentiation of digital products from other types of digitalized products allows you to differentiate the areas of research, and also helps to investigate individual categories of certain forms of products based on their differentiation. The implication is that understanding these differences can create a clearer picture of the perception of a complex technological world.

Innovations in the digital world are increasingly being developed in the field of open platforms consisting of basic technology and a large number of additional products developed by an ecosystem of independent complementary companies. The literature on the platform ecosystem mainly focuses on indirect network effects arising from the number of add-ons, with little attention to the quality of add-ons. Joint actions of platform owners and users are needed to respond to opportunities, failures and obsolescence. ■

References

1. Hui K.L., Chau P.Y.K. (2002) Classifying digital products. *Communications of the ACM*, vol. 45, no. 6, pp. 73–79. https://doi.org/10.1145/508448.508451

2. Adisorn T., Tholen L., Götz T. (2021) Towards a digital product passport fit for contributing to a circular economy. *Energies,* vol. 14, no. 8. https://doi.org/10.3390/en14082289

3. Almeida F., Duarte S.J., Monteiro A.J. (2020) The challenges and opportunities in the digitalization of companies in a post-COVID-19 world. *IEEE Engineering Management Review*, vol. 48, no. 3, pp. 97–103. https://doi.org/10.1109/EMR.2020.3013206

4. Corbet S., Hou Y.G., Hu Y., Larkin C., Lucey B., Oxley L. (2022) Cryptocurrency liquidity and volatility interrelationships during the COVID-19 pandemic. *Finance Research Letters*, vol. 45, article 102137. https://doi.org/10.1016/j.frl.2021.102137

5. Fairgrieve D., Feldschreiber P., Howells G., Pilgerstorfer M.Q.C. (2020) Products in a pandemic: Liability for medical products and the fight against COVID-19. *European Journal of Risk Regulation*, vol. 11, no. 3, pp. 565–603. https://doi.org/10.1017/err.2020.54

6. Jin L., Hao Z., Huang J., Akram H.R., Saeed M.F., Ma H. (2021) Depression and anxiety symptoms are associated with problematic smartphone use under the COVID-19 epidemic: The mediation models. *Children and Youth Services Review*, vol. 121, article 105875. https://doi.org/10.1016/j.childyouth.2020.105875

7. Hu Y., Li W. (2011) Document sentiment classification by exploring description model of topical terms. *Computer Speech and Language*, vol. 25, no. 2, pp. 386–403. https://doi.org/10.1016/j.csl.2010.07.004

8. De Sordi J.O., Nelson R.E., Meireles M., da Silveira M.A. (2016) Development of digital products and services: Proposal of a framework to analyze versioning actions. *European Management Journal*, vol. 34, no. 5, pp. 564–578. https://doi.org/10.1016/j.emj.2016.01.009

9. Kim M. (2019) Digital product presentation, information processing, need for cognition and behavioral intent in digital commerce. *Journal of Retailing and Consumer Services*, vol. 50, pp. 362–370. https://doi.org/10.1016/j.jretconser.2018.07.011

10. Kleinsmann M., Ten Bhömer M. (2020) The (new) roles of prototypes during the co-development of digital product service systems. *International Journal of Design*, vol. 14, no. 1, pp. 65–79.

11. U.S. Department of Commerce (1998) *The Emerging Digital Economy*. Available at: https://www.commerce.gov/sites/default/files/migrated/reports/emergingdig_0.pdf (accessed 17 April 2023).

12. Avinadav T., Chernonog T., Perlman Y. (2014) Analysis of protection and pricing strategies for digital products under uncertain demand. *International Journal of Production Economics*, vol. 158, pp. 54–64. https://doi.org/10.1016/j.ijpe.2014.07.021

13. Gustafsson E., Jonsson P., Holmström J. (2019) Digital product fitting in retail supply chains: Maturity levels and potential outcomes. *Supply Chain Management*, vol. 24, no. 5, pp. 574–589. https://doi.org/10.1108/SCM-07-2018-0247

14. Feng J., Yu K. (2020) Moore's law and price trends of digital products: The case of smartphones. *Economics of Innovation and New Technology*, vol. 29, no. 4, pp. 349–368. https://doi.org/10.1080/10438599.2019.1628509

15. Makkonen H., Komulainen H. (2018) Explicating the market dimension in the study of digital innovation: A management framework for digital innovation. *Technology Analysis and Strategic Management*, vol. 30, no. 9, pp. 1015–1028. https://doi.org/10.1080/09537325.2018.1433823

16. Christensen C.M. (1997) *The innovator's dilemma: when new technologies cause great firms to fail*. Boston: Harvard Business School Press.

17. Abrosimov Y., Mingaleev G., Snegurenko A. (2020) Organization of enterprise digital infrastructure. *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, Vladivostok, Russia, pp. 1–4. https://doi.org/10.1109/FarEastCon50210.2020.9271088

18. Kim C., Kim D.J. (2017) Uncovering the value stream of digital content business from users' viewpoint. *International Journal of Information Management*, vol. 37, no. 6, pp. 553–565. https://doi.org/10.1016/j.ijinfomgt.2017.05.004

19. Demsetz H. (1970) The private provision of public goods. *The Journal of Law and Economics*, vol. 13, no. 2, pp. 293–306. https://doi.org/10.1086/466695

20. Kim M., Lennon S. (2008) The effects of visual and verbal information on attitudes and purchase intentions in Internet shopping. *Psychology and Marketing*, vol. 25, no. 2, pp. 146–178.

21. Popkova E.G. (2020) Quality of digital product: Theory and practice. *International Journal for Quality Research*, vol. 14, no. 1, pp. 201–218. https://doi.org/10.24874/IJQR14.01-13

22. Martínez-Caro E., Cegarra-Navarro J.G., Alfonso-Ruiz F.J. (2020) Digital technologies and firm performance: The role of digital organisational culture. *Technological Forecasting and Social Change*, vol. 154, article 119962. https://doi.org/10.1016/j.techfore.2020.119962

23. Alawneh A., Al-Refai H., Batiha K. (2013) Measuring user satisfaction from e-government services: Lessons from Jordan. *Government Information Quarterly*, vol. 30, no. 3, pp. 277–288. https://doi.org/10.1016/j.giq.2013.03.001

24. Clark B.Y., Brudney J.L., Jang S. (2013) Coproduction of government services and the new information technology: Investigating the distributional biases. *Public Administration Review*, vol. 73, no. 5, pp. 687–701. https://doi.org/10.1111/puar.12092

25. Chen C., Lin Y., Chen W., Chao C., Pandia H. (2021) Role of government to enhance digital transformation in small service business. *Sustainability*, vol. 13, no. 3, article 1028. https://doi.org/10.3390/su13031028

26. Hienerth C., Lettl C., Keinz P. (2014) Synergies among producer firms, lead users, and user communities: The case of the LEGO producer-user ecosystem. *Journal of Product Innovation Management*, vol. 31, no. 4, pp. 848–866. https://doi.org/10.1111/jpim.12127

27. Howells G. (2020) Protecting consumer protection values in the fourth industrial revolution. *Journal of Consumer Policy*, vol. 43, no. 1, pp. 145–175. https://doi.org/10.1007/s10603-019-09430-3

28. Mohammad A.A.S. (2012) The effect of brand trust and perceived value in building brand loyalty. *International Research Journal of Finance and Economics*, vol. 85, pp. 111–126.

29. Moore-Russo D., Grantham K., Lewis K., Bateman S.M. (2010) Comparing physical and cyber-enhanced product dissection: Analysis from multiple perspectives. *International Journal of Engineering Education*, vol. 26, no. 6, pp. 1378–1390.

30. Martin R.S., Newkirk J.W., Koss R.S. (2003) *Agile software development: Principles, patterns, and practices*. Upper Saddle River, NJ: Prentice Hall.

31. Sutherland J. (2014) *Scrum: The art of doing twice the work in half the time*. Sydney: Currency.

32. Chernonog T., Avinadav T. (2019) Pricing and advertising in a supply chain of perishable products under asymmetric information. *International Journal of Production Economics*, vol. 209, pp. 249–264. https://doi.org/10.1016/j.ijpe.2017.10.002

33. Gong J., Smith M.D., Telang R. (2015) Substitution or promotion? The impact of price discounts on cross-channel sales of digital movies. *Journal of Retailing*, vol. 91, no. 2, pp. 343–357. https://doi.org/10.1016/j.jretai.2015.02.002

34. Peitz M., Waelbroeck P. (2006) Piracy of digital products: A critical review of the theoretical literature. *Information Economics and Policy*, vol. 18, no. 4, pp. 449–476. https://doi.org/10.1016/j.infoecopol.2006.06.005

35. McAfee A., Brynjolfsson E. (2017) *Machine, platform, crowd. Harnessing our digital future*. W.W. Norton & Company.

36. Möhring M., Keller B., Schmidt R., Pietzsch L., Karich L., Berhalter C. (2018) Using smart edge devices to integrate consumers into digitized processes: The case of amazon dash-button. *BPM, Workshops, LNBIP*, pp. 374–383.

37. Metcalfe B. (1996) There oughta be a law. *The New York Times*, July 15, 1996, Section D, p. 7. Available at: https://www.nytimes.com/1996/07/15/business/there-oughta-be-a-law.html (accessed 17 April 2023).

38. Metcalfe B. (2013) Metcalfe's law after 40 years of Ethernet. *IEEE Computer*, vol. 46, no. 12, pp. 26–31. https://doi.org/10.1109/MC.2013.374

39. Jugel D., Schweda C.M., Zimmermann A. (2015) Modeling decisions for collaborative enterprise architecture engineering. *10th Workshop Trends in Enterprise Architecture Research (TEAR)*, CAISE, Stockholm, Sweden, pp. 351–362.

40. Vargo S.L., Lusch R.F. (2008) Service-dominant logic: continuing the evolution. *Journal of the Academy of Marketing Science*, vol. 36, no. 1, pp. 1–10. https://doi.org/10.1007/s11747-007-0069-6

41. Vargo S.L., Lusch R.F. (2016) Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science,* vol. 44, no. 4, pp. 5–23. https://doi.org/10.1007/s11747-015-0456-3

42. Zhu C., Yao Z., Luan J., Zhao F. (2016) Network externality on retailer and supplier pricing strategies for competitive products. *The Pacific Asia Conference on Information Systems, PACIS 2016.*

43. Henfridsson O., Mathiassen L., Svahn F. (2014) Managing technological change in the digital age: The role of architectural frames. *Journal of Information Technology*, vol. 29, no. 1, pp. 27–43. https://doi.org/10.1057/jit.2013.30

44. Ahmed C.M., Zhou J. (2020) Challenges and opportunities in cyberphysical systems security: A physics-based perspective. *IEEE Security and Privacy*, vol. 18, no. 6, pp. 14–22. https://doi.org/10.1109/MSEC.2020.3002851

45. Agaram V. (2017) Knowledge system based design-for-reliability for developing connected intelligent products. *SAE Technical Papers*, SAE International. https://doi.org/10.4271/2017-01-0196

46. Bajic E., Cea A. (2005) Smart objects and services modeling in the supply chain. *IFAC Proceedings Volumes (IFAC – Papers Online)*, vol. 16, pp. 25–30. https://doi.org/10.3182/20050703-6-cz-1902.01488

47. Miche M., Schreiber D., Hartmann M. (2009) Core services for smart products. *3rd European Workshop on Smart Products*, pp. 1–4.

48. Cronin M.J. (2010) *Smart products, smarter services: Strategies for embedded control*. Cambridge University Press. https://doi.org/10.1017/CBO9780511761928

49. Porter M.E., Heppelmann J.E. (2015) *How smart, connected products are transforming companies*. Harvard Business Review.

50. Dumitrescu R. (2018) Utilizing opportunities for the industrial location. *The Internet of Things* (ed. U. Sendler). Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-54904-9_12

51. Anand A., Brix J. (2022) The learning organization and organizational learning in the public sector: A review and research agenda. *Learning Organization*, vol. 29, no. 2, pp. 129–156. https://doi.org /10.1108/tlo-05-2021-0061

52. Gianvito L., Pesce D., Tucci C.L. (2021) The digital transformation of search and recombination in the innovation function: Tensions and an integrative framework. *Journal of Product Innovation Management*, vol. 38, no. 1, pp. 90–113.

