Article

Not peer-reviewed version

The Impact of the Human Factor on Communication During a Collision Situation in Maritime Navigation

Leszek Misztal * and Paulina Hatlas-Sowinska

Posted Date: 3 February 2025

doi: 10.20944/preprints202502.0016.v1

Keywords: safety; metaontology; automation of communication; data mining; natural language processing; human factor

Article

The Impact of the Human Factor on Communication During a Collision Situation in Maritime Navigation

Leszek Misztal [1,*] and Paulina Hatlas-Sowinska [2]

[1] Faculty of Computer Science and Telecommunication, Maritime University of Szczecin, Poland; l.misztal@am.szczecin.pl

[2] The Institute of Mathematics, Physics and Chemistry; Department of Mathematics, Maritime University of Szczecin; p.hatlas-sowinska@pm.szczecin.pl

* Correspondence: l.misztal@am.szczecin.pl

Abstract: In the article, the authors draw attention to the significant impact of the human factor during collision situations in maritime navigation. The problems in the communication process between ship drivers are so excessive that the authors propose automatic communication. It would be an alternative method to the current one. The presented system comprehensively performs communication tasks during a sea voyage. To reach mentioned goal AI methods of natural language processing and additional properties of metaontology (ontology supplemented with objective functions) are applied. Dedicated to maritime transport application, the model for translating a natural language into an ontology consists of multiple steps and uses AI methods of classification for the recognition of a message from the ship's bridge. The reverse model is also multi-stage and uses the created rule-based knowledge base to create natural language sentences built on the basis of the ontology. Validation of the model's accuracy results was conducted through accuracy assessment coefficients for information classification, commonly used in science. Receiver operating characteristic (ROC) curves represent the results in the datasets. Our study confirmed the correctness of the assumptions, the designed system architecture and the algorithms developed in the prototype.

Keywords: safety; metaontology; automation of communication; data mining; natural language processing; human factor

1. Introduction

Human errors are the most common cause of navigational accidents. One classification of these errors was introduced by Reason [1]. The work [2] describes the '80 - 20' rule, which states that up to 80% of accidents are due to human error and 20% are technical accidents. The classification given herein is a general form of listed human errors [3]. Other approaches to this issue can be found in [4] or [5].

1.1. Analysis of Marine Accidents

The analysis of maritime accidents was carried out using official reports from maritime accident investigation bodies and the Maritime Statistical Yearbooks for the years 2019 - 2022 [6].

Statistics collected between 2019 and 2022 [7] including marine accidents broken down by type, and causes of marine accidents and incidents, show that there has been a noticeable improvement in the level of safety in shipping and accident statistics are on a downward trend. However, the humans still remain the main factor causing most accidents. The specific causes of maritime accidents are listed in Table 1.

Table 1. Causes of accidents.

CAUSES OF ACCIDENTS	2019	2020	2021	2022

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

Error in navigation or maneuvering	**24**	**10**	**4**	**16**
Loss of control	0	2	0	0
Damage to equipment	20	14	8	8
Slipping, loss of balance, unfortunate fall, hitting	1	4	0	0
Bad weather conditions	2	3	3	0
Hull leakage	0	0	0	0
Mechanical defect	0	0	0	0
Lack of caution at work	4	11	41	28
Immobilization by fishing nets	0	0	0	8
Other	22	6	7	8
Total	73	50	63	68
TOTAL OF HUMAN ERRORS	**28**	**23**	**45**	**44**

Source: Authors' study based on "Data by the State Commission on Maritime Accident Investigation".

To summarize the statistics in Table 1, below are the percentages representing the human factor in the listed causes of marine accidents and incidents. Human errors accounted for:

- 38.3% of all causes in 2019;
- 46% of all causes in 2020;
- 71.4% of all causes in 2021;
- 64% of all causes in 2022.

According to the latest data from Allianz, 38 large vessels were completely lost in 2022, a drop from 59 in 2021. The 2023 Safety and Shipping Review shows that maritime safety has improved over the past 10 years, but the human factor - as a cause of accidents - still ranks highest. The region including southern China, Indochina, Indonesia and the Philippines had the highest number of shipping losses in 2022 – a total of 10, ranking first in terms of transport losses over the past decade [8].

The latest edition of the Annual Overview of Marine Casualties and Incidents [9] states that in 2022, 2510 marine accidents and incidents were reported, the figure smaller by 182 compared to 2021, and by 84 compared to 2020. After the traffic was reduced in 2020 due to the COVID pandemic, it increased in 2021, along with resumed cruise ship and ferry operations, and reached pre-pandemic levels in 2022. However, the number of casualties and incidents reported that year is 5.1% below the annual average and below the average of 2,670 events before the pandemic.

Table 2. Global shipping losses by number of vessels by region, 2013-2022. Total losses, vessels over 100 gross tons.

REGION	**2010-2018**	**2019**	**2020**	**2013-2022**
S. China, Indochina, Indonesia and Philippines	216	12	16	204
Baltic	136	1	Data not available	Data not available
Japan, Korea and N. China	102	2	2	76
British Isles, North Sea, English Channel and Bay of Biscay	68	2	3	54
Arabian Gulf and approaches	49	0	4	45
West African Coast	36	3	3	34
West Mediterranean	38	0	Data not available	118
East African Coast	30	0	1	34
Bay of Bengal	24	2	Data not available	27
Russian Arctic and Bering Sea	23	Data not available	3	Data not available

 doi:10.20944/preprints202502.0016.v1

All other regions	188	19	10	174
The whole world	**910**	**41**	**42**	**766**

Source: Authors' research, based on: Allianz Commercial, Safety and Shipping Review 2023 Copyright © 2023.

