

DOI: 10.21821/2309-5180-2020-12-5-877-883

METHOD OF COMPARATIVE IDENTIFICATION OF MARINE OBJECTS TAXONOMY

A. S. Sobolev, V. A. Dorovskoy, N. P. Smetyh

Kerch State Maritime University, Kerch, Russian Federation

Information organizational systems that have been successfully used in the various areas of economy for more than twenty years are now divided into two main groups: integrated and highly specialized systems. The class of highly specialized systems includes information systems for banking automation, statistics, financial and accounting (for example, IC, FinExpert, SoNet), marketing, and investment management (for example, Project Expert). Integrated organizational systems include corporate information systems, which today successfully displace traditional automated enterprise management systems in the field of production. The purpose of the research is to improve the efficiency and quality of intellectual processing of poorly formalized information in the organizational information systems based on the implementation of information and logical models and methods of marine objects; the object of research is information processing processes in the organizational information systems. The subject of the research is information technologies for intellectual processing of poorly formalized information used in the organizational systems based on the algebraic-logical models. Research methods are based on the complex use of the theory of intelligence, category theory, algebra of finite predicates and the method of comparative identification for the development of information and logical models and methods of knowledge identification. Predicate algebra and category theory are used to formalize knowledge, describe natural language relations, and model the Manager's intellectual activity in documents processing tasks. The comparator identification method is used to describe the intellectual functions of a corporate information system user. Thus, the method of comparative identification allows you to automatically divide the texts of documents in full-text databases into identical ones (in relation to the areas of knowledge of the Manager of a marine object), and divide information semantic units expressed by CS, rubrics, and UDC classes into equivalence classes that define top-rank taxa, allowing you to automate the process of developing a corporate taxonomy.

Keywords: comparative identification, taxonomy, marine objects, intelligent functions, corporate information system, marine object, marine manager.

For citation:

Sobolev, Alexandr S., Vladimir A. Dorovskoy, and Nadezhda P. Smetyh. "Method of comparative identification of marine objects taxonomy." *Vestnik Gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala S. O. Makarova* 12.5 (2020):877–883. DOI: 10.21821/2309-5180-2020-12-5-877-883.

УДК 681.5:004

МЕТОД КОМПАРАТОРНОЙ ИДЕНТИФИКАЦИИ ТАКСОНОМИИ МОРСКИХ ОБЪЕКТОВ

А. С. Соболев, В. А. Доровской, Н. П. Сметюх

ФГБОУ ВО «Керченский государственный морской технологический университет»,
Керчь, Российская Федерация

Отмечается, что информационные организационные системы, успешно используемые в различных областях больше двадцати лет, делятся на две основные группы: интегрированные и узкоспециализированные системы. К классу узкоспециализированных систем можно отнести информационные системы для автоматизации банковской деятельности, в статистике, для финансового и бухгалтерского учета (например, IC, FinExpert, SoNet), в маркетинге, в инвестиционном менеджменте (например, Project Expert). К интегрированным организационным системам относятся корпоративные информационные системы, которые сегодня успешно вытесняют в сфере производства традиционные автоматизированные системы управления предприятием. Целью исследований является повышение эффективности и качества интеллектуальной обработки слабо формализованной информации в организационных информационных системах на основе реализации информационно-логических моделей и методов морских

объектов. Объектом исследования являются процессы обработки информации в организационных информационных системах. Предметом исследования являются информационные технологии интеллектуальной обработки слабо формализованной информации, применяемые в организационных системах, основанные на алгебро-логических моделях. Рассматриваемые в работе методы исследования базируются на комплексном использовании теории интеллекта, теории категорий, алгебры конечных предикатов и метода компараторной идентификации для разработки информационно-логических моделей и методов идентификации знаний. Алгебра предикатов и теория категорий используется для формализации знаний, описания естественно-языковых отношений и моделирования интеллектуальной деятельности менеджера применительно к задачам обработки документов. Рассмотрен метод компараторной идентификации, используемый для описания интеллектуальных функций пользователя корпоративной информационной системы. Таким образом, метод компараторной идентификации позволяет автоматически разделять тексты документов полнотекстовых баз данных на тождественные (по отношению к областям знаний менеджера морского объекта) и разбивать информационные смысловые единицы, выражаемые компараторной сетью, рубриками и классами УДК, на классы эквивалентностей, определяющие таксоны верхнего ранга, позволяющие автоматизировать процесс разработки корпоративной таксономии.