53. Vial G. (2019) Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, vol. 28, no. 2, pp. 118–144.

54. Verganti R, Vendraminelli L., Iansiti M. (2020) Innovation and design in the age of artificial intelligence. *Journal of Product Innovation Management*, vol. 37, no. 3, pp. 212–271.

55. Ahmad W., Neil D.T. (1994) An evaluation of landsat thematic mapper (TM) digital data for discriminating coral reef zonation: Heron reef (GBR). *International Journal of Remote Sensing*, vol. 15, no. 13, pp. 2583–2597. https://doi.org/10.1080/01431169408954268

56. Back M.D., Küfner A.C., Egloff B. (2011) Automatic or the people? *Psychological Science*, vol. 22, no. 6, pp. 837–838. https://doi.org/10.1177/0956797611409592

57. Bellanova R. (2017) Digital, politics, and algorithms: Governing digital data through the lens of data protection. *European Journal of Social Theory*, vol. 20, no. 3, pp. 329–347. https://doi.org/10.1177/1368431016679167

58. Hedberg T.D., Krima S., Camelio J.A. (2019) Method for enabling a root of trust in support of product data certification and traceability. *Journal of Computing and Information Science in Engineering*, vol. 19, no. 4. https://doi.org/10.1115/1.4042839

59. Oberweis A., Pankratius V., Stucky W. (2007) Product lines for digital information products. *Information Systems*, vol. 32, no. 6, pp. 909–939. https://doi.org/10.1016/j.is.2006.09.003

60. Mencarelli R., Rivière A., Lombart C. (2021) Do myriad e-channels always create value for customers? A dynamic analysis of the perceived value of a digital information product during the usage phase. *Journal of Retailing and Consumer Services*, vol. 63, article 102674. https://doi.org/10.1016/j.jretconser.2021.102674

61. Eckert T., Hüsig S. (2022) Innovation portfolio management: A systematic review and research agenda in regards to digital service innovations. *Management Review Quarterly*, vol. 72, no. 1, pp. 187–230. https://doi.org/10.1007/s11301-020-00208-3

62. Wang Y., Wang K.L., Yao J.T. (2009) Marketing mixes for digital products: A study of the marketspaces in China. *International Journal of Technology Marketing*, vol. 4, no. 1, pp. 15–42. https://doi.org/10.1504/IJTMKT.2009.023554

About the author

Ansel I. Shaidullin

PhD student, Department of Business Informatics, Higher School of Business, HSE University, 26, Shabolovka st., Moscow 119049, Russia;

E-mail: aishajdullin@hse.ru

ORCID: 0000-0002-2653-1745

DOI: 10.17323/2587-814X.2023.2.71.84

An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus*

Ivan E. Nikolaev
E-mail: ivan_nikolaev@csu.ru

Chelyabinsk State University
Address: 129, Kashirin Brothers Str., Chelyabinsk 454001, Russia

Abstract

Online recruitment systems have accumulated a huge amount of data on the real labor market in recent years. Of particular interest to the study are the data on the real requirements of the labor market contained in the texts of online vacancies, as well as the process of extracting and structuring them for further analysis and use. The stage of compiling an up-to-date list of requirements for a position profile in the recruitment process is very time-consuming and requires a large amount of effort from an HR specialist related to monitoring changes in entire industries and professions, as well as analyzing relevance of existing requirements on the market. In this article, the author proposes a conceptual model of a recommendation system that allows one to reduce the burden on an HR specialist at the stage of forming an up-to-date list of requirements for a position profile in the recruitment process. The model is based on a combination of the following components: a graph model of labor market requirements based on the ESCO taxonomy adapted for the Russian language; and an intelligent method of forming

* The article is published with the support of the HSE University Partnership Programme

recommendations for compiling an up-to-date list of requirements in the recruitment process based on neural network models of the language on the architecture of transformers, ESCO skills taxonomy and corpus online vacancies of the Russian labor market. The article also provides a conceptual algorithm for the work of the recommendation system and possible options for recommendations on updating the list of requirements of the position profile in the recruitment process based on an analysis of the needs of the real labor market.

Keywords: labor market analysis, labor market requirements, human resources, job profile, data mining, natural language processing, neural network language models

Citation: Nikolaev I.E. (2023) An intelligent method for generating a list of job profile requirements based on neural network language models using ESCO taxonomy and online job corpus. *Business Informatics*, vol. 17, no. 2, pp. 71–84. DOI: 10.17323/2587-814X.2023.2.71.84

Introduction

Currently, the vast majority of companies cover a significant part of their staffing needs by posting online job advertisements in online recruitment systems. In such systems, a huge amount of structured and semi-structured data about vacancies and applicants (resumes) is generated and accumulated daily. For example, there are two Russian-language online recruitment systems hh.ru and superjob: the first one has about 57 million resumes and more than 40 million vacancies for the period from 2010 to 2020; the second contains more than 12 million vacancies for the period from 2010 to 2020.

In this regard, the problem of processing and extracting information from online job data becomes particularly relevant, since its solution will allow modeling and understanding complex phenomena in the labor market (see, for example, [1–6]).

Although online recruitment systems have de facto become the main source for co-founders and recruitment managers, they still show shortcomings in search, relevance and accuracy, since job offers are presented in natural language and often in several syntactically and lexically different, but semantically similar forms. This leads to the fact that in the search process search queries are subject to natural language ambiguity and do not compare well with job descriptions in online vacancies. In particular, queries that are overly defined or inconsistent often do not return matches, while relevant job offers could still be found if the problem of consistency or specificity of search queries was solved. If there are not enough exact matches, it is often necessary to accept the worst alternatives or compromise with the initial requirements.

Another problem is the lack of tools that allow a person to use the data extracted from online vacancies in their professional activities. For example, when developing a position profile in the process of creating a vacancy for a position in a company, an HR specialist or the head of the corresponding department needs to spend significant efforts in order to select several dozen, or even hundreds, of similar vacancies, analyze them, extract from them a list of published requirements and responsibilities, conduct their analysis and ranking, and compare them with the job responsibilities of the position for which a new vacancy is being developed.

Modern recommendation systems in general, and in the field of the labor market in particular, largely depend on large amounts of manual data processing and expert knowledge, which makes them expensive, difficult to update and error-prone.

This article proposes a conceptual model of a recommendation system based on the following components:

- graph model of labor market requirements based on the ESCO taxonomy adapted for the Russian language;
- an intelligent method of forming recommendations for compiling a list of requirements in the recruitment process based on neural network language models using the ESCO skills taxonomy and the corpus of online vacancies of the Russian labor market;
- a model and conceptual algorithm of the automated recommendation system for the formation of recommendations;
- possible options for recommendations on updating the list of requirements of the job profile based on an analysis of the needs of the labor market.

1. Analysis of the results of previous works

In recent years, there has been an increasing interest in the use of artificial intelligence (AI) methods for analyzing data on the labor market – "labor market intelligence" (LMI). LMI means the development and use of AI methods, algorithms and structures for analyzing labor market data that help with policy planning and decision-making [7–9].

For example, in [10, 11], research is aimed at creating recommendation systems that determine the correspondence between the applicant's resume and the vacancy for specific competencies at the level of determining the position. Other works are aimed at determining the demand for certain skills [12], which can help students determine their educational trajectory or direction of retraining and increase their level of competitiveness.

Due to the rapid development of computational linguistics and tools for analyzing texts in natural language, some scientists are trying to analyze changes in the labor market based on the texts of online vacancies at the level of individual competencies [13–16]. This approach has many advantages, as it allows us to identify changes at the level of specific professions and specialties, as well as the requirements of employers. For example, it allows one to monitor online vacancies in different regions and countries in real time, predicting the demand for individual skills, competencies and technologies within specific professions or industries, as well as quickly comparing similar labor markets in different countries and regions.

The project of the European Center for the Development of Vocational Education (Cedefop) deserves special attention, since its goal is to collect and classify online vacancies for the entire EU using machine learning [17–19]. In addition, within the framework of this project, research is being conducted to identify trends in the labor market and predict the demand for individual skills. For example, in [20], the authors, using the methods of intellectual text analysis, analyze the literature in the category "fourth technological revolution" and compare the results with the new version of the ESCO skills classification to determine to what extent the new version of the ESCO skills classification, created by experts manually, reflects the trends occurring in the real market labor.

2. Methods and materials

2.1. Overview of the European Taxonomy of ESCO Skills

The ability to extract valuable knowledge from large amounts of data, such as online recruitment systems, strongly depends on the availability of up-to-date knowledge bases, taxonomies and thesauri. Such resources are necessary for the effective application of machine learning methods and for solving most NLP (natural language processing) and NLU (natural language understanding) tasks.

Currently, a large number of labor market analysis projects are based on the European ESCO Skills Classification. ESCO (European Skills, Competencies and Professions) is a multilingual classification of European skills, competencies, qualifications and professions. It defines and classifies skills, competencies, qualifications and professions corresponding to the EU labor market, education and vocational training, in 25 European languages. The system provides professional profiles showing the relationship between professions, skills, competencies and qualifications. ESCO was developed in an open IT format, is available for free use by everyone and is available through an online portal.

ESCO is structured on the basis of three interrelated components, representing a searchable database in 28 languages. These main elements are: a) Professional Profiles (professions), b) Skills/competencies/knowledge and c) Qualifications, as shown in *Fig. 1* of the ESCO data model. The first component – professional profiles (or professions) contains the name, description of the profession and shows whether skills and competencies and knowledge are necessary or optional, and which qualifications are relevant to each profession. The second component contains information about knowledge, skills and competence, as well as some group concepts. ESCO vl contains about 13 500 concepts (and if you include alternative names, then almost 100k formulations) organized into a hierarchy and is also structured through communication with professions. The third component, the qualifications component, allows States and assigning authorities to provide data on qualifications that are collected in ESCO. Qualifications are structured using the European Qualifications Framework (EQF) and ISCED Fields of Education and Training 2013.

Currently, separate groups of scientists are proposing approaches and algorithms for automatically expanding the taxonomy of ESCO skills based on open online job data [21, 22].

For example, for the profession "computer equipment engineer", ESCO code 2152.1.1, 22 alternative names of the profession are defined in the taxonomy; an incomplete list of alternative names is given in *Table 1*. Also, a number of entities skill/competence and knowledge that are necessary for this profession are defined for this profession, each of which also includes a list of alternative names in natural language. For example, 47 basic and 25 additional skills/competencies, 16 basic and 20 additional knowledge entries are defined for this profession; examples of names for the entities of skills/competencies and knowledge, as well as their alternative names are presented in *Table 2*.

The most important advantage of this classification is that it uses the wording of the names of the profession, the names of skills / competencies and knowledge, as well as their alternative names – in natural

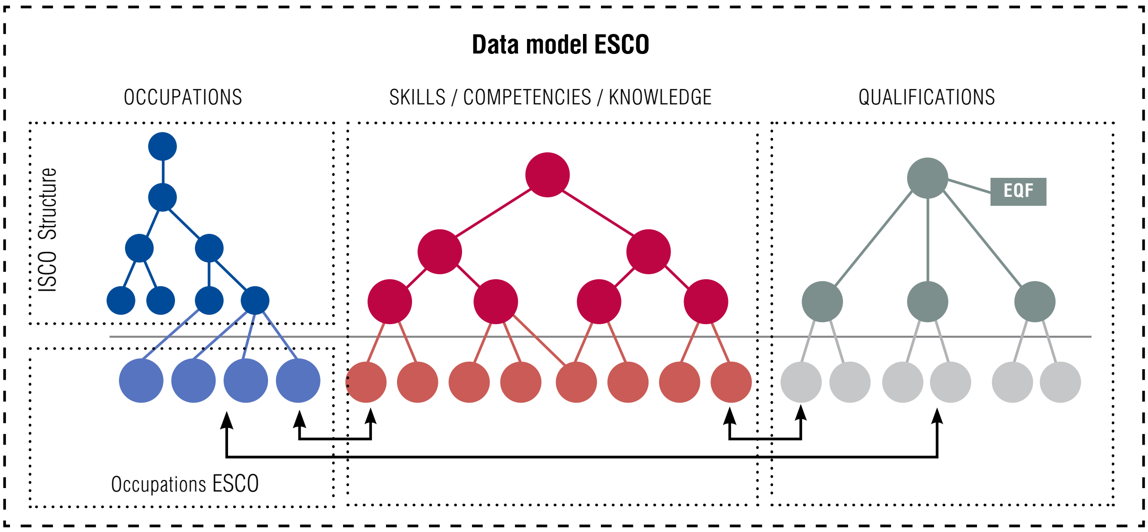

Fig. 1. Data Model ESCO.
Source: https://ec.europa.eu/esco/portal/escopedia/ESCO_data _model

Table 1.