Aiming to contribute to the improvement of maritime safety, the authors continue their research on this topic. The above statistics indicate that human errors still account for a significant percentage as a cause of maritime accidents. In previous publications [10], the authors proposed a model for the translation of natural language into ontology and vice versa in a maritime vessel autonomous navigation system. The system comprehensively implemented communication tasks during a voyage at sea. Machine learning methods derived from text mining and basic and additional properties of ontologies were used. The authors developed a new ontology formula, specifically dedicated to maritime transport applications. The key elements of the prototype were the sequence of communication messages from the navigation bridge, the decomposition and extraction of the communication sequence and the rule base. The presented model was implemented and verified on selected scenarios of collision situations at sea. The results of the research confirmed the correctness of the assumptions made, the designed system architecture and the developed algorithms in the prototype. The current research aims to implement a reverse translation model, i.e. a transition from natural language to ontology, in a system for autonomous, machine-to-machine (M2M) and semi-autonomous (human-machine) navigation of a sea-going vessel.

1.2. Analysis of the State of Knowledge

The process of conducting a ship requires continuous exchange and processing of navigational information. The correctness of the decisions made is influenced by both the extent, accuracy and reliability, as well as the correct perception of the information. Ship navigators are required to use all available means to assess the navigational situation: ship's equipment and systems, including AIS, ARPA, ECDIS, voice communication and other facilities. Voice communication provides a communication channel for obtaining additional information and, where appropriate, making arrangements. An analysis of the maritime court decisions shows that, in the case of a collision, the failure to establish voice communication with the other vessel was one of the allegations made against the vessels involved in the accident. Wrong decisions may be caused by the failure to establish voice communication, the misconduct of the communication or the misunderstanding of the information thus transmitted. Such errors can be due to fatigue and stress, which in turn reduces mental resilience and personal safety, lowers self-esteem and situational awareness, impairs leadership qualities, and increases decision-making time. Errors may also result from poor English language skills. The disadvantages of spoken communication include the problem of decoding the message at the semantic level, polarisation (the tendency to express extreme opinions), labelling (noticing problems by naming them rather than analysing them), mixing facts and conclusions, and static evaluation (i.e. the lack of verification of opinion required in constantly changing circumstances).

The primary task of navigation is to ensure safe navigation by avoiding hazards during the execution of a sea voyage. Establishing direct automated communication between ships may reduce wrong decisions and consequent wrong actions leading to maritime accidents. This mainly refers to dangerous situations requiring a decisive action to avoid a collision, i.e. close quarters situations, where avoiding a collision requires agreed joint action by the navigators of the meeting vessels. These actions, called the last-minute manoeuvre, are aimed at avoiding a collision or, when this is not possible, minimising its consequences.

1.3. Communication Process

A communication model prototype for the transformation from natural language messages issued by a ship's navigator to an ontological form understood by an autonomous system was presented in [1]. It consists of nine steps: input in the form of natural language messages, tokenization, removal of redundant information, transformation of words to the basic form,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

lemmatization, part-of-speech recognition, message type classification [29-32, 34-40], rule set generation using artificial intelligence and, finally, output in the form of an ontology. The algorithm of the communication process is shown in Figure 1.

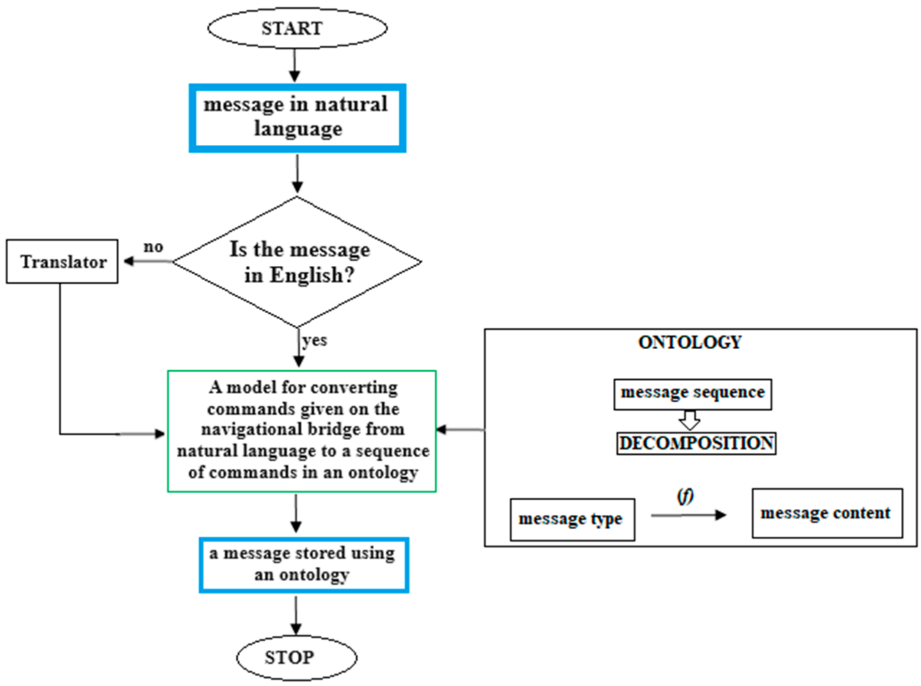

Figure 1. Algorithm of the communication process.