Ключевые слова: компараторная идентификация, таксономия, морские объекты, интеллектуальные функции, корпоративная информационная система, менеджер морского объекта.

Для цитирования:

Соболев А. С. Метод компараторной идентификации таксономии морских объектов / А. С. Соболев, В. А. Доровской, Н. П. Сметюх // Вестник Государственного университета морского и речного флота имени адмирала С. О. Макарова. — 2020. — Т. 12. — № 5. — С. 877–883. DOI: 10.21821/2309-5180-2020-12-5-877-883.

Введение (Introduction)

Актуальность исследований основана на применении полученных методов разработки систем общения с ЭВМ на языке, понятном компьютеру и менеджеру морского объекта. Повышение эффективности вывода в автоматизированных информационных морских системах морской информации обусловлено непрерывно увеличивающимся объемом различных данных во всех отраслях человеческой деятельности и соответствующими потребностями менеджеров оперативно, полно и качественно получать документные и электронные первоисточники. С возрастанием объемов накопленной информации возникает задача представления знаний и интеллектуального анализа данных с целью предоставления пользователю возможности правильно ориентироваться в среде электронных ресурсов большого объема, а морскому эксперту (сотруднику организации) — возможности автоматизировать некоторые участки своей работы.

Современная морская организационная система — это рациональное объединение менеджеров морских объектов и привлечение ресурсов в обособленном целенаправленном коллективе, обеспечивающем распределение обязанностей, координацию действий морских менеджеров для достижения поставленной цели и их ориентацию на конечный результат.

В рамках данного исследования под *морской системой* понимается такая организационная форма деятельности, при которой ее структура предполагает наличие множества территориально распределенных объектов, взаимодействующих между собой в процессе деятельности и обменивающихся большими объемами информации.

Методы и материалы (Methods and Materials)

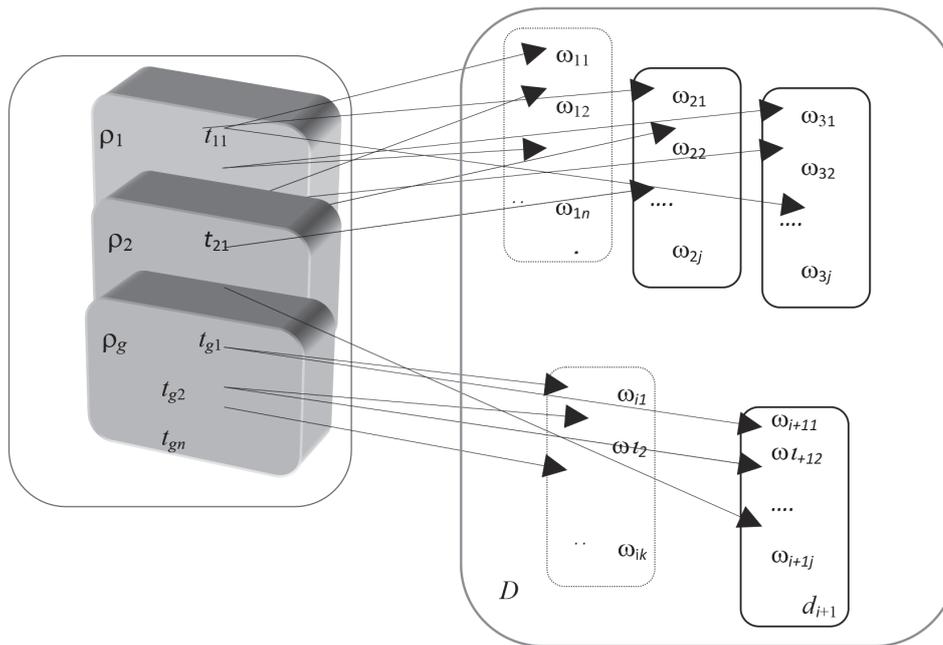
Методы исследования базируются на комплексном использовании теории интеллекта, теории категорий, алгебры конечных предикатов и метода компараторной идентификации для разработки информационно-логических моделей и методов идентификации знаний. Алгебра предикатов и теория категорий используются для формализации знаний в задачах обработки документов. Метод компараторной идентификации используется для описания интеллектуальных функций пользователя корпоративной информационной системы.