Examples of alternative names for the profession "computer equipment engineer"

Name
computer equipment specialist computer
computer engineer
engineer PC hardware engineer
IT equipment specialist

language, which greatly simplifies and expands the possibilities of its application for analyzing the texts of online vacancies of the real labor market, which are also presented in natural language, modern machine learning methods.

2. The model of the recommendation system for the formation of current requirements of the profile of the position

2.1. Generalized graph model of labor market requirements based on ESCO classification and data from online vacancies

Let's imagine a labor market model as a directed graph. Let's take the ESCO taxonomy as a basis.

Definition of a graph model of the labor market based on the ESCO taxonomy. The graph model is represented as a tuple of three elements $E = (O, R, S)$, where $O = \{o_1, ..., o_n\}$ – set of occupations, $S = \{s_1, ..., s_m\}$ – set of multiple entities of skills/competencies and knowledge, and $R : O \cdot S \to B$ – the relation that connects occupation o with skill s, namely $r(o, s) = 1$ if skill s is associated with occupation o in ESCO, and 0 otherwise.

It is worth noting that one profession can be associated with several entities of skills/competencies and knowledge, and one entity of skills/competencies or knowledge can relate to several professions (*Fig. 2*).

Definition of an online vacancy. Online vacancy j is represented as a tuple $V = (i, c, p, t)$, where $i \in N$ is a unique identifier, $c \in C$ is a unique identifier of the industry, $p \in P$ is a unique identifier of the profession of the vacancy, $t \in T$ is the text of the requirements from the vacancy.

2.2. Calculating the importance of skills for a profession

Since the same skill in the ESCO structure can be associated with several professions, a tool is needed to assess the importance of a skill for a particular profession.

The importance of skills for each profession can be assessed using the RCA tool, originally used in the context of research in the USA [23], where the authors used the O*NET skill classification (the American equivalent of ESCO) to take into account the importance of each skill for each profession. The importance (frequency) of skills for professions $o_i \in O$ and skills $s_l \in S$ is determined by the formula (1), where I denotes the indicator function. The rca function is calculated by the formula (2), where sf is the frequency of the skill s_l for the profession o_i.

To get a more understandable measure, we calculate the normalized rca by formula (3), normalizing the *rca* with respect to the maximum value obtained for the profession in question, so that the most sought-after skill for each profession has a normalized *rca* equal to 1.

$$sf(o_i, s_l) = \frac{\sum_{k=1}^{m} I(o_k = o_i) \cdot I(s_i = s_l)}{\sum_{k=1}^{m} I(o_k = o_i)}, \quad (1)$$

$$rca(oi, sl) = \frac{sf(o_i, s_l) / \sum_{j=1}^{p} sf(o_{ij}, s)}{\sum_{k=1}^{m} sf(o_k, s_l) / \sum_{k=1}^{m} \sum_{j=1}^{p} sf(o_k, s_j)}, \quad (2)$$

$$rca_N(o_i, s_l) = \frac{rca(oi, sl)}{max_j rca(oi, sl)}. \quad (3)$$

Table 2.

Examples of the name for the entities skill/competence and knowledge for the profession "computer equipment engineer"

Priority name	Alternative names	Type
assemble hardware components	assembly of computer equipment, installation of equipment, assembly of computer components, assembly of computer components	skill / competence
installing the software	computer software installation, software download, computer software download, computer software, installation, software installation, software download, …	skill / competence
create technical plans	create plans for technical details, create industrial plans, create technical drawings	skill / competence
principles of electricity	electric current, voltage, electricity physics, electricity science, electricity theory, resistance, voltage	knowledge
hardware components	hardware components of the system, types of hardware components, hardware components, components for hardware systems, parts for hardware systems, components of hardware systems, hardware parts of the system, typology of hardware components	knowledge

2.3. Formation of recommendations based on semantic comparison of the initial list of requirements from the position profile with the graph model of the labor market

The idea of the method of forming recommendations for updating the list of requirements when compiling a position profile comes from the semantic comparison of individual entities from the initial list of requirements and entities from the graph model of the labor market.

In the proposed method, the following stages can be distinguished:

1. Creating a graph model of the labor market based on the classification of ESCO skills.

2. Expansion of the graph model due to information from the texts of online vacancies of the real labor market.

3. Comparison of the initial list of requirements with the entities of the graph model of the labor market.

4. Ranking of matching results based on the RCA metric.

5. Formation of recommendations for inclusion in the initial list of requirements.

Steps 3–5 can be repeated several times, which will allow you to form a more accurate and up-to-date list of requirements with each new iteration.

The recommendation system model is shown in *Fig. 3.*

3. Conceptual algorithm of the system's functioning

A conceptual algorithm is an abstract description of the process of solving a problem or performing a certain action without specifying detailed instructions or

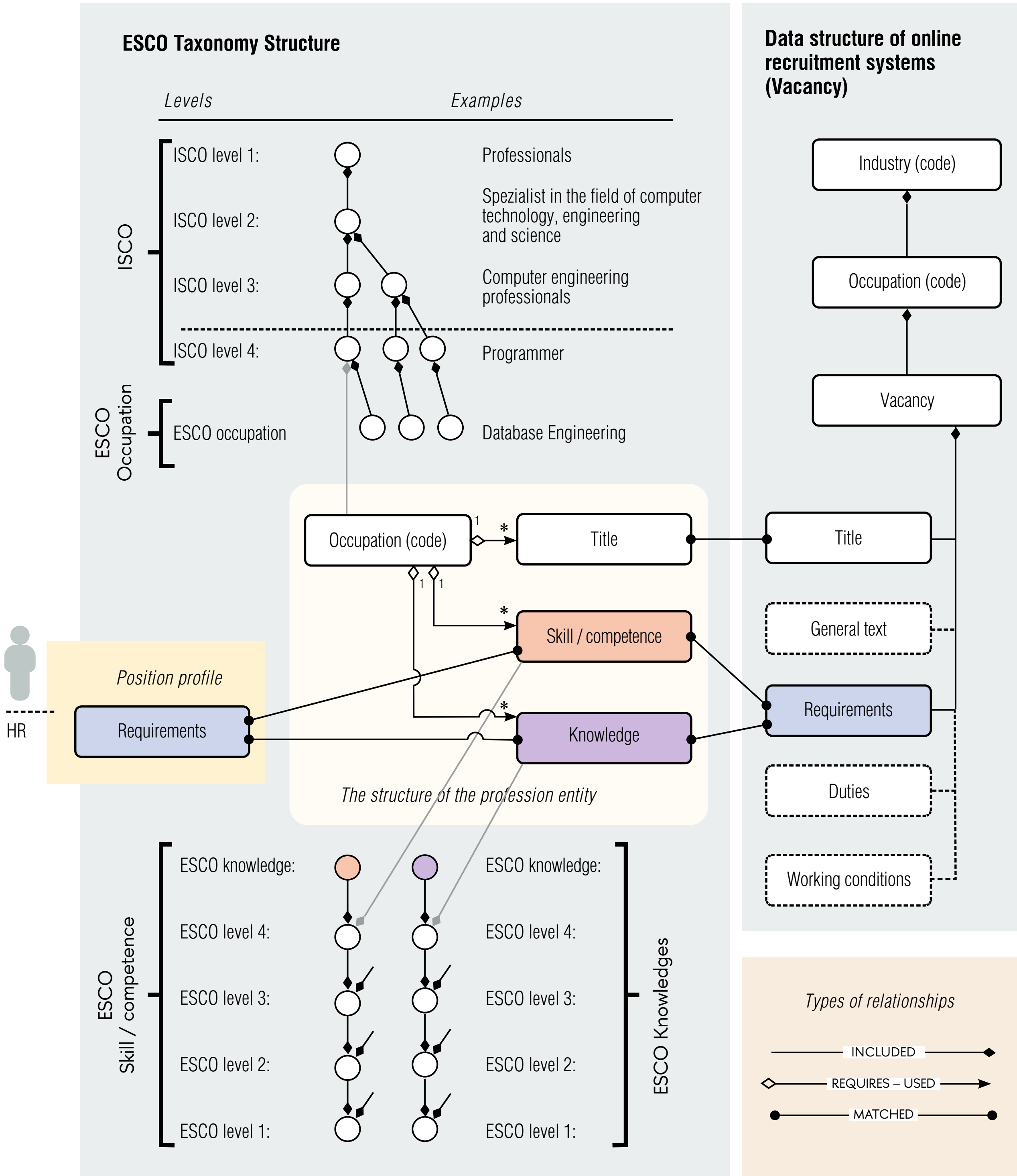

Fig. 2. The main entities and relationships in the generalized graph model of labor market requirements based on the ESCO classification.

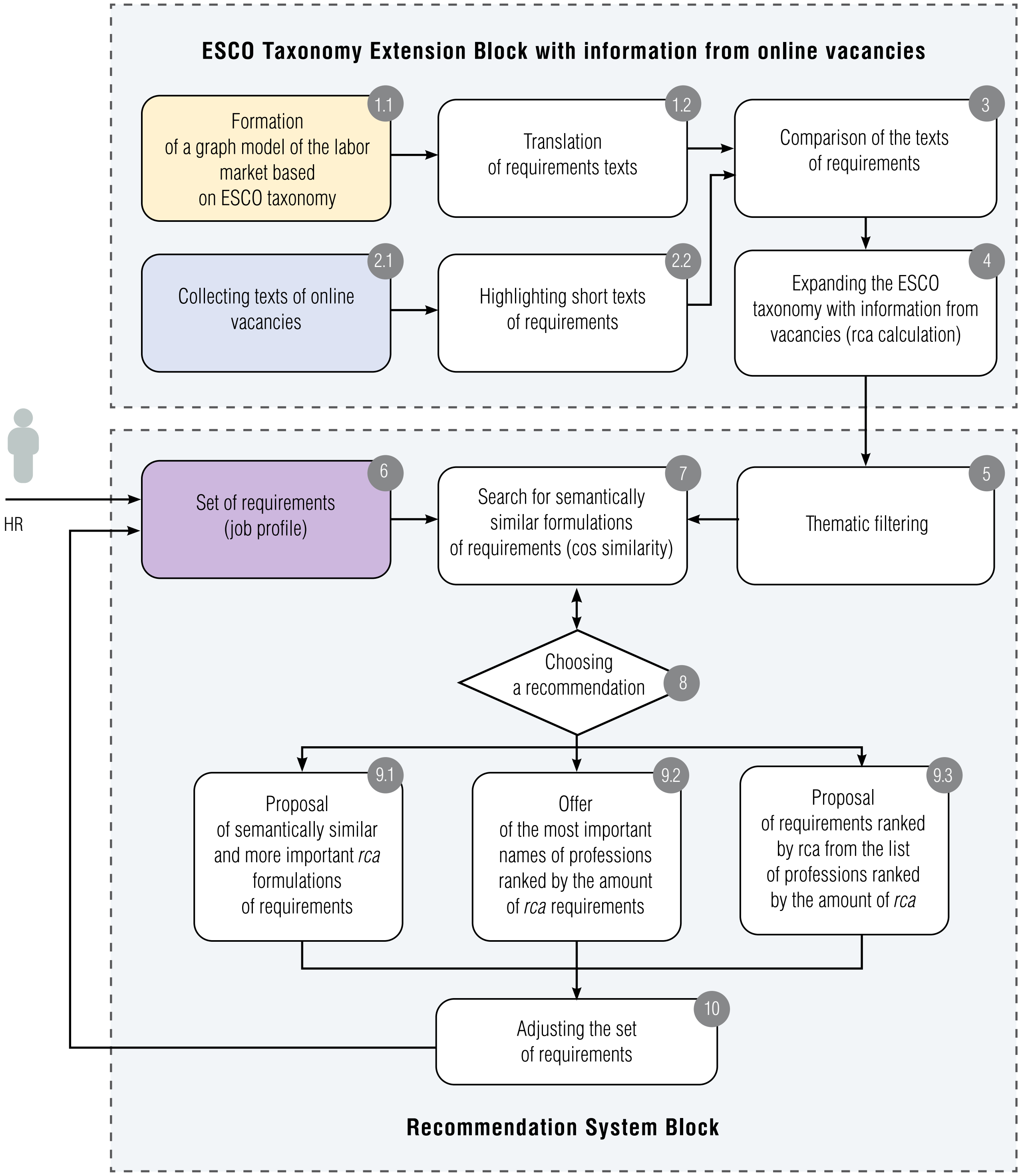

Fig. 3. Model of the recommendation system for updating the list of requirements when compiling a position profile.

programming language. This concept is used in computer science, mathematics and other scientific fields where the process of solving a problem is important, and not a specific code or programming language. In scientific publications, this term can be used to discuss general approaches to solving problems, without reference to specific technologies or implementations.