2. Materials and Methods

2.1. Metaontology

A number of scientific publications introduce metaontology [11-14]. Generalizing some definitions and applications of metaontology, we can state that it contributes, inter alia, to the understanding of a given ontology and introduces a formal model for understanding a given ontology-based system. For the model to work comprehensively, we must define the rules, formalize and conceptualize the metaontology. The authors, for the purpose of their research, formulated an objective function containing lexical rules that will enable the improvement of the maritime transport communication (M2M) system.

The ontology was built using the Protégé program in English (maritime language required by the STCW Convention). It represents the structure of individual classes, instances and entities. The communication ontology reflects the real-world processes of acquiring and sharing information and the negotiations carried out between traffic participants. The navigation ontology contains navigational terms, which are its elements for creating messages. In order to send messages that are unambiguous, each type of message has been assigned a category.

The whole ontology can have an tree structure (Figure 2), which naturally represents the hierarchy of data (physical and abstract objects, concepts, etc.). Trees, on the other hand, make searching easier and faster, and allow the sorted data to be easily operated on. Therefore, the latter method uses a data structure representing a mathematical tree, where the root of the tree is the ontology and the path is the sequence of edges connecting the tree.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

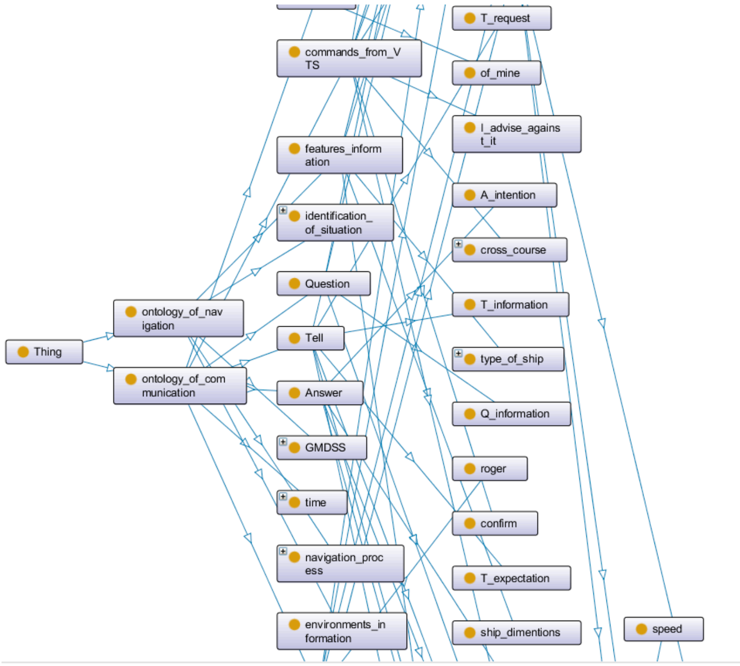

Figure 2. A mathematical tree presenting a fragment of the navigation and communication ontology.

The developed communication ontology is universal enough to work with any navigation ontology: marine, aerial, submarine, etc.

The ontology definition, formula and its abstraction classes are discussed by the authors in previous works [1,6,15]. The set O defines the structure of the concepts, the relations between them, as well as the theory about the model being defined. It contains: axiom of choice, class of abstraction, relationship, message-generating function and message interpretation function.

Presented below is the authors' original objective function that, in combination with the existing ontology, forms a metaontology applicable to automated communication. The objective function contains the message type and specific category, and determines which connections are possible and which are rejected by the function.

$$f_O = m_t(A) \cdot \sum_{n=1}^{4} A_n + m_t(Q) \cdot \sum_{n=1}^{3} Q_n + m_t(T) \cdot \sum_{n=1}^{7} T_n, \quad (1)$$

where:

f_O - objective function;

m_t - message type:

m_t (A) - Answer,

m_t (Q) - Question,

m_t (T) - Tell;

A_n - n-th category from type: Answer,

Q_n - n-th category from type: Question,

T_n - n-th category from type: Tell.

The communication ontology takes into account real-world processes of acquiring and sharing information and negotiations carried out between traffic participants. The main classes in this ontology include message structure, objects, time and short phrases. The message structure consists of the body and the header. There are three message types: answer, question and tell, which have specific categories, e.g.: question, demand, or request. Only the type: 'Answer' has one defined category which informs the recipient that the message they sent was not understood, i.e.: "A_ do not understand".

In order to make the message to be sent unambiguous, each type of the message was assigned a specific category, as summarized in Table 3.

Table 3. Message types with categories.

Type	Category	Meaning
Answer		
	A_information	Response by sending information
	A_intention	Response expressing an intention
	A_permission	Response expressing permission
	A_do not understand	Response informing that message was misunderstood
Question		
	Q_information	Request for information
	Q_intention	Asking about intention
	Q_permission	Question about permission
Tell		
	T_information	Sending information
	T_intention	Message expressing intention
	T_permission	Granting a permission
	T_demand	Sending a request
	T_expectation	Message expressing expectation
	T_request	Sending a request
	T_warning	Sending a warning

2.2. The New Conversation Model

The creation of a communication model, the foundation for operations performed in the ontology, was based on the referenced literature, and experts' knowledge of marine communication. A number of communication models have been created, such as Lasswell's persuasive act model, Shannon and Weaver's signal transmission model, Newcomb's triangular model, Schramm's community of experience model, the hypodermic needle model, the two-stage communication model, McCombs and Showa's agenda-setting process model and other. None of the aforementioned models is suitable enough for building a communication ontology for shipping.