Результаты (Results)

В процессе интеллектуальной обработки документов и смысловых единиц компаратор реализует бинарный предикат персонификации интеллектуального ресурса морской компании $P(t, r)$ (). Полагаем, что выбор значения предиката $\varepsilon = P(t, r)$, $\varepsilon = \{0, 1\}$ полностью определяется смыслом, раскрываемым в документе: $\omega = g(d)$, и концептом $\rho = f(t)$, соответствующим КС (компараторной сети), рубрике или классу УДК — t . В результате понимания смысла документа и значения понятия определяется соответствие ($\varepsilon = 1$) или несоответствие ($\varepsilon = 0$) смысла, раскрываемого в морском документе ω , понятию ρ . Применяем предикат аналитико-синтетической обработки документа [1]–[4] $\varepsilon = Z(\tau, \rho)$, реализующий функцию понимания менеджером морского объекта смысла текста и значения смысловых единиц. Предикаты P и Z , функции f и g , переменные d, t, ω, ρ связаны зависимостью

$$P(d, t) = Z(g(d), f(t)) = Z(\omega, \rho). \quad (1)$$

Установление конкретного вида предиката $Z(\omega, \rho)$ позволяет группировать смысловые единицы по классам *условной* и *безусловной эквивалентности*. Под классификацией смысловых единиц в данном случае понимаем процесс распределения концептов, определяемых данными единицами, на взаимно исключающие таксоны¹, представляющие множество объектов, имеющих общие содержательные признаки. Корпоративная таксономия, приведенная на логической схеме, строится таким образом, чтобы в получившейся иерархической системе каждый таксон занимал определенное относительно других классов точно фиксированное место.



Логическая схема формирования корпоративной таксономии

Основной принцип деления смысловых единиц основан на лексической синонимии (безусловной эквивалентности) и семантической родственности слов. Использование равенства (1) позволяет перейти от субъективного восприятия смыслов к объективному отношению между документами и смысловыми единицами (представленными ключевыми словами и словосочетаниями, классами УДК и рубриками), соответствующими $P(d, t) = 1$ или $P(d, t) = 0$, несоответствующей тексту документа. Доказано, что введенные на множествах $D * D$ и $T^k * T^k$ предикаты документов и смысловых единиц соответствуют персонифицированному интеллектуальному ресурсу k -го менеджера морского объекта:

¹ Таксон — группа объектов, связанных степенью общности свойств и признаков.

$$G_1(d_1, d_2) = \forall t^k \in T^k (P(d_1, t^k) \sim P(d_2, t^k)); \quad (2)$$

$$G_2(t_1^k, t_2^k) = \forall d \in D (P(d, t_1^k) \sim P(d, t_2^k)). \quad (3)$$

Эти группы объектов являются предикатами эквивалентности, так как удовлетворяют условиям рефлексивности, транзитивности и симметричности и однозначно определяются предикатом персонифицированного интеллектуального ресурса морской компании $P(d, t^k)$. Предикат, устанавливающий разбиение ψ_1 , — G_1 , выполняет также разбиение множества D на слои текстов документов. Все документы, принадлежащие одному слою разбиения, относятся к одной области знаний менеджера морского объекта. Любые документы, взятые из разных слоев разбиения ψ_1 , содержат информацию, относящуюся к различным областям знаний менеджеров морского объекта. Разбиение ψ_2 множества T на слои смысловых единиц осуществляет предикат G_2 который определяет при разбиении КС классы классификаторов, характеризуемые определенной степенью общности. Все смысловые единицы, принадлежащие одному слою разбиения, относятся к одному таксону. Вместе с тем любые две смысловые единицы, взятые из разных слоев разбиения ψ_2 , относятся к различным таксонам.

Функция $f(t)$ связывает КС, классы-классификаторы или рубрику t , отображаемые лексическими единицами естественного языка, с тем понятием ρ , которое они выражают. Т. е. понимая смысловую единицу, менеджер компании осознает обозначаемое ею и объем [3]–[6]. Тогда предикат эквивалентности G_1 на декартовом произведении $D^* D$ можно представить в виде

$$G_1(d_1, d_2) = T_1(g(d_1), g(d_2)), \quad (4)$$

где g — сюръективная функция, представляющая каноническое отображение множества D на разбиение ψ_1 ; T_1 — предикат равенства на ψ_1 .

Документы d_1 и d_2 будут тождественны, если им соответствует один и тот же образ: $g(d_1) = g(d_2)$. Таким образом, осуществляется моделирование процесса понимания смысла текстов документов, тождественных относительно одной области знаний менеджера, с помощью компараторной идентификации предиката G_1 путем сравнения значений функции понимания текстов документов g . В данной модели компаратор в текущий момент времени анализирует (сравнивает) ощущения в понимании смысла двух текстов документов, принимая на этой основе решение об их эквивалентности или неэквивалентности. Реакция компаратора может принимать только два булевых значения: $\{0, 1\}$.