The result of our work was the development of a conceptual algorithm for the operation of the recommendation system:

1.1. Formation of a graph model of the labor market based on the European taxonomy of ESCO skills (see description of the graph model).

1.2. Adaptation of the graph model of labor market skills for the Russian language. With the help of automatic translation services, all formulations of the European taxonomy of skills are translated, while the structure of relations between the entities of the taxonomy is preserved.

2.1. Collecting online vacancies from Russian online recruitment systems. Many online recruitment systems support an open API for obtaining job data (for example api.hh.ru, api.superjob.ru and others).

2.2. Selection of short texts of requirements from texts of vacancies. There are several possible options for selecting short texts of requirements from the texts of vacancies:

♦ search for direct matches of entity formulations from the graph model of the labor market with texts from online vacancies;
♦ developing rules and finding matches based on rules (for example, using the yargy parser from the natasha library for python (https://natasha.github.io/);
♦ and the third option, preparing a training dataset and training a model for the NER task. For example, using neural network models to analyze natural language texts from the deeppavlov library (https://deeppavlov.ai /) developed by an innovative AI company – iPavlov, a spin-off of MIPT. Continuation of the successful project "Neurointellect iPavlov", implemented within the framework of the NTI, with the industrial support of Sberbank.

3. Comparison of the texts of requirements from online vacancies with the entities of the graph model of labor market requirements. At this step, it is supposed to use a simple comparison and search for matches between normalized (reduced to normal form) texts of the requirements extracted from the texts of online vacancies and the entities of the graph model.

One of the possible improvements of this step can be suggested using the Russian version of the ruwordnet thesaurus [24], which contains relations of hyponyms, hyperonyms, as well as a dictionary of synonyms for the Russian language. Using this information will allow you to expand the list of variations of the texts of requirements when comparing.

4. For all formulations of the graph model, the *rca* parameter is considered, which actually indicates the importance of a skill for a particular profession.

It is worth noting that the expansion of the ESCO taxonomy may occur not just with statistical data, but may also represent a more complex process, for example, the search for new formulations of skills/competencies or knowledge and their integration into an existing graph model [21, 22].

5. Thematic filtering. At this step, using thematic modeling tools, the user can be offered terms and concepts automatically grouped by topic. The user can select a list of words or concepts that must necessarily be contained or absent in the texts of the final output. The possibility of using such a thematic filter was demonstrated by the author in the article [25].

6. The user of the system (for example, an HR specialist) forms an initial set of requirements and this is submitted to the system input.

7. Comparison of the texts of requirements from the user's request and entities from the graphical model of the labor market. In this case, it is assumed that we use a more intelligent matching process which can be divided into two stages: with the help of modern neural network models built on the architecture of transformers (RuBERT, Robert and others), vector representations are obtained for the texts of user requirements and for the entities of the graph model. Then, using a cone proximity measure, vector rep-

resentations are matched in pairs. Further, all texts of requirements that lie beyond a certain permissible distance (determined experimentally) are cut off, and which are the most semantically close to the text of the original requirement. This is how semantically similar texts of requirements from the graph model are determined for all texts from the user's original list of requirements.

The algorithm for choosing the most effective neural network model that would allow generating the best (from the point of view of compactness) vector representations for semantically similar texts of requirements was considered by the author in the article [26].

8. The user chooses which type of recommendation he would like to form for the initial list of requirements.

9.1. Semantically similar texts from the graph model selected in step 7 are ranked by the *rca* parameter. The arranged formulations are demonstrated to the user, indicating the *rca* parameter and profession. In fact, at this stage, the user gets the opportunity to select the most important and semantically similar requirements, and he decides to add the proposed requirements to his initial list.

9.2. Semantically similar texts from the graph model selected in step 6 are ranked by the *rca* parameter. For the totality of the requirements of the initial list, the formulas with the highest *rca* are selected. For all requirements and professions, the amount is considered. The sessions are ranked by the amount of *rca* for the initial list of requirements (*Fig. 4*). From the graph model, *n* profession names are selected based on the largest amount of *rca* for the initial list of user requirements and are shown to the user.

9.3. Just like in 9.2, the *rca* sum is calculated. *N* profiles are selected from the graph model based on the largest amount of *rca* for the initial list of user requirements (*Fig. 4*). From each selected profession, the *M* most important *rca* texts of skills/competencies or knowledge are selected and demonstrated to the user.

10. After studying the proposed recommendations, the system user can choose the options that will be added to the initial list of requirements.

Steps 5 through 10 can be repeated over again, which will iteratively refine and improve the initial list of requirements set by the user.

Examples of possible options for modifying the set of requirements of the position profile based on the recommendations formed:

- Propose a new alternative formulation for the existing requirement based on a higher *rca*.
- Propose to include a new requirement in the position profile based on a high *rca*, taking into account the relationship with existing requirements.
- Recommend to exclude a requirement from the position profile based on a low *rca*, taking into account the relationship with existing requirements.
- Break down the requirements into categories by profession
- Suggest the title of the position based on the list of requirements of the position profile.
- Ranking of requirements according to their demand based on the *rca* indicator.

The list of recommendations can be adjusted towards expansion by adding new functional blocks to the system.

The computational complexity of the entire system is estimated as low. The most time-consuming stages, such as extracting the entities of knowledge, skills and competencies from the texts of vacancies, as well as their vector representation, occur once and can occur in the background. Thematic filtering and *rca* counting are relatively simple computational operations. The most difficult, from the point of view of computational complexity, is the operation of ranking a large number of requirements relative to each other based on a cosine measure of proximity. To perform a highly productive search for similar vectors of requirements, there are plans to use the FAISS library (Facebook AI Similarity Search) [26, 27]. This library provides a set of algorithms for indexing large sets of vectors and quickly searching for nearest neighbors in these sets. The library was developed by Facebook AI Research and is distributed under the terms of the Apache 2.0 license.

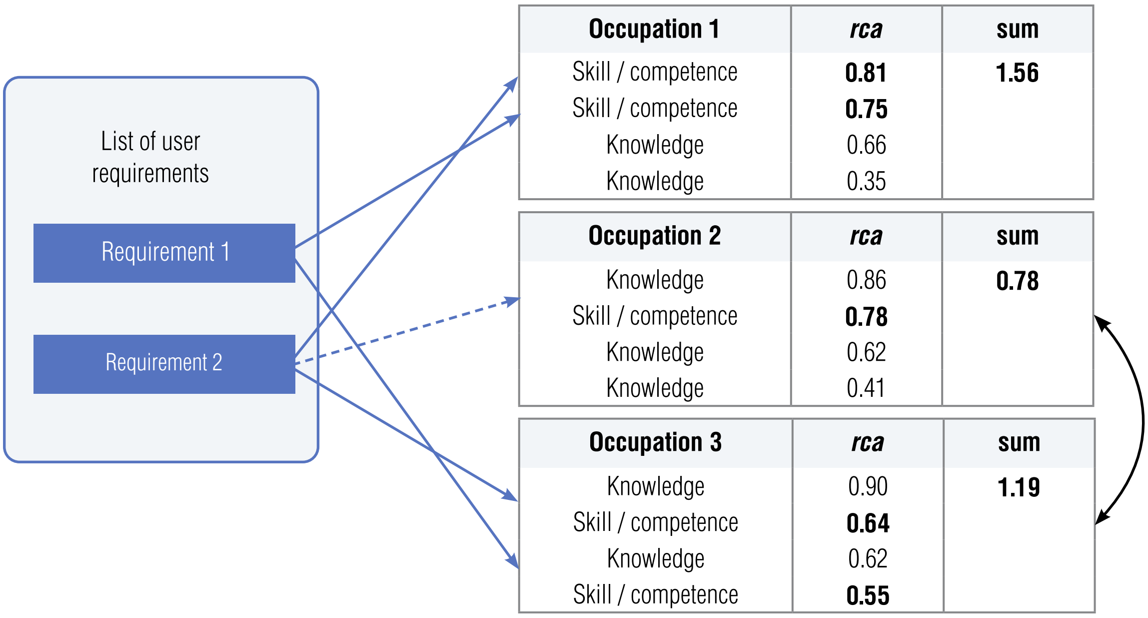

Fig. 4. Ranking of the profession by the amount of *rca* for the initial list of requirements.

Conclusion

The development of a recommendation system for the formation of an up-to-date list of job profile requirements based on an analysis of the real requirements of the labor market is an important step in the development of HR technologies and will allow us to: significantly reduce the labor costs of HR specialists; more accurately and systematically determine the requirements for candidates for various positions; identify potential opportunities for retraining employees; better understand how changes in the requirements of the labor market affect companies and their personnel; form more flexible and adaptive strategies for attracting and managing personnel.

The proposed conceptual model of the recommendation system includes:

- Graph model of labor market requirements based on ESCO taxonomy adapted for the Russian language;
- An intelligent method of forming recommendations for compiling a list of requirements in the recruitment process based on neural network language models using the ESCO skills taxonomy and the corpus of online vacancies of the Russian labor market. Within the framework of the method, we propose to use neural network models of the language built on the architecture of transformers (models of the BERT family) to assess the semantic proximity of the entities of the initial list of requirements with the graph model of the labor market;
- Model and conceptual algorithm of the automated recommendation system for the formation of recommendations.

This article also provides possible options for recommendations on updating the list of requirements of the position profile based on an analysis of the needs of the real labor market.

To improve this model, the author additionally plans to: develop a method for extracting individual short texts of knowledge and skills from the texts of the requirements of online vacancies of the real labor market; develop a system for automatically expanding the graph model with texts of knowledge and skills; integrate an extended graph model of labor market requirements through the international system of classification of occupations (ISCO) with all-Russian classifiers (OKZ, the all-Russian classifier of occupations, OKVED, the all-Russian classifier of economic activities), and professional standards of the Russian Federation. In addition, a separate and important task is to develop a method for assessing the quality and investigating the effectiveness of using a recommendation system based on the proposed model in various sectors of the economy and in various labor markets. ■

References

1. Boselli R., Cesarini M., Mercorio F. (2017) Using machine learning for labour market intelligence. *Machine Learning and Knowledge Discovery in Databases: European Conference*, pp. 330–342. https://doi.org/10.1007/978-3-319-71273-4_27

2. Colombo E., Mercorio F., Mezzanzanica M. (2019) AI meets labor market: Exploring the link between automation and skills. *Information Economics and Policy*, vol. 47, pp. 27–37. https://doi.org/10.1016/j.infoecopol.2019.05.003

3. Giabelli A., Malandri L., Mercorio F. (2022) GraphLMI: A data driven system for exploring labor market information through graph databases. *Multimedia Tools and Applications*, vol. 81, pp. 3061–3090. https://doi.org/10.1007/s11042-020-09115-x

4. Mezzanzanica M., Boselli R., Cesarini M., Mercorio F. (2015) A model-based approach for developing data cleansing solutions. *Journal of Data and Information Quality*, vol. 5, no. 4, pp. 1–28. https://doi.org/10.1145/2641575

5. Xu T., Zhu H., Zhu C., Li P., Xiong H. (2018) Measuring the popularity of job skills in recruitment market: A multi-criteria approach. Proceedings of the *AAAI conference on artificial intelligence*, vol. 32, no. 1. https://doi.org/10.1609/aaai.v32i1.11847

6. Zhang D., Liu J., Zhu H., Liu Y., Wang L., Wang P., Xiong H. (2019) Job2Vec: Job title benchmarking with collective multi-view representation learning. Proceedings of the *28th ACM International Conference on Information and Knowledge Management*, pp. 2763–2771. https://doi.org/10.1145/3357384.3357825

7. *The importance of LMI* (2015) UK Commission for Employment and Skills. Available at: https://www.gov.uk/government/publications/the-importance-of-labour-market-intelligence (accessed 30.01.2023).