Lexical analysis is important in developing an ontology-based message structure, as this will allow data with a specific syntax to be read unambiguously. When loading such data, the syntax must be recognized before it can be processed. The coding action consists in dividing the loaded string of words into smaller syntactic elements, lexemes, then analyzing a string of lexemes. The separate module that deals with this task is aimed at increasing the efficiency of communication. During the lexical analysis of the ontology, a string of words is loaded and split into lexemes. However, what is passed on is not exactly lexemes. It is information representing the meaning of a lexeme by a letter symbol and an optional attribute. The symbol represents information about the lexeme type. If lexemes of a given type carry a certain 'value', an attribute, equal to this value, is attached to the symbol.

By analyzing previous studies, the authors have created a new communication model that extends the previous one (Figure 3).

 doi:10.20944/preprints202502.0016.v1

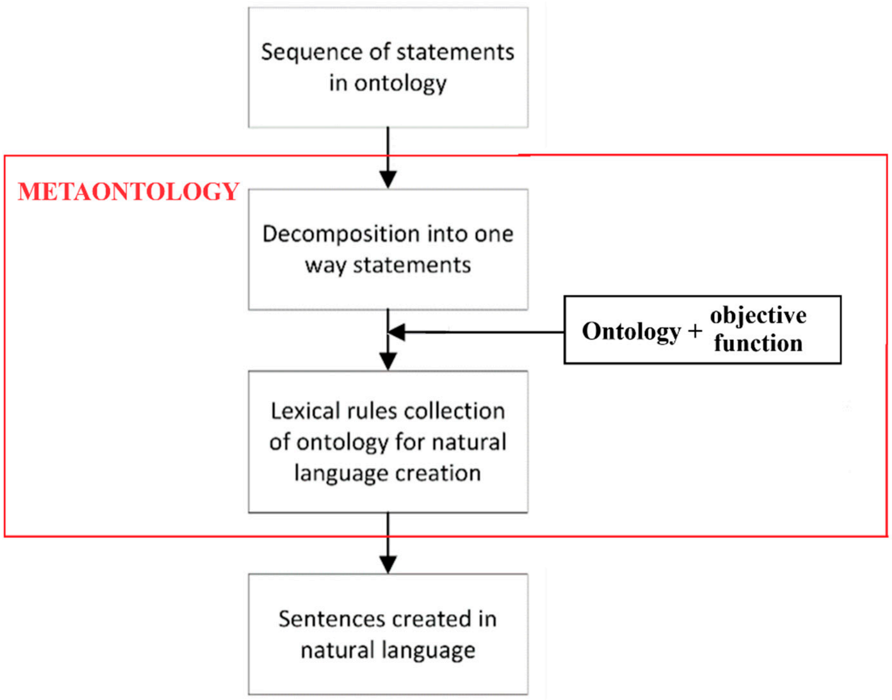

Figure 3. New conversation model.

2.3. Data Validation

In the model transforming natural language to ontology and ontology to natural language, a two-level method of measuring the accuracy of the proposed model was adopted:

- the first level includes checking the accuracy of the transformation at the level of single ontological items to natural language sentences or at the level of single natural language sentences to ontology,
- the second level checks the accuracy of the transformation of all maritime communication scenarios from ontology to natural language and from natural language to ontology.

To measure the quality, universally recognized qualitative coefficients [19] based on the uncertainty matrix [20] used in classification [21] studies were adopted. Due to mentioned necessity for accuracy measurement, the division of learning sets and test sets was carried out. Datasets for the first level considering single natural language sentences or single ontology items are smaller. The second level dataset of entire scenarios is significantly bigger than previously mentioned set. For both types of scenarios the same division of sets was adopted, allowing the accuracy to be estimated in the objective and reliable way. The accuracy of the measurements for the two levels was determined using a confusion matrix, presented in the Table 4.

Table 4. The confusion matrix for accuracy measurements.

Predicted data / Actual data	1	2
1	TP	TN
2	FN	FP

The accuracies of the methods for both types of transformation will be verified for the first and second level using the following coefficients:

$$PPV = \frac{TP}{TP + FP} \tag{2}$$

where:

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

PPV - positive predictive value,
TP - true positive rate,
FP - false positive rate.

$$NPV = \frac{TN}{TN + FN} \tag{3}$$

where
NPV - negative predictive value,
TN - true negative rate,
FN - true positive rate.

$$SE = \frac{TP}{TP + FN} \tag{4}$$

where
SE - sensitivity rate.

$$SP = \frac{TN}{TN + FP} \tag{5}$$

where
SP - specificity rate.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where
ACC - the total accuracy rate.

$$ERR = \frac{FP + FN}{FP + FN + TP + TN} \tag{7}$$

where
ERR - total error rate.

$$F = \frac{(\beta^2 + 1) \cdot P \cdot TP}{(\beta^2 \cdot P) + TP} \tag{8}$$

$$\beta = \frac{FN}{TP + FN} \tag{9}$$

where
F - F-measure,
β = FN / (TP+FN) - beta false positive rate,
P - sensitivity rate.

$$FOM = \frac{W \cdot L_{PP}}{(W+1) \cdot L_{WP}} + \frac{L_{PN}}{(W+1) \cdot L_{WN}} \tag{10}$$

where
FOM - prediction quality rate,
W - ratio of the cost of negative predictions to the cost of positive predictions,
LPP - number of correct positive predictions,
LWP - number of all positive predictions,
LPN - number of correct negative predictions,
LWN - number of all negative predictions.