Предикат эквивалентности G_2 на декартовом произведении $T^k * T^k$ можно представить в виде

$$G_2(t_1^k, t_2^k) = D_1(f(t_1), f(t_2)), \quad (5)$$

где f — сюръективная функция, представляющая каноническое отображение множества T на разбиение ψ_2 ; D_1 — предикат равенства на декартовом произведении $\psi_2 * \psi_2$.

Терминологические понятия t_1^k и t_2^k называются *тождественными*, если $f(t_1) = f(t_2)$. Формула (4) моделирует процесс распознавания КС, рубрик и классов УДК, тождественных относительно одной области знаний менеджера. Осуществляя компарацию предиката G_2 путем сравнения значений функции понимания терминологического понятия f , компаратор анализирует (сравнивает) понимание смысла двух смысловых единиц, принимая решение об их эквивалентности или неэквивалентности [7]–[9]. Если два терминологических понятия эквивалентны относительно одной области знаний менеджера, то реакция компаратора соответствует значению 1 предиката G_2 , и значению 0 — в противоположном случае.

Можно выразить предикат аналитико-синтетической обработки документов, поступающих на вход КИС Z (1) через предикат персонификации интеллектуального ресурса компании и функции понимания документа и терминологического понятия:

$$Z(\omega, \rho) = \exists d \in D \exists t \in T (P(d, t) g(d, \omega) f(t, \rho) = P(g^{-1}(\omega), f^{-1}(\rho))), \quad (6)$$

где $g^{-1}(\omega)$ — один из элементов $d \in D$, удовлетворяющих условию $\omega = g(d)$, а $f^{-1}(\rho)$ обозначает один из элементов $t \in T$, удовлетворяющих условию $\rho = f(t)$.

Из равенства (1) можно определить

$$P(d, t) = \exists \omega \in D' \exists \rho \in T' (Z(\omega, \rho) g(d, \omega) f(t, \rho)). \quad (7)$$

Будем считать смысловые единицы t_1 и t_2 условно эквивалентными, если они соответствуют концептам $\rho_1 = f(t_1)$ и $\rho_2 = f(t_2)$, относящимся к одному дескриптору, раскрывающему смысл текста одного документа: $\omega_1 \in d$ и $\omega_2 \in d$.

Значениями переменных ω и ρ служат не сами слои разбиения ψ_1 и ψ_2 , а их имена D' и T' . Следует отличать множества слоев разбиения ψ_1 и ψ_2 от множеств имен этих слоев D' и T' , подтемы, отражаемые документами, смысл, понимание которого соответствует области знаний менеджера морского объекта, а также подрубрики и классы, объединяющие смысловые единицы и таксоны, так называемые *данные объединения*.

Имена слоев разбиения можно выбрать произвольным образом, но при этом каждому классу разбиения должно соответствовать в точности одно имя. Каждая область знаний менеджера морского объекта из множества D' характеризуется как слой разбиений ϑ_c и ψ_1 множества D , состоящего из всевозможных текстов документов $d' \in D$, тождественных по смыслу или по осмыслению ситуации тексту документа d . О тексте документа d известно, что он освещает смысл ω , $\omega = g(d)$. Множество D' всех областей знаний, рассматриваемых с помощью данной корпоративной информации, характеризует разбиение ψ_1 [9]–[13]. В свою очередь, понятие $\rho = f(t)$ из множества T' характеризуется как слой Λ_b разбиения ψ_2 , состоящий из ключевых слов, классов УДК и рубрик $t' \in T$, тождественных смысловой единице t .

Предикат g связывает предметную переменную ω , служащую именами слоев разбиения ψ_1 , и переменную d , обозначающую текст документа из множества D . Разбиение ψ_1 выступает в роли множества D' всех областей знаний менеджера, определенных осмыслением текстов документов, взятых из множества D .

Предикат $f(t, \rho)$ отображает связь между слоями разбиения смысловых единиц ψ_2 . В роли множества T' выступает совокупность имен всех слоев разбиения ψ_2 , представляющих собой множество таксонов. Имена таксонов следующего уровня таксономии могут выбираться менеджером вручную, если разбиение на области знаний осуществляется в диалоговом режиме, либо в автоматическом режиме. Имя слоя разбиения (имя таксона), который объединяет смысловые единицы, представляющие таксоны нижнего ранга, может определяться по наиболее общему понятию, входящему в данный слой разбиения.