8. Mezzanzanica M., Mercorio F. (2019) Big data enables labor market intelligence. *Encyclopedia of Big Data Technologies*, pp. 226–236. https://doi.org/10.1007/978-3-319-63962-8_276-1

9. Eggertsson T. (1990) *Economic behavior and institutions: Principles of neoinstitutional economics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511609404

10. Qin C., Zhu H., Xu T., Zhu C., Jiang L., Chen E., Xiong H. (2018) Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. Proceedings of the *41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*, pp. 25–34. https://doi.org/10.1145/3209978.3210025

11. Zhu C., Zhu H., Xiong H., Ma C., Xie F., Ding P., Li P. (2018) Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Transactions on Management Information Systems*, vol. 9, no. 3, pp. 1–17. https://doi.org/10.1145/3234465

12. Xu T., Zhu H., Zhu C., Li P., Xiong H. (2018) Measuring the popularity of job skills in recruitment market: A multi-criteria approach. Proceedings of the *AAAI Conference on Artificial Intelligence*, vol. 32, no. 1. https://doi.org/10.1609/aaai.v32i1.11847

13. Vinel M., Ryazanov I., Botov D., Nikolaev I. (2019) Experimental comparison of unsupervised approaches in the task of separating specializations within professions in job vacancies. *Artificial Intelligence and Natural Language. AINL 2019. Communications in Computer and Information Science* (eds. D. Ustalov, A. Filchenkov, L. Pivovarova), vol. 1119. Springer, Cham. https://doi.org/10.1007/978-3-030-34518-1_7

14. Nikolaev I., Ryazanov I., Botov D. (2020) The comparison of distributive semantics models applied to the task of short job requirements clustering for the Russian labor market. Proceedings of the *8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)*. Atlantis Press, pp. 295–301. https://doi.org/10.2991/aisr.k.201029.056

15. Colombo E., Mercorio F., Mezzanzanica M. (2018) Applying machine learning tools on web vacancies for labour market and skill analysis. *Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence. Technology Policy Institute Conference on the Economics and Policy Implications of Artificial Intelligence*. Available at: https://techpolicyinstitute.org/wp-content/uploads/2021/03/Colombo_paper.pdf (accessed 30 May 2023).

16. O'Kane L., Narasimhan R., Nania J., Taska B. (2020) *Digitalization in the German labor market: Analyzing demand for digital skills in job vacancies*. Gütersloh: Bertelsmann Stiftung.

17. *Real-time labour market information on skill requirements: Setting up the EU system for online vacancy analysis (2016)* Cedefop. Available at: https://www.cedefop.europa.eu/it/about-cedefop/public-procurement/real-time-labour-market-information-skill-requirements-setting-eu (accessed 30 January 2023).

18. Boselli R., Cesarini M., Mercorio F., Mezzanzanica M. (2018) Classifying online job advertisements through machine learning. *Future Generation Computer Systems*, vol. 86, pp. 319–328. https://doi.org/10.1016/j.future.2018.03.035

19. Boselli R., Cesarini M., Marrara S., Mercorio F., Mezzanzanica M., Pasi G., Viviani M. (2018) WoLMIS: A labor market intelligence system for classifying web job vacancies. *Journal of Intelligent Information Systems*, vol. 51, pp. 477–502. https://doi.org/10.1007/s10844-017-0488-x

20. Chiarello F., Fantoni G., Hogarth T., Giordano V., Baltina L., Spada I. (2021) Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach. *Technological Forecasting and Social Change*, vol. 173, article 121177. https://doi.org/10.1016/j.techfore.2021.121177

21. Giabelli A., Malandri L., Mercorio F., Mezzanzanica M., Seveso A. (2020) NEO: A tool for taxonomy enrichment with new emerging occupations. *The Semantic Web – ISWC 2020. Lecture Notes in Computer Science*, vol. 12507, pp. 568–584. https://doi.org/10.1007/978-3-030-62466-8_35

22. Malandri L., Mercorio F., Mezzanzanica M., Nobani N. (2020) Meet: A method for embeddings evaluation for taxonomic data. Proceedings of the *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 31–38. Sorrento, Italy. https://doi.org/10.1109/ICDMW51313.2020.00014

23. Alabdulkareem A., Frank M.R., Sun L., AlShebli B., Hidalgo C., Rahwan I. (2018) Unpacking the polarization of workplace skills. *Science Advances*, vol. 4, no. 7, eaao6030. https://doi.org/10.1126/sciadv.aao6030

24. *The Russian language Thesaurus RuWordNet*. Available at: https://ruwordnet.ru/ru (accessed 30 May 2023).

25. Nikolaev I., Botov D., Dmitrin Y., Klenin J., Melnikov A. (2019) Use of topic modelling for improvement of quality in the task of semantic search of educational courses. Proceedings of the *21st International Workshop on Computer Science and Information Technologies (CSIT 2019)*, pp. 104–111. Atlantis Press. https://doi.org/10.2991/csit-19.2019.18

26. Nikolaev I., Melnikov A. (2022) Comparison of neural network models based on transformer architecture in the context of the task of evaluating the compactness of vector representations of semantically similar texts of the requirements of the European ESCO skills classification. *Bulletin of the South Ural State University. Series: Computer Technology, Control, Radio Electronics*, vol. 22, no. 3, pp. 19–29.

27. *Faiss. A library for efficient similarity search and clustering of dense vectors*. Available at: https://faiss.ai/ (accessed 30 May 2023).

28. Johnson J., Douze M., Jégou H. (2019) Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, vol. 7, no. 3, pp. 535–547.

About the author

Ivan E. Nikolaev

Senior Lecturer, Department of Information Technology, Institute of Information Technologies, Chelyabinsk State University, 129, Kashirin Brothers Str., Chelyabinsk 454001, Russia;

E-mail: ivan_nikolaev@csu.ru

ORCID: 0000-0002-9686-2435

DOI: 10.17323/2587-814X.2023.2.85.97

The present and future of the digital transformation of real estate: A systematic review of smart real estate

Fengchen Wang
E-mail: fwang@hse.ru

HSE University
Address: 3A-1, Kantemirovskaya Street, St. Petersburg 194100, Russia

Abstract

The contribution of the real estate industry to the global and regional economy is remarkable, yet in today's evolving digital technology and digital economy, the digital transformation of the real estate industry is lagging behind other industries. This is, on the one hand, due to the solidified processes and systems linked to the upstream and downstream real estate industries, and, on the other hand, due to the fact that digital technology disrupts traditional ways of doing business, making the industry full of uncertainty. The digital transformation of the real estate industry is a broad and emerging concept. Various related research fields are concerned with the penetration and application of different innovative technologies to the industry. This study provides a systematic review focusing on the field of smart real estate using the bibliometric analysis approach under the guidance of PRISMA. The bibliometric analyses were performed in RStudio by utilizing 22 scientific documents indexed in Scopus and Web of Science that were published from 2012 to 2022. The findings suggest that: (i) smart real estate research is still a new but rapidly emerging field; (ii) only limited academic institutions from a few countries, such as the University of New South Wales in Australia, have shown significant contributions; (iii) the research exhibits specific collaborative network characteristics, leading to a high concentration of authors and citations; and (iv) data-driven topics such as "machine learning," "information management," "data analytics" and "big data" indicate a high degree of research density and centrality.

Keywords: smart real estate, digital transformation, digital economy, innovative technology, data-driven, property technology (PropTech), systematic review, bibliometric analysis, PRISMA, RStudio

Citation: Wang F. (2023) The present and future of the digital transformation of real estate: A systematic review of smart real estate. *Business Informatics*, vol. 17, no. 2, pp. 85–97. DOI: 10.17323/2587-814X.2023.2.85.97

Introduction

Digital technologies are driving worldwide innovation and disruption across numerous industries. However, the real estate industry has been slow to embrace technology, and this trend may take some time to catch on [1]. The real estate industry is thought to be difficult to innovate in due to the presence of hard barriers in solidified processes and systems [2], and digital technologies also present disruption to the old way of doing business and necessitate significant changes by organizations to compete in the new environment [3]. Smartness is always an indicator for innovative digital technology, and the real estate industry is empowered with such smartness through the applications of several technologies, including virtual and augmented realities (VR and AR), big data, robotics, 3D scanning drones, clouds, software as a service (SaaS), wearable technologies [1], digital twin, and CyberGIS [4]. The real estate industry plays a pivotal role in economic development at both global and regional level; therefore, it is imminent to promote the digital transformation of the real estate industry and conform to the transformation of Industry 4.0 [5] and Marketing 4.0 [6].

The purpose of this study is to systematically review the extant literature on smart real estate associated with a bibliometric analysis, to investigate current theoretical developments in the literature, and to provide future guidance for both academic scholars and industry practitioners. In line with the purpose, this study develops a novel approach to the design science research methodology of business informatics by introducing the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) [7].

This study is organized as follows: the first section introduces the research background and purpose of the study; followed by the materials and method section, which illustrates the PRISMA flowchart of identification and selection of scientific documents used for bibliometric analysis; the third section provides in-depth insights through the aspects of (i) the development trend of the research field; (ii) the performance of countries and institutions; (iii) trends in authorship and collaborations; and (iv) the analysis of keyword co-occurrences and thematic evolution. Finally, the conclusions and limitations of the study are shown.

1. Materials and method

A systematic review is a review that uses specific, systematic approaches to collect and synthesize the findings of studies that address a clearly defined research purpose [8]. This study employed the PRISMA approach (hereafter referred to as the PRISMA 2020 statement), one of the most frequently consulted approaches for mixed quantitative and qualitative systematic reviews [7]. Unlike previous established design science research of purpose-oriented study, such as Archer's six-step design science research of programming, data collection and analysis, synthesis of objectives and analysis results, development, prototyping, and documentation [9, 10], the PRISMA approach contains a 27-item checklist with detailed explanations of each item considered essential for reporting in systematic reviews, which make such an approach distinctive and enable researchers to provide a transparent, complete, and accurate process while exploring the state of knowledge in a chosen research field and identifying future research priorities [7].

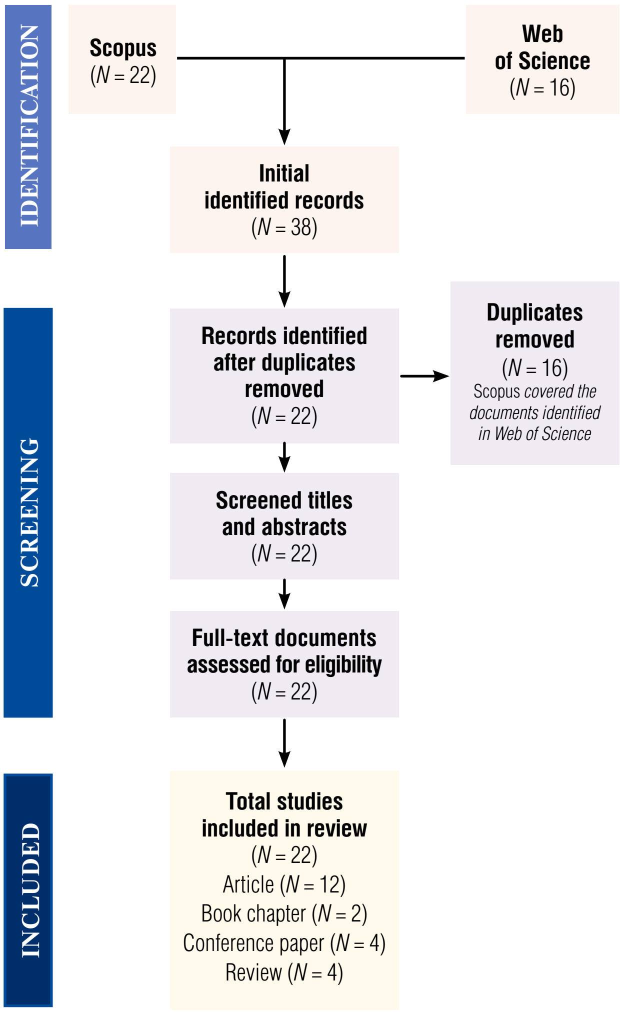

Fig. 1. PRISMA flowchart for scientific documents identification and selection.