The results of the estimation of the ontology transformation into natural language using a learning curve is presented below (Figure 4).

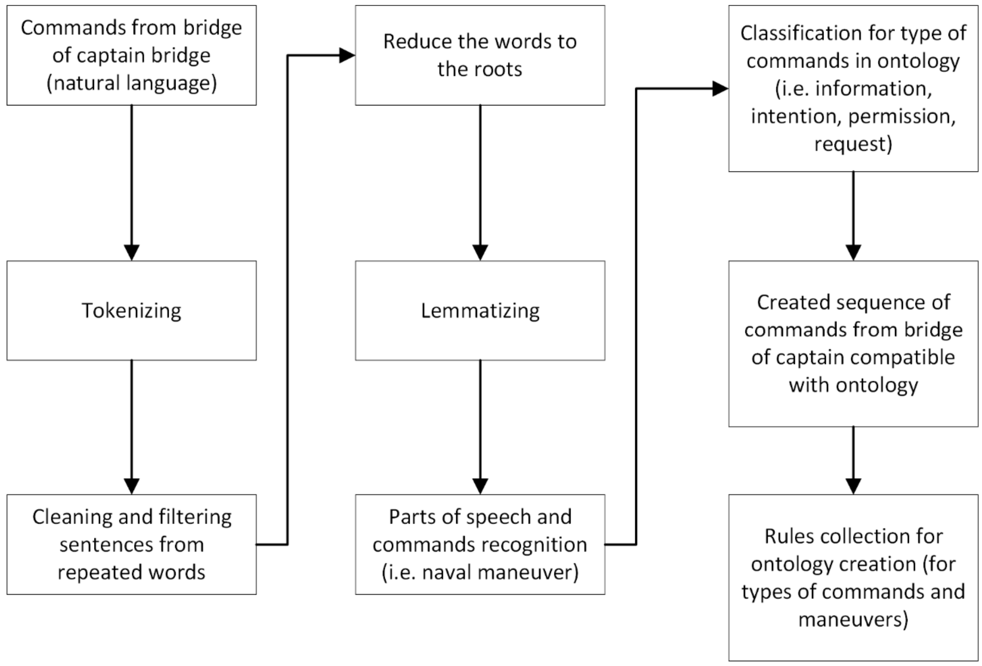

Figure 4. Model of natural language transformation into ontology.

The input data for the method will be shown in the relevant tables in Chapter 3.

3. Results

3.1. Example Collision Situations Using the Proposed Model

In order to achieve the objectives set in this study, the method of simulation research using an ECDIS simulator was applied, allowing the implementation of pre-planned scenarios of ship encounters. Each ship model allows full use of the ship's equipment. Visualization of the sea area enables the user to carry out visual observation. The navigator can also use any of the systems that the modern navigational bridge is equipped with, including ARPA and AIS, used in our scenarios. The ship models have full course and speed maneuvering capabilities.

Two ship encounter situations involving two vessel models (non-autonomous real-time) were selected for the study:

- crossing courses:

vessel A sails on course 270°, vessel B on course 000°; the meeting vessels are of medium size (length approx. 170 m), both proceed full ahead and are visible to each other;

- opposite or reciprocal courses:

vessel A sails on course 180°, vessel B on course 000° ; the meeting vessels are of medium size (length approx. 170 m), both proceed full ahead and are visible to each other;

The selected scenarios are given as standard ship encounter situations as described in the Collision Regulations. At the same time, these situations raise many interpretation issues regarding the safety of the maneuver being carried out.

Each situation was repeated several times on the simulator, depending on how the situation had been developing. The recordings showed that each time the communication processes and related maneuvers by the ships ran differently. For these simulations, appropriate scenarios were prepared and recorded, allowing the initial situation to be replayed several times, and navigators had the possibility to take subsequently various actions.

3.2. Example Collision Situations Using the Proposed Model

Wind force 3° B. Sea state 2. Ship A (heading 270°) sent a message to ship B (heading 000°) inquiring about its intentions. Ship B replied with two messages. The first in the form of a request for ship A to change course slightly to the left. The second one was in the form of permission to pass in front of him. Ship A agreed and changed course 30 degrees to port. The ships in good visibility passed each other by a distance of 0.5 nautical miles, which is acceptable given the maneuver and good visibility.

Table 5. Scenario of an example collision situation on intersecting courses.

Standard message in natural language (based on Standard Marine Communication Phrases)	Expected outcome in the ontology
What are your intentions?	1. {ontology of communication, message structures, body, Question, Q_intention}
I want to pass ahead of you (1,2). Change course a little to port.(3 ,4)	1. {ontology of communication, message structures, body, Tell, T_request}, 2. {ontology of navigation, features information, Ship maneuvering, Passing ,ahead of you}. 3. {ontology of communication, message structures, body, Tell, T_permission}, 4. {ontology of navigation, features information, navigation information, course, alter course, to port}.
Roger.	1. {ontology of communication, phrase, roger}

The basic statistics of data prepared in the scenarios are given below, with the count of sentences, words and ontology items:

Table 6. Part of speech statistics

ALL.	PRP	NN	DT	JJ	VB	WP	RB
14	2	4	1	1	4	1	1

Table 7. Statistics of the ontologies.