Критерием выбора имени таксона высшего ранга из класса эквивалентных смысловых единиц может считаться полнота выражения смыслового значения данного класса, формально выраженная его весовой нагруженностью и информационной значимостью.

Выводы (Summary)

На основании проведенных исследований сделан вывод о том, что метод компараторной идентификации может разделять тексты документов полнотекстовых баз данных на тождественные. Кроме того, использование данного метода позволяет разбивать информационные смысловые единицы, выражаемые компьютерными системами, рубриками и классами в виде универсального десятичного кода, на классы эквивалентностей. Это разбиение определяет таксоны верхнего ранга, позволяя автоматизировать процесс разработки корпоративной таксономии.

СПИСОК ЛИТЕРАТУРЫ

1. Бондаренко М. Ф. Модели языка / М. Ф. Бондаренко, В. А. Чикина, Ю. П. Шабанов-Кушнаренко // Бионика интеллекта. — 2004. — № 61. — С. 27–37.
2. Бондаренко М. Ф. Об общей теории компараторной идентификации / М. Ф. Бондаренко, С. Ю. Шабанов-Кушнаренко, Ю. П. Шабанов-Кушнаренко // Бионика интеллекта. — 2008. — № 2(69). — С. 13–22.
3. Петров Э. Г. Компараторная структурно-параметрическая идентификация моделей скалярного многофакторного оценивания: моногр. / Э. Г. Петров, В. В. Крючковский. — Херсон: Олди-плюс, 2009. — 294 с.

4. Хайрова Н. Ф. Использование элементов метаданных OPENDOCUMENT FORMAT для решения задачи структурирования и извлечения информации / Н. Ф. Хайрова // Вестник Херсонского государственного технического университета. — 2009. — № 6(29). — С. 42–45.
5. Шабанов-Кушнарченко Ю. П. Теория интеллекта: математические средства / Ю. П. Шабанов-Кушнарченко. — Х.: Вища шк., 1984. — 143 с.
6. Шабанов-Кушнарченко Ю. П. Компараторная идентификация лингвистических объектов: монография / Ю. П. Шабанов-Кушнарченко, Н. В. Шаронова. — К.: ИСДО, 1993. — 116 с.
7. Шаронова Н. В. Модель извлечения глубинных знаний для систем организационного управления / Н. В. Шаронова, В. А. Тарловский, Н. Ф. Хайрова // Вестник Херсонского национального технического университета. — 2010. — № 2 (38). — С. 97–102.
8. Von Bertalanffy L. An Outline of General System Theory / L. Von Bertalanffy // British Journal for the Philosophy of Science. — 1950. — Vol. 1. — P. 134–165. DOI: 10.1093/bjps/I.2.134.
9. Bar-Yossef Z. Efficient search engine measurements / Z. Bar-Yossef, M. Gurevich // Proceedings of the 16th international conference on World Wide Web. — 2007. — С. 401–410. DOI: 10.1145/1242572.1242627.
10. Hammer M. Reengineering the Corporation. A Manifesto for Business Revolutions / M. Hammer, J. Champy. — 1st Edition. — New York: HarperBusiness, 1993. — 240 p.
11. Черный С. Г. Увеличение степени отказоустойчивости в программно-аппаратных системах сетевого управления на примере мягкого облачного хранилища / С. Г. Черный, А. А. Жиленков // Научно-техническая информация. Серия 2: Информационные процессы и системы. — 2020. — № 1. — С. 28–35. DOI: 10.36535/0548-0027-2020-01-4.
12. Биденко С. И. Информационная технология компенсаторной идентификации пространственных данных от оптических систем наблюдения беспилотных аппаратов / С. И. Биденко, Д. А. Елизаров, С. Г. Черный, В. М. Шестаков // Информация и космос. — 2020. — № 1. — С. 120–126.
13. Жиленков А. А. Повышение степени отказоустойчивости в сложных программно-аппаратных системах сетевого управления / А. А. Жиленков, С. Г. Черный // Датчики и системы. — 2019. — № 12 (242). — С. 11–17. DOI: 10.25728/datsys.2019.12.2.