On the first stage of scientific document identification, the two main scientific bibliographic databases [11], Web of Science (WoS) and Scopus, were selected for bibliographic data retrieval. The search query "smart real estate" was entered into the topic search in Web of Science and "TITLE-ABS-KEY" in Scopus. In order to ensure data synchronization, data retrieval was done for both databases on the same day, January 9, 2023. Differing from some similar systematic studies (e.g., [12]), the search was not limited by the Web of Sciences core bibliographic collections (e.g., the Social Sciences Citation Index (SSCI)) and no time constraints were set, which were operated similarly in Scopus, and the language of the documents was limited to English for both databases. A total of 38 initially identified scientific records of smart-real estate relevant studies were obtained, of which 22 were included in Scopus and 16 in the Web of Sciences. Both bibliographic metadata were imported into RStudio during the second stage of document screening in order to detect duplicate records and merge the two bibliometric datasets using the *bibliometrix* package in RStudio. *Bibliometrix* is the most widely used R package for bibliometric studies [13], which are increasingly referenced in scientific publications. Following the five-step bibliometric dataset merging with duplicates using the *bibliometrix* package (see *Table 1*), 16 duplicates were identified and found to be covered by the Scopus databases, after which 22 records of scientific documents were confirmed eligible for systematic analysis through the screening of titles and abstracts for each.

On the third stage, 22 scientific documents were included for further analysis in the systematic review, including 12 journal articles, 2 book chapters, 4 conference papers, and 4 reviews. Adhering to the purpose of this study, the comprehensive bibliometric analysis was performed in the Biblioshiny app of the *bibliometrix* package in RStudio, which provides a web interface for the *bibliometrix* package that is used to analyze the bibliographic data in a visualized graphical format and provide insights into the conceptual themes [11].

2. Results and discussion

2.1. Trend of development

Figure 2 shows the trend of research in the field of smart real estate. It is observed that the first research in this field appeared in 2012, followed by a gap in the five years from 2013 to 2017. Since 2018, the research in this field shows a gradual increase in the trend, reaching its

Table 1.

Five-step bibliometric dataset merging with duplicates using the R package "bibliometrix"

Step 1: Download and install package "bibliometrix" in RStudio

```
>install.packages ("bibliometrix")
```

Step 2: Run the "bibliometrix" and "xlsx" package in RStudio

```
>library (bibliometrix)
>library (xlsx)
```

Step 3: Import and convert bibliographic files

```
>wos_data <- convert2df ("wos.txt", dbsource = "wos", format = "plaintext")
#Import and convert the Web of Science dataset "wos.txt" and name the converted dataset as "wos_data"
>scopus_data <- convert2df ("scopus.bib", dbsource = "scopus", format = "bibtex")
#Import and convert Scopus dataset "scopus.bib" and name the converted dataset as "scopus_data"
```

Step 4: Merge the WoS and Scopus database and remove the duplicated data

```
>merged_data <-mergeDbSources (wos_data, scopus_data, remove.duplicated = T)
#Merge the converted datasets and name the merged dataset as "merged_data"
#Duplicates with a count of 16 are automatically removed from the merged dataset
```

Step 5: Export the merged database to "xlsx" file and write the file name as "merged database"

```
>write.xlsx (merged_data, "merged database.xlsx")
#Export the "xlsx" file and name it "merged database.xlsx"
```

peak in 2020, and a slow decrease from 2020 to 2022. Scientific documents published in 2018 received the most average citations per year (36), followed by documents published in 2020 and 2021, which received 14.75 and 20.2 average citations per year, respectively. As shown in *Table 2*, the bibliographic metadata consists of 22 scientific documents contributed by 50 researchers, with an average of 3.05 co-authors per document and an international co-authorship rate of 31.82%, including 12 journal articles, 2 book chapters, 4 conference papers, and 4 review articles. From 2012 to 2022, the annual growth rate was 14.87%, and the average number of citations per document reached 14.45 times.

2.2. Most contributing countries, institutions and authors

2.2.1. Most productive countries and institutions

A systematic review at the country or institutional level indicates the degree of internationalization of a specific research field [14]. A scientific publication from a country when at least one author is affiliated with an institution located in that country [15], and the individual collaboration in the research field adds up to an observable change in the structure of science [14].

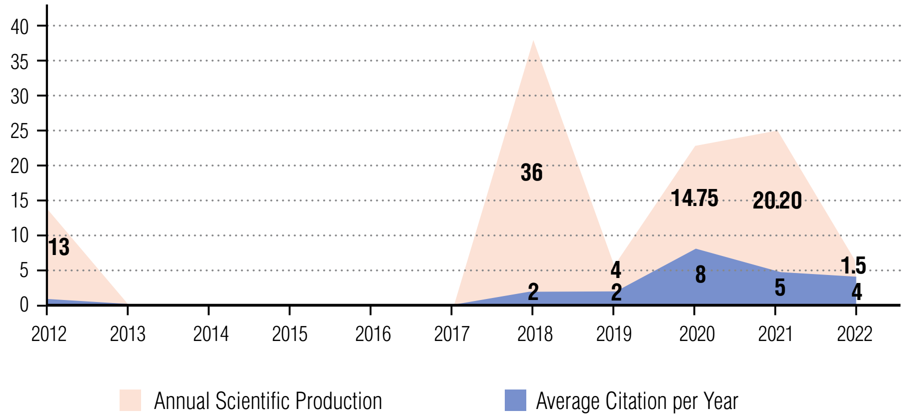

Fig. 2. Annual scientific production and average citations per year.

A total of 14 countries have made their contributions in the research field of "smart real estate," of which 6 countries have produced two or more scientific documents (*Table 3*). Among all the countries, Australia made the most significant contribution, cumulatively contributing nine documents; the total number of citations reached 283 times, and the average number of citations per document reached 31.44 times. Turkey came in second with four documents, 77 citations, and an average citation of 19.25 per document. In addition, the Netherlands and Malaysia produced 3 documents each, and Pakistan and the United Kingdom produced 2 documents each.

Table 4 shows the institutions that contributed more than two publications and the leading authors affiliated with those institutions. The 22 identified scientific documents were contributed by scholars from a total of 26 institutions. Among them, the University of New South Wales in Australia has contributed 8 journal articles, ranking first, co-authored by researchers such as Ullah, Sepasgozar, and Shirowzhan. The Near East University in Turkey ranked second and was led by Al-Turjman, who contributed three journal articles; it is worth noting that all three articles were co-authored with Ullah from the University of New South Wales in Australia. The University of Reading Malaysia ranked third, with two journal articles contributed by Lecomte in 2019 and 2020. In addition, Lecomte published one journal article with the University of Quebec in Montreal in Canada in 2022 in the research filed of smart real estate [16].

2.2.2. Trends in authorship and collaborations

Analysis of authorship and trends in collaboration provides deep insights into the structure and practice of a particular scholarly research field, and the intensity of collaboration between authors and the impact of collaboration on scientific citations vary widely at the international and domestic levels [14].

Research in the field of smart real estate shows a high intensity of author collaboration. As shown in *Fig. 3*, there are ten groups of collaborative networks, among which the collaborative network centered on Ullah and Sepasgozar has the highest collaboration density and the largest number of scientific publications (*Table 3*). Furthermore, as illustrated in *Fig. 4*, seven publications co-authored by Ullah and Sepasgozar (including the co-authorship with Low et al.) ranked among the top ten most cited in the research field; a journal article published in the journal *Sustainability* in 2018 was cited 66 times, ranking first. The remaining 9 groups of collaborative networks exhibited the characteristics of multiple authors cooperating on a single publication: collaborative networks such as Ahmed et al. [26], Allameh et al.[27], Kempeneer et al. [28], Azmi et al. [29], Sandeep Kumar and Talasila [30] each collaborated on one journal paper; collaborative networks such as Hapuarachchi

Table 2.

Descriptive bibliographic metadata

Description	Results
Main information	
Timespan	2012–2022
Documents (Articles, Reviews, etc.)	22
Annual Growth Rate %	14.87
Document Average Age	3.05
Average citations per doc	14.45
References	1590
Document types	
Article	12
Book chapter	2
Conference paper	4
Review	4
Document contents	
Keywords Plus (ID)	200
Author's Keywords (DE)	103
Authors	
Authors	50
Single-authored docs	5
Co-Authors per Doc	3.05
International co-authorships %	31.82

Table 3.

Most productive countries

Country	No. of Documents	Total citations	Average citations per doc
Australia	9	283	31.44
Turkey	4	77	19.25
Netherlands	3	17	5.67
Malaysia	3	16	5.33
Pakistan	2	14	7.00
United Kingdom	2	1	0.50

Note:
The documents identified and included may be co-authored by several scholars from different institutions located in different countries; therefore, the number of documents and total citations corresponding to each country are cumulative.

Table 4.

Most productive institutions with leading authors

Institution	Country	N	Leading author(s)	Documents
University of New South Wales	Australia	8	Ullah, Sepasgozar and Shirowzhan	[1,17–23]
Near East University	Turkey	3	Al-Turjman	[19, 21, 22]
University of Reading Malaysia	Malaysia	2	Lecomte	[24, 25]

Note:
N = number of documents; % = percentage of contributed scientific documents.
"Leading authors" refer to authors who have authored (including co-authored) more than two documents with their corresponding affiliated institutions.

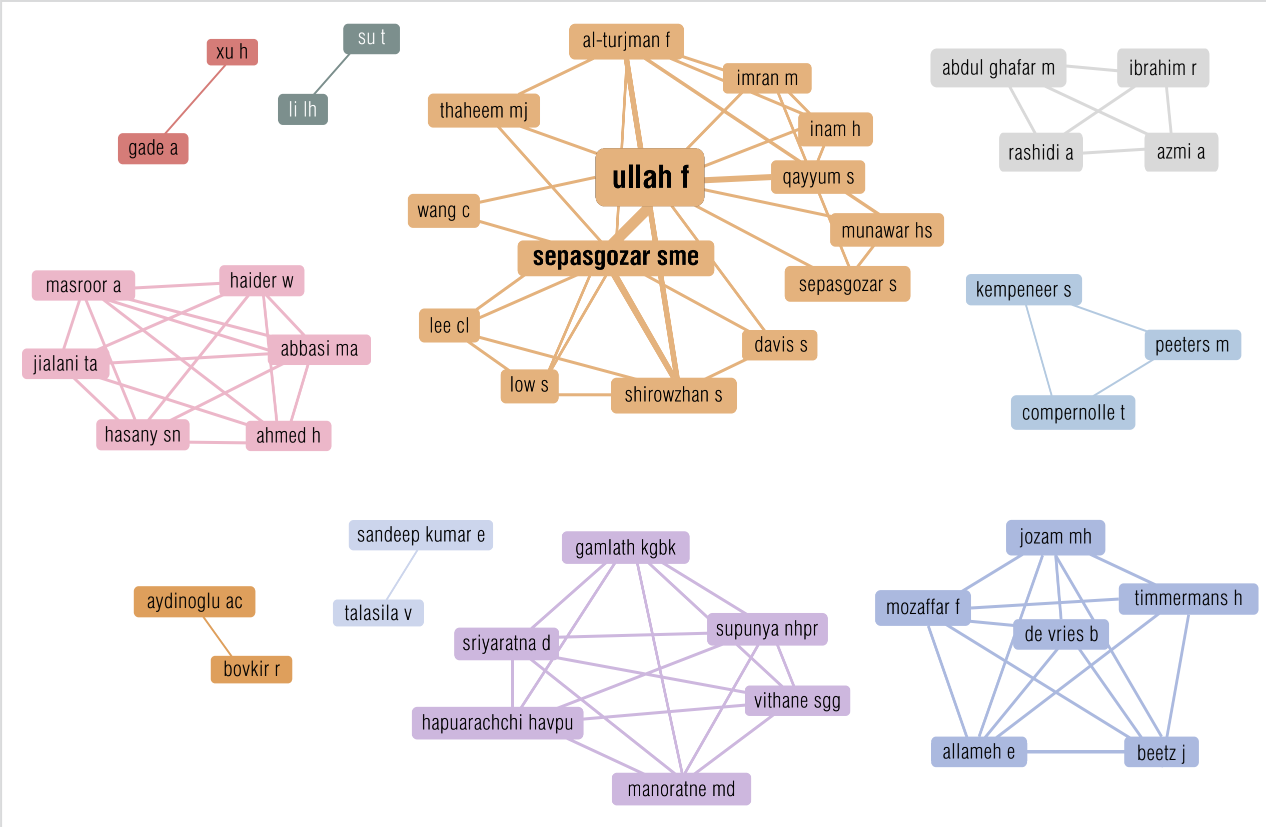

Fig. 3. Collaboration network of authors.

et al. [31], Aydinoglu and Bovkir [32], Xu and Gade [33], and Su and Li [34] each produced one conference paper collaboratively.