Ontology of communication	Ontology of navigation
4	2

Table 8. Statistics of the most frequently occurring words in the root form

Word	Roger	Are	Pass	Ahead	Change	Course	Port	Little
Quantity	1	1	1	1	1	1	1	1

3.3. Example Situation: Opposite or Reciprocal Courses

Good visibility. Ship A heading 180°, ship B heading 000°. Ship A informs ship B that they are almost straight ahead and proposes to change course to starboard. Ship B rejects this offer, sending a reply saying it is safe. Ship B proposes to pass each other on their starboard sides and maintain course and speed. Ship A agrees and sends a confirmation. Ship B also sends an acknowledgment.

Table 9. Scenario of an example collision situation on reciprocal courses.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

Scheme 1.	Expected outcome in the ontology
We have head on situation. (1,2) Let's change both to starboard. (3,4)	1. {ontology of communication, message structures, body, Tell, T_information}, 2. {ontology of navigation, features information, navigation information, course, head on}, 3. {ontology of communication, message structures, body, Tell, T_permission}, 4. {ontology of navigation, features information, navigation information, course, alter course, both to starboard}
NO. (1) The courses are safe.(2,3) We pass each other on the starboard to starboard. (4,5) Keep the course and speed. (6,7)	1. ontology of communication, phrase, negativ}, 2. {ontology of communication, message structures, body, Answer, A_information}, 3. {ontology of navigation, features information, navigation information, course, safe course} 4. {ontology of communication, message structures, body, Tell, T_information}, 5. {ontology of navigation, features information, Ship maneuvering, passing, starboard to starboard}, 6. {ontology of communication, message structures, body, Tell, T_information}, 7. {ontology of navigation, identification of situation, keep_the_course_and_speed}.
Roger	1. {ontology of communication, phrase, OK}
Roger	1. {ontology of communication, phrase, OK}

The basic statistics of data prepared in the scenarios are given below, with the count of sentences, words and ontology items:

Table 10. Part of speech statistics

ALL	PRP	NN	DT	JJ	VB
23	1	8	5	1	8

Table 11 . Statistics of the ontologies

Ontology of communication	Ontology of navigation
8	5

Table 12 . Statistics of the most frequently occurring words in the root form

Word	Let	Starboard	Change	Course	Safe	Ok	Speed	Keep
Quantity	2	2	2	2	1	2	1	1

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

3.4. Accuracy Results

Accuracy results of model for the translation of a message written in an ontology into natural language are given below:

Table 13. Accuracy measurements for the first level, which includes single sentences for the second scenario

Predicted data / Actual data	**1**	**2**
1	4	0
2	1	3

Table 14. Accuracy rates for the model of translation from natural language to ontology

Factor	*PPV*	*NPV*	*SE*	*SP*	*ACC*	*ERR*	*F*
Accuracy	*1*	*0,75*	*0,8*	*1*	*0,875*	*0,125*	*0,825*

Accuracy measurements for the second level, which include complete marine collision scenarios with all sentences in all the scenarios:

Table 15. Accuracy measurements for the second level, which include complete marine collision scenarios with all sentences in all the scenarios

Projected data / Actual data	***1***	***2***
1	*72*	*12*
2	*4*	*61*

Accuracy obtained using the learning curve (Figure 5):

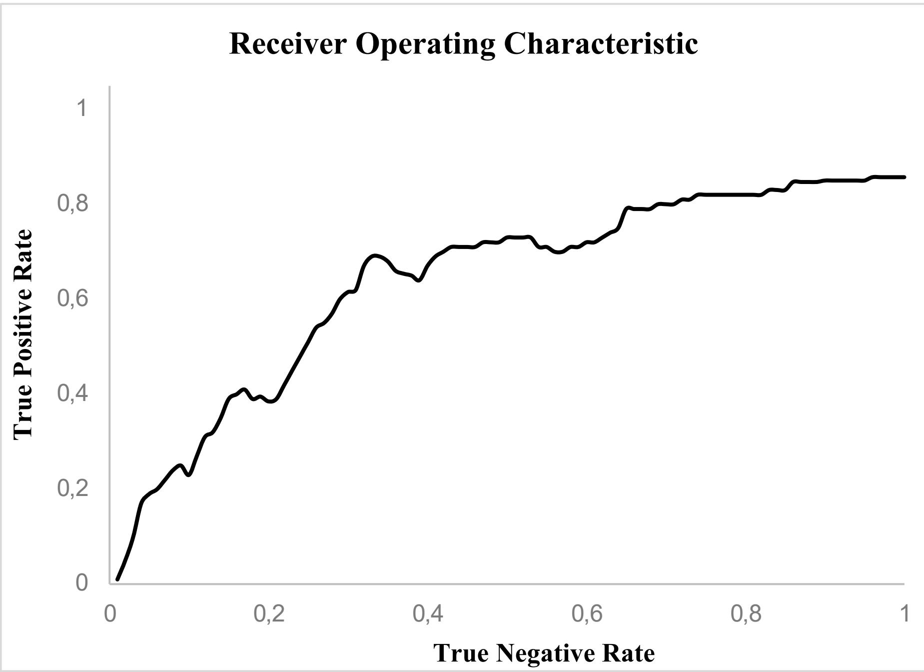

Figure 5. ROC curve for the given test scenarios.

Table 16. Accuracy rates for the model of translation from natural language to ontology.