REFERENCES

1. Bondarenko, M. F., V. A. Chikina, and Ju. P. Shabanov-Kushnarenko. “Modeli jazyka.” *Bionika intellekta* 61 (2004): 27–37.
2. Bondarenko, M. F., S. Ju. Shabanov-Kushnarenko, and Ju.P. Shabanov-Kushnarenko. “Ob obshhej teorii komparatornoj identifikacii.” *Bionika intellekta* 2(69) (2008): 13–22.
3. Petrov, Je. G., and V. V. Krjuchkovskij. *Komparatornaja strukturno-parametricheskaja identifikacija modelej skaljarnogo mnogofaktornogo ocenivaniya: Monografija*. Herson: Oldi-pljus, 2009.
4. Hajrova, N. F. “Ispol’zovanie jelementov metadannyh OPENDOCUMENT FORMAT dlja reshenija zadachi strukturirovaniya i izvlechenija informacii.” *Vestn. Herson. gos. tehn. un-ta* 6(29) (2009): 42–45.
5. Shabanov-Kushnarenko, Ju. P. *Teorija intellekta: matematicheskie sredstva*. H.: Vishha shk., 1984.
6. Shabanov-Kushnarenko, Ju. P., and N. V. Sharonova. *Komparatornaja identifikacija lingvisticheskikh ob#ektov: monografija*. K.: ISDO, 1993.
7. Sharonova, N. V., V. A. Tarlovskij, and N. F. Hajrova. “Model’ izvlechenija glubinyh znaniy dlja sistem organizacionnogo upravlenija.” *Vestnik Hersonskogo nacional’nogo tehničeskogo universiteta* 2(38) (2010): 97–102.
8. Von Bertalanffy, Ludwig. “An Outline of General System Theory.” *British Journal for the Philosophy of Science* 1 (1950): 134–165. DOI: 10.1093/bjps/I.2.134.
9. Bar-Yossef, Ziv, and Maxim Gurevich. “Efficient search engine measurements.” *Proceedings of the 16th international conference on World Wide Web*. 2007. 401–410. DOI: 10.1145/1242572.1242627.
10. Hammer, Michael, and James Champy. *Reengineering the Corporation. A Manifesto for Business Revolutions*. 1st edition. New York: HarperBusiness, 1993.
11. Chernyi, S. G., and A. A. Zhilenkov. “Enhanced fault tolerance in software and hardware network control systems as the case of soft cloud storage.” *Automatic Documentation and Mathematical Linguistics* 1 (2020): 28–35.
12. Bidenko, Sergey Ivanovich, Dmitry Aleksandrovich Elizarov, Sergey Grigoryevich Cherny, and Vasily Mikhaylovich Shestakov. “Information technology for compensatory identification of spatial data from optical surveillance systems of unmanned vehicles.” *Information and Space* 1 (2020): 120–126.
13. Zhilenkov, Anton A., and Sergei G. Chernyi. “Increasing fault tolerance in complex hardware-software network control systems.” *Sensors & Systems* 12(242) (2019): 11–17. DOI: 10.25728/datsys.2019.12.2.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Соболев Александр Сергеевич —
инженер
ФГБОУ ВО «КГМТУ»
298309, Российская Федерация, Керчь,
ул. Орджоникидзе, 82
e-mail: sobolev.alexandr1496@gmail.com
Доровской Владимир Алексеевич —
доктор технических наук, профессор
ФГБОУ ВО «КГМТУ»
298309, Российская Федерация, Керчь,
ул. Орджоникидзе, 82
e-mail: dora1943@mail.ru
Сметюх Надежда Павловна —
кандидат технических наук
ФГБОУ ВО «КГМТУ»
298309, Российская Федерация, Керчь,
ул. Орджоникидзе, 82
e-mail: nadya.s.2011@yandex.ru

INFORMATION ABOUT THE AUTHORS

Sobolev, Alexandr S. —
Engineer
Kerch State Maritime University
82 Ordzhonikidze Str., Kerch, 298309,
Russian Federation
e-mail: sobolev.alexandr1496@gmail.com
Dorovskoy, Vladimir A. —
Dr. of Technical Sciences, professor
Kerch State Maritime University
82 Ordzhonikidze Str., Kerch, 298309,
Russian Federation
e-mail: dora1943@mail.ru
Smetyh, Nadezhda P. —
PhD
Kerch State Maritime University
82 Ordzhonikidze Str., Kerch, 298309,
Russian Federation
e-mail: nadya.s.2011@yandex.ru

*Статья поступила в редакцию 6 октября 2020 г.
Received: October 6, 2020.*