2.3. Analysis of keyword co-occurrence and thematic evolution

When analyzing the knowledge of the research field, keyword analysis allows for more systematic insights into the current state and trends for future development. Keywords Plus, containing the author's keywords and words or phrases automatically generated by the computer algorithms that appear frequently in the titles of an article's references, are recommended for co-occurrence analysis of keywords and the evolution of the research themes [35]. In this section, the conceptual structure map using the multiple correspondence analysis method was employed to distinguish the present research themes by categorizing the selected keywords, whereas the thematic map was used to consult the comprehensiveness of each clustered theme by Keywords Plus, thus summarizing the evolution of themes by the degree of development and relevance of each theme cluster.

The bibliographic metadata contains 200 Keywords Plus spread across 22 scientific documents, 29 of which appeared more than twice and were chosen for keyword co-occurrence analysis and thematic evolution. As shown in *Fig. 5*, 29 keywords were categorized into four groups according to the conceptual structures, and the detailed categories of keywords and corresponding documents with the highest contribution are summarized in *Table 5*.

The four different types of themes are distributed in different quadrants according to their degree of development (density on the *y*-axis) and relevance (centrality on

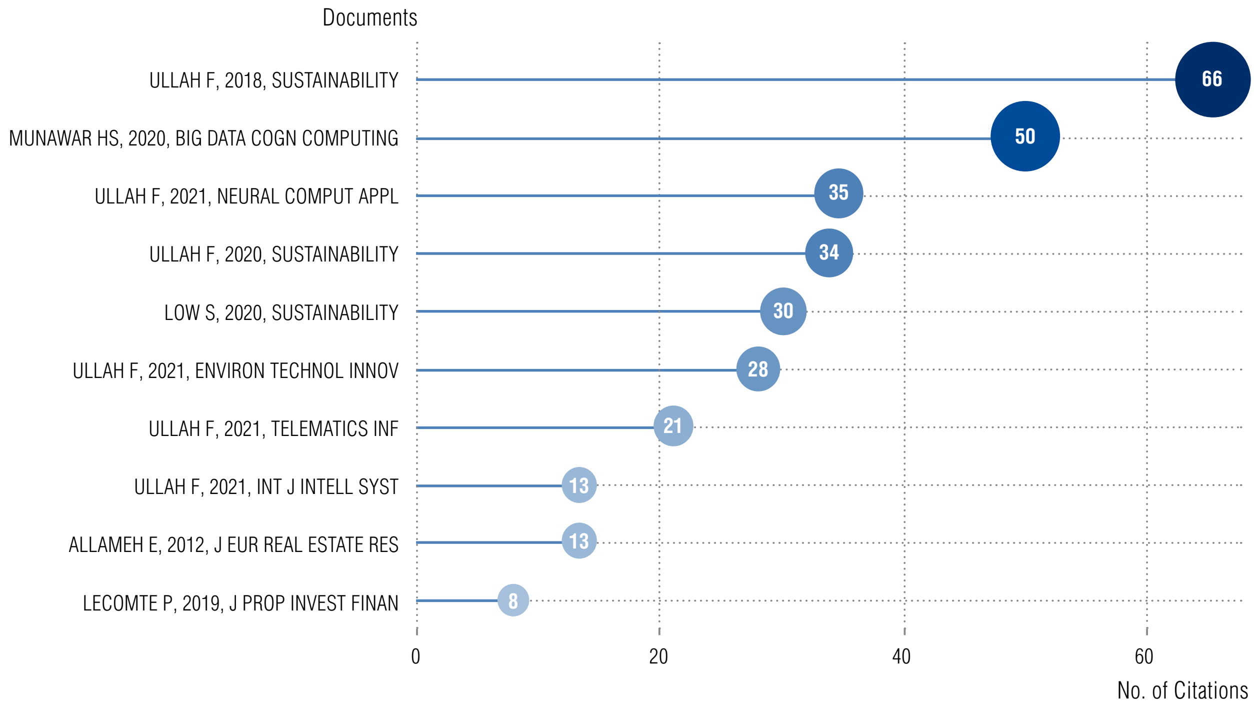

Fig. 4. Most cited documents in the research field of smart real estate.

the *x*-axis): niche themes are located in the upper-left quadrant, motor themes are located in the upper-right quadrant, emerging or declining themes are located in the lower-left quadrant, and basic themes are located in the lower-right quadrant [36]. Shown in *Fig. 6*, we observe:

The cluster consisting of "deep learning," "neural networks," "smart cities," and "taxation" is the only cluster located in the niche theme quadrant; according to [36], those topics are narrowly focused and peripheral in nature, with strong internal linkages but weak external ties, and thus have only a minimal impact on the research field.

The merging or declining themes are both weakly developed and peripheral, showing low density and low centrality. Three clusters are located in the quadrant of emerging or declining themes. These are: the cluster containing "sustainability," "property market," and "software," and the cluster composed of "property," "real estate industry," "sales," and the cluster composed of "real estate agents." According to *Table 5*, the publications that contributed to those keywords were published between 2018 and 2022, resulting in the emerging themes. It is worth noting that the cluster composed of the single keyword "real estate agents" is less developed than the other two located in the same quadrant.

The motor themes exhibit high density and strong centrality; themes locate in this quadrant, suggesting that they are both well-developed and essential for structuring a study topic, as well as tied externally to theories that are relevant to other conceptually related themes [36]. The cluster composed of "machine learning," "information management," "data analytics," "big data," and "life cycle" possesses the highest development degree and relevance degree. The cluster composed of "decision stress," "housing," "investments," "machine learning," "techniques" and "real estate investment" has a medium degree of relevance and a higher degree of development. Meanwhile, the cluster composed of "architectural design," "marketing" and "real estate" has a higher degree of relevance and a medium degree of development.

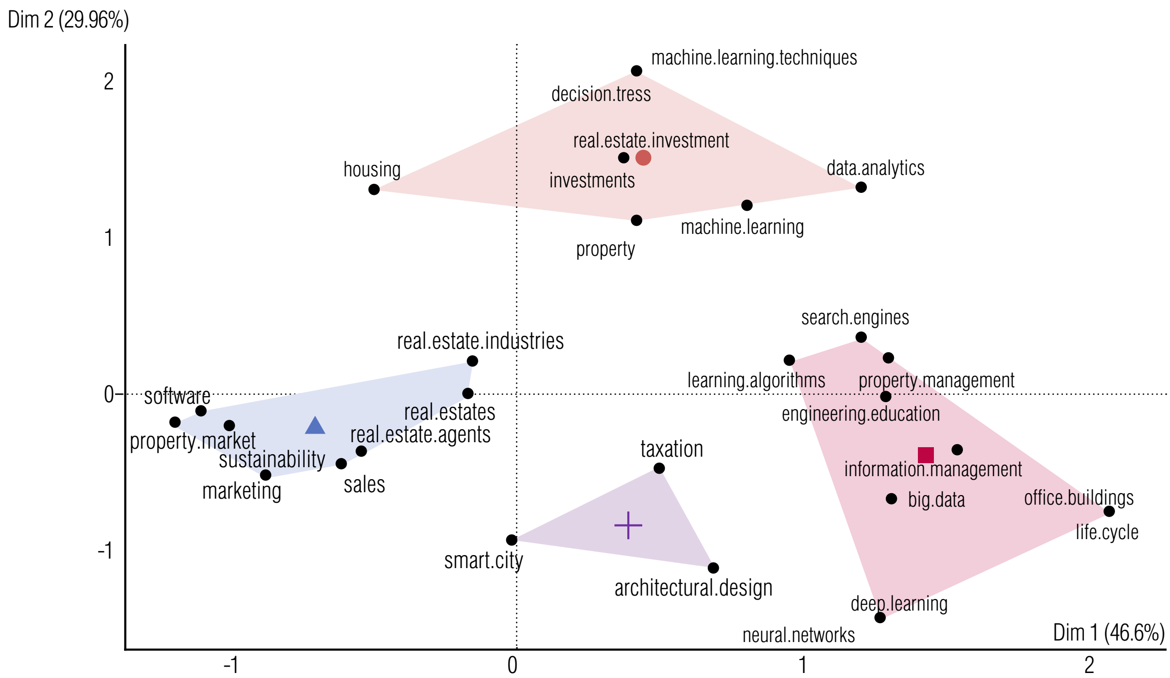

Fig. 5. Conceptual structure map using multiple correspondence analysis (MCA) method.

The keywords "learning algorithms" and "engineering education" formed the only cluster in the quadrant of basic themes, indicating their importance for the research field of smart real estate but not yet fully formed.

By comparing *Fig. 5* and *Fig. 6*, we found that most of the keywords in Categories I and III are either motor themes with a higher degree of development and relevance or basic themes with a higher degree of relevance but have not been explored in depth. Category III overlaps with all keywords of the three clusters located in the quadrant of emerging or declining themes; keywords of Category IV, "smart city" and "taxation," are located in the niche themes quadrant, and "architectural design" is located in the motor themes quadrant.

Conclusion

Innovative technology is important and influential, and the development of technology has driven the transformation of the industry. As pointed out by [37], the driving force of transformation is strategy, not technology, and it applies to the real estate industry. This study provides a systematic review of the digital transformation of the real estate industry, focusing on the aspect of smart real estate. It provides a comprehensive understanding of current trends in theoretical development, taking "smart real estate" as a research field, and it guides academic scholars with future research directions and industry practitioners with strategy or policy-making. The key findings are summarized as follows:

i. Research on smart real estate is a relatively new research field. The relevant literature first appeared in 2012 and has shown a rapid growth trend since 2018.

ii. The University of New South Wales in Australia made the most contributions to this field of study, followed by the Near East University in Turkey and the University of Reading in Malaysia. There appeared to be significant gaps in this research field in other developed regions such as the European Union as well as in emerging economies such as China and Russia.

iii. Research in the field of smart real estate exhibits a strong co-authorship characteristic, with the most prominent

Table 5.

Categories of keywords and corresponding documents with the highest contribution

Keywords distribution per category	No. of keywords	Time of appearance	Most contributed documents
Category I			
machine learning techniques, decision trees, data analytics, machine learning, property, investments, real estate investment, housing	8	2020	[26, 30]
Category II			
real estate, real estate industries, real estate agents, sales, marketing, sustainability, property market, software	8	2018–2022	[1, 17, 20, 31]
Category III			
search engines, learning algorithms, property management, engineering education, information management, big data, office buildings, life cycle, deep learning, neural networks	10	2020	[18]
Category IV			
taxation, smart city, architectural design	3	2018–2019	[33, 34]

Note:
The most contributed documents were identified by the factorial analysis in the Biblioshiny app, which presents the most weighted documents in influencing the corresponding research category [11].