Rate	**PPV**	**NPV**	**SE**	**SP**	**ACC**	**ERR**	**F**
Accuracy	0,8571	0,9384	0,9473	0,8356	0,8926	0,1073	0,9499

The prototype was developed using the Python programming language [22], object-oriented programming techniques [33] and the Natural Language Tookit programming library [25,26] for natural language processing. The scikit-learn library [23,33] was also used to enable the use of machine learning methods for the proposed translation model. The TKinter library [24] served for programming a user-friendly graphical user interface (GUI) in the form of a desktop application.

The prototype implements the model for converting sentences from natural language to sequences of instructions in an ontology [27]. The model consists of a number of steps, where the input contains navigator's messages in the form of a sequence of natural language sentences, while the output is a sequence of messages in the form of an ontology. At the start of the operation, the model tokenizes [28] the data into individual words and sentences. It then discards words that should not be relevant to the spoken messages, but are typical of human verbal interactions (e.g. please, however, but, etc.). The next step is stemming, i.e. reducing the number of similar words that derive from a single parent (root). Lemmatization, in turn, aims to reduce the number of words by using a root word, thus reducing words with different endings, for example. The reduction of information is followed by labelling the speech parts, such as nouns, verbs, adjectives and other, as well as identifying the types of messages. This is an important step before using the classifier to recognise the types of messages issued from the bridge, such as information, request, or other. Then, on the basis of the knowledge base in the form of rules, an operation will be carried out to create a sequence of messages in the form of classes from a set of ontologies. This form will be unambiguously readable by the autonomous system.

The prototype, a programmed application, has the following functionalities performed by classes programmed in Python: Loaddata class - enables loading individual scenarios into defined complex data structures, Prepword class - performs tasks such as Savedata and Classify. Savedata transforms the prepared and decoded data to the format necessary for classifications in artificial intelligence, while Classify defines a machine learning model and implements the learning of the model including the determination of the accuracy of the natural language transformation to the ontological notation. Another class, AppGUI, was also programmed to create a graphical user-friendly interface for the proposed prototype.

4. Discussion

The presented measurements are based on 20 maritime traffic collision scenarios. The model was implemented and verified by using selected maritime collision scenarios, which were repeatedly modified and replicated to obtain a significant number of statistical tests.

For each scenario, messages produced by the navigator (human voice) and by means of speech synthesis techniques were used, transmitted by the radio voice communication methods available on board vessels. The above sentences were then transformed by the prototype model from natural language to ontology, described in this article. Our model performs a linguistic analysis of the sentence identifying the parts of speech. In subsequent steps the model matches the selected parts of speech with specific items in the communication or navigation ontology using classification models trained by means of machine learning methods.

The prototype software successfully executed the translations for the communication scenarios loaded for testing. The introduced metaontology as well as the objective function performed well to complete the task. The authors will continue to explore and widen its applications, e.g. land and air modes of transport.

Accuracy results by presented transformation model in the overall perspective specified by coefficient total accuracy rate yields around 90%. The coefficient total error rate that indicate level of error is about 10%. The receiver operating characteristic (ROC) curve also shows high accuracy for training set that is significantly over average line. In the summary it means reaching satisfying accuracy results with usage of proposed transformation model for presented accuracy coefficients. Because domain of application of the method is critical for safety in maritime traffic it is necessity for further scientific researches of presented model for achieving greater overall accuracy.

 doi:10.20944/preprints202502.0016.v1

Natural language processing on artificial intelligence, enabling a computer to understand, use and generate human language externally, already visible. Due to the safety requirements in maritime transport, NLP can be implemented through an automatic communication system, which is proposed by the authors [16]. Ontology as a structure contains a set of concepts and their interrelationships in a specific area that can help resolve and standardize terminology, which is key to the connection between systems, entities and the connection in the transport chain [17,18]. Further work aimed at developing and improving the developed models and operating algorithms. The goal is to access sources that can be connected to a level of maritime security.

4.1. Survey Research

A survey was conducted to verify the hypothesis put forward by the authors: the use of the developed model of the communication process for an automatic communication system in maritime navigation will increase the effectiveness of communication by improving the exchange and clear interpretation of the transmitted information. 32 people participated in the research. The respondents, aged 31-50, held the rank of senior deck officer. The study aimed to compare communication taking place during collisions (through the current use of VHF) and those implemented through automatic ontology-based communication. Here are some of them (Figures 6-11).

- Do you think that the use of automatic communication was correct in the presented collision situation?

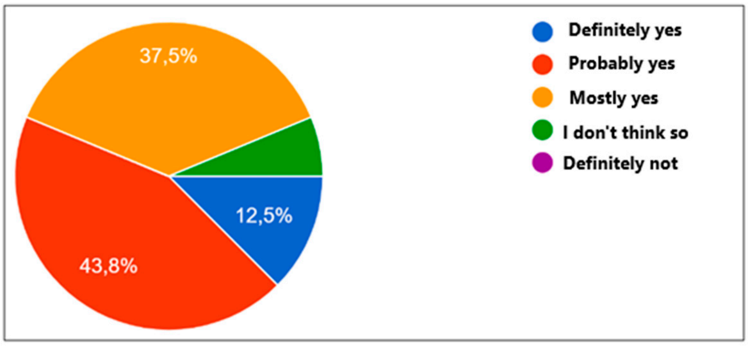

Figure 6. Answers regarding the use of semi-automatic communication.

- Do you think that voice communication (except VHF) on the ship is sufficient?