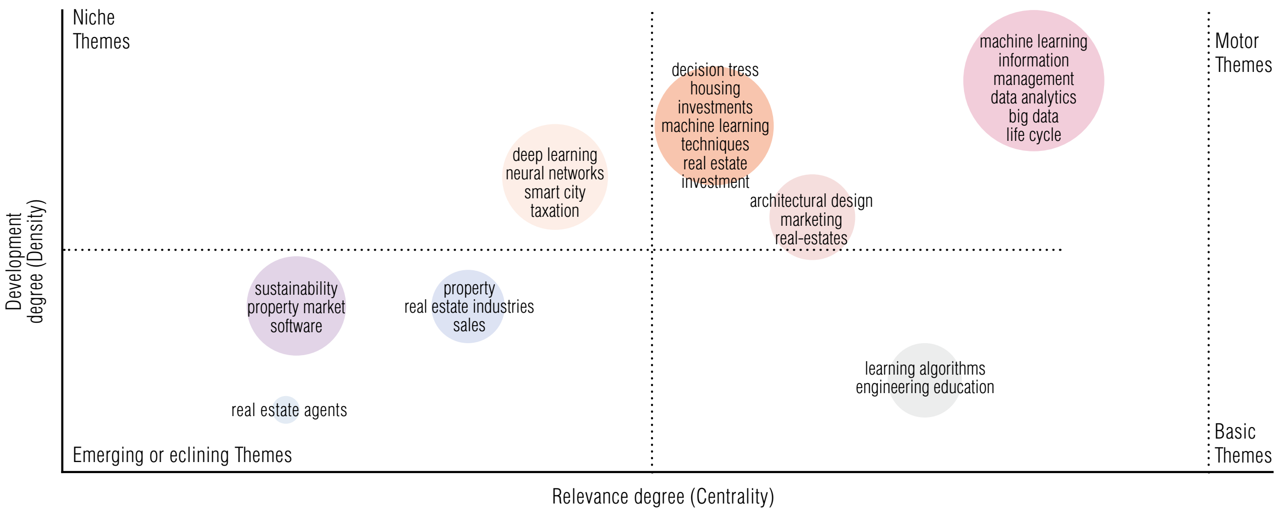

Fig. 6. Thematic map by Keywords Plus.

contributions coming from a collaborative network of Ullah, Sepasgozar, and Shirowzhan et al., including the total number of publications and citations.

iv. Through two different bibliometric mapping methods that categorized and clustered themes using the Keywords Plus, we found that keywords from two categories, such as "machine learning techniques," "decision trees" and "data analytics" from Category I, and "information management," "big data" and "life cycle" from Category III, have a higher degree of thematic development and relevance.

The study employing the bibliometric analysis method provides the most objective results for a systematic review, which avoids the bias of individual subjective factors. Nevertheless, there are two major limitations that need to be acknowledged. First, the digital transformation of real estate is a very broad concept and involves a number of related research fields such as smart city, property technology (ProTech), digital real estate, smart housing and smart homes, etc.; thus it is highly recommended for future researchers to compare and summarize all related research concepts or frameworks. Second, Scopus and Web of Science were limited as sources of scientific documents; other bibliographic databases, such as Google Scholar and ProQuest, should be researched in the future for the development of bibliometric studies. ■

References

1. Ullah F., Sepasgozar S.M.E., Wang C. (2018) A systematic review of smart real estate technology: Drivers of, and barriers to, the use of digital disruptive technologies and online platforms. *Sustainability*, vol. 10, no. 9, 3142. https://doi.org/10.3390/su10093142
2. Kassner A.J., Cajias M., Zhu B. (2022) The PropTech investors' dilemma – What are the key success factors that secure survival? *Journal of Property Investment & Finance*, vol. 41, no. 1, pp. 76–91. https://doi.org/10.1108/JPIF-01-2022-0007
3. Vassileva B. (2017) Marketing 4.0: How technologies transform marketing organization. *Óbuda University e-Bulletin*, vol. 7, no. 1, pp. 47–56.
4. Shirowzhan S., Tan W., Sepasgozar S.M. (2020) Digital twin and cyberGIS for improving connectivity and measuring the impact of infrastructure construction planning in smart cities. *ISPRS International Journal of Geo-Information*, vol. 9, no. 4, 240. https://doi.org/10.3390/ijgi9040240
5. Starr C.W., Saginor J., Worzala E. (2020) The rise of PropTech: Emerging industrial technologies and their impact on real estate. *Journal of Property Investment & Finance*, vol. 39, no. 2, pp. 157–169. https://doi.org/10.1108/JPIF-08-2020-0090
6. Dash G., Kiefer K., Paul J. (2021) Marketing-to-Millennials: Marketing 4.0, customer satisfaction and purchase intention. *Journal of Business Research*, vol. 122, pp. 608–620. https://doi.org/10.1016/j.jbusres.2020.10.016
7. Page M.J., McKenzie J.E., Bossuyt P.M. et al. (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *The BMJ*, vol. 372(n71). https://doi.org/10.1136/bmj.n71
8. Higgins J.P., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., Welch V.A. (editors) (2019) *Cochrane handbook for systematic reviews of interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons.
9. Archer L.B. (1984) Systematic method for designers. *Developments in design methodology* (ed. N. Cross), pp. 57–82. New York: John Wiley & Sons.
10. Peffers K., Tuunanen T., Rothenberger M.A., et al. (2007) A design science research methodology for information systems research. *Journal of management information systems*, vol. 24(3), pp. 45–77.
11. Aria M., Cuccurullo C. (2017) Bibliometrix: An R-Tool for comprehensive science mapping analysis. *Journal of Informetrics*, vol. 11, no. 4, pp. 959–975. https://doi.org/10.1016/j.joi.2017.08.007

12. Zhu J., Liu W. (2020) A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*, vol. 123, no. 1, pp. 321–335. https://doi.org/10.1007/s11192-020-03387-8
13. Linnenluecke M.K., Marrone M., Singh A.K. (2020) Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management*, vol. 45, no. 2, pp. 175–194. https://doi.org/10.1177/0312896219877678
14. Demaine J. (2022) Fractionalization of research impact reveals global trends in university collaboration. *Scientometrics*, vol. 127, no. 5, pp. 2235–2247. https://doi.org/10.1007/s11192-021-04246-w
15. Anand A., Kringelum L.B., Madsen C.Ø., Selivanovskikh L. (2021) Interorganizational learning: A bibliometric review and research agenda. *The Learning Organization*, vol. 28, no. 2, pp. 111–136. https://doi.org/10.1108/TLO-02-2020-0023
16. Lecomte P. (2022) On the economic nature of behavioural control in smart real estate. *Journal of General Management*. https://doi.org/10.1177/03063070221131980
17. Low S., Ullah F., Shirowzhan S., Sepasgozar S.M., Lin Lee C. (2020) Smart digital marketing capabilities for sustainable property development: A case of Malaysia. *Sustainability*, vol. 12, no. 13, 5402. https://doi.org/10.3390/su12135402
18. Munawar H.S., Qayyum S., Ullah F., Sepasgozar S. (2020) Big data and its applications in smart real estate and the disaster management life cycle: A systematic analysis. *Big Data and Cognitive Computing*, vol. 4, no. 2, 4. https://doi.org/10.3390/bdcc4020004
19. Ullah F., Al-Turjman F. (2023) A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities. *Neural Computing and Applications*, vol. 35, pp. 5033–5054. https://doi.org/10.1007/s00521-021-05800-6
20. Ullah F., Sepasgozar S.M.E. (2020) Key factors influencing purchase or rent decisions in smart real estate investments: A system dynamics approach using online forum thread data. *Sustainability*, vol. 12, no. 11, 4382. https://doi.org/10.3390/su12114382
21. Ullah F., Al-Turjman F., Qayyum S., Inam H., Imran M. (2021) Advertising through UAVs: Optimized path system for delivering smart real-estate advertisement materials. *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3429–3463. https://doi.org/10.1002/int.22422
22. Ullah F., Sepasgozar S.M.E., Thaheem M.J., Al-Turjman F. (2021) Barriers to the digitalisation and innovation of Australian Smart Real Estate: A managerial perspective on the technology non-adoption. *Environmental Technology and Innovation*, vol. 22, article 101527. https://doi.org/10.1016/j.eti.2021.101527
23. Ullah F., Sepasgozar S.M.E., Shirowzhan S., Davis S. (2021) Modelling users' perception of the online real estate platforms in a digitally disruptive environment: An integrated KANO-SISQual approach. *Telematics and Informatics*, vol. 63, article 101660. https://doi.org/10.1016/j.tele.2021.101660
24. Lecomte P. (2019) New boundaries: Conceptual framework for the analysis of commercial real estate in smart cities. *Journal of Property Investment and Finance*, vol. 37, no. 1, pp. 118–135. https://doi.org/10.1108/JPIF-10-2018-0083
25. Lecomte P. (2020) iSpace: principles for a phenomenology of space user in smart real estate. *Journal of Property Investment and Finance*, vol. 38, no. 4, pp. 271–290. https://doi.org/10.1108/JPIF-07-2019-0091
26. Ahmed H., Jilani T.A., Haider W., Hasany S.N., Abbasi M.A., Masroor A. (2020) Producing standard rules for smart real estate property buying decisions based on web scraping technology and machine learning techniques. *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 489–505. https://doi.org/10.14569/IJACSA.2020.0110363
27. Allameh E., Jozam M.H., de Vries B., Timmermans H., Beetz J., Mozaffar F. (2012) The role of smart home in smart real estate. *Journal of European Real Estate Research*, vol. 5, no. 2, pp. 156–170. https://doi.org/10.1108/17539261211250726

28. Kempeneer S., Peeters M., Compernolle T. (2021) Bringing the user back in the building: An analysis of ESG in real estate and a behavioral framework to guide future research. *Sustainability*, vol. 13, no. 6, 3239. https://doi.org/10.3390/su13063239

29. Azmi A., Ibrahim R., Abdul Ghafar M., Rashidi A. (2022) Smarter real estate marketing using virtual reality to influence potential homebuyers' emotions and purchase intention. *Smart and Sustainable Built Environment*, vol. 11, no. 4, pp. 870–890. https://doi.org/10.1108/SASBE-03-2021-0056

30. Sandeep Kumar E., Talasila V. (2020) A combined data analytics and network science approach for smart real estate investment: Towards affordable housing. *Smart Governance for Cities: Perspectives and Experiences. EAI/Springer Innovations in Communication and Computing* (ed. N. Lopes). Springer, Cham. https://doi.org/10.1007/978-3-030-22070-9_8

31. Hapuarachchi H.A.V.P.U., Manoratne M.D., Gamlath K.G.B.K., Vithane S.G., Sriyaratna D., Supunya N.R. (2022) Realty scout – smart system for real estate analysis & forecasting with interactive user interface. Proceedings of the *2022 IEEE 7th International Conference for Convergence in Technology (I2CT 2022), Pune, India, 7–9 April 2022*, pp. 56–61.

32. Aydinoglu A.C., Bovkir R. (2020) Developing a mobile application for smart real estate information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIV-4/W3-2020, pp. 89–94. https://doi.org/10.5194/isprs-archives-XLIV-4-W3-2020-89-2020

33. Xu H., Gade A. (2017) Smart real estate assessments using structured deep neural networks. Proceedings of the *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, San Francisco, CA, USA, 04–08 August 2017*, pp. 1–7. https://doi.org/10.1109/UIC-ATC.2017.8397560.

34. Su T., Li L.H. (2019) BIM - Based machine learning engine for smart real estate appraisal. Proceedings of the *2nd International Conference on Sustainable Smart Manufacturing (S2M 2019), Manchester, United Kingdom, 9–11 April 2019*, pp. 63–68.

35. Zhang J., Yu Q., Zheng F., Long C., Lu Z., Duan Z. (2016) Comparing keywords plus of WoS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology*, vol. 67, no. 4, pp. 967–972. https://doi.org/10.1002/asi.23437

36. Cobo M.J., López-Herrera A.G., Herrera-Viedma E., Herrera F. (2011) An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, vol. 5, no. 1, pp. 146–166. https://doi.org/10.1016/j.joi.2010.10.002

37. Kane G.C., Palmer D., Phillips A.N., Kiron D., Buckley N. (2015) *Strategy, not technology, drives digital transformation*. MIT Sloan Management Review and Deloitte University Press.

About the author

Fengchen Wang

Graduate student, St. Petersburg School of Economics and Management, HSE University – Campus in St. Petersburg, 3A-1, Kantemirovskaya Street, St. Petersburg 194100, Russia;

E-mail: fwang@hse.ru

ORCID: 0000-0003-3103-5049