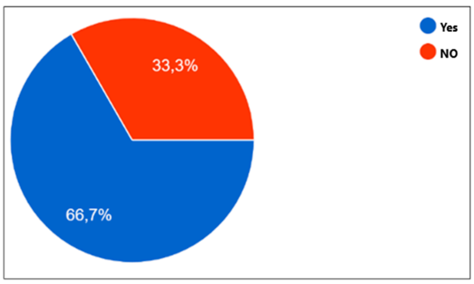

Figure 7. Voice communication replies (via FM).

- Do you think that communication with the second navigator via VHF is understandable?

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 February 2025 doi:10.20944/preprints202502.0016.v1

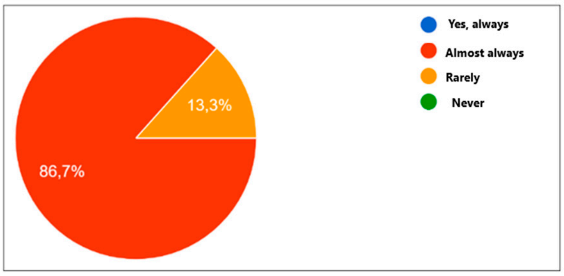

Figure 8. Answers regarding VHF communication.

- Do you think that talking through an automatic communication system can shorten the time it takes for navigators to communicate, e.g. in a collision situation?

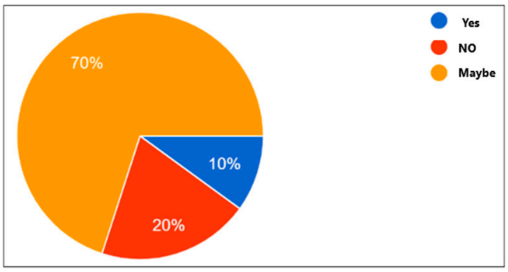

Figure 9. Automatic communication talk time replies.

- Do you think that the current communication system at sea should be modernized?

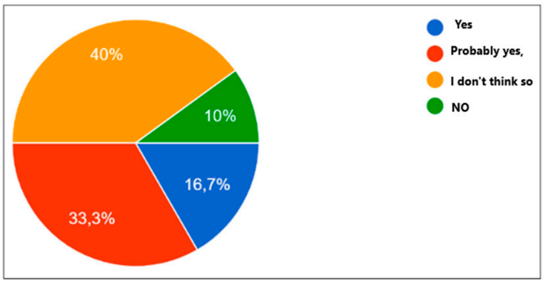

Figure 10. Answers regarding the modernization of the communication system.

- Do you think that sending clearly understandable messages during the automatic communication process will increase the level of safety in maritime navigation?

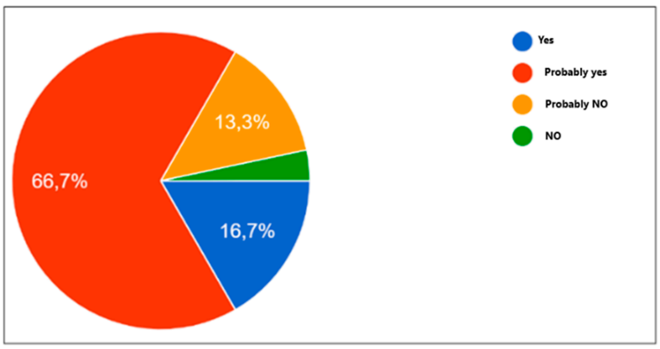

Figure 11. Answers regarding security level enhancement for semiautomatic communication.

The analysis of the survey results showed that:

- 70% of respondents believe that an additional method of communication between navigators is needed (beyond VHF);
- 63.3% of respondents would use communication based on sending messages with another ship;
- 80% of respondents state that talking through an automatic communication system can shorten the time it takes for navigators to communicate, e.g. in a collision situation;
- 83.4% of respondents believe that sending clearly understandable messages during the semi-automatic communication process will increase the level of safety in maritime navigation.

5. Conclusions

The authors have demonstrated that it is feasible to successfully apply metaontology and machine learning methods in the proposed prototype software for ship-to-ship communication. The application of metaontology for a semi-autonomous as well as machine-to-machine communication systems may contribute to avoiding collisions (or significantly reduce collision consequences). Notably, apart from collision and close-quarter situations, the proposed solutions can also be used to conduct routine navigation. Early established communication in the indicated manner will allow potentially dangerous situations to be resolved much earlier, give navigators more time to analyze and observe how the situation develops, and reduce stress associated with performing last-minute maneuvers.

Testing the data for the model of translation from natural language to the ontological form uses common quality coefficients, which measure the accuracy of machine learning using classification methods. Seven coefficients, or rates, were used, i.e. total accuracy rate, total error rate, positive prediction rate, negative prediction rate, sensitivity, specificity and F-Measure.

The prototype described herein will be improved to increase the accuracy obtained in the tests, in particular in terms of the machine learning methods used, taking into account the aforementioned coefficients while increasing the input database in the form of natural language sentences uttered by ship navigators, to take into account real communication situations in shipping.

Author Contributions: Conceptualization, Paulina Hatlas-Sowinska and Leszek Misztal.; methodology, Paulina Hatlas-Sowinska.; software, Leszek Misztal.; validation, Paulina Hatlas-Sowinska and Leszek Misztal.; formal analysis, Paulina Hatlas-Sowinska.; investigation, Paulina Hatlas-Sowinska.; resources, Paulina Hatlas-Sowinska and Leszek Misztal.; data curation, Leszek Misztal.; writing—original draft preparation, Paulina Hatlas-Sowinska and Leszek Misztal.; writing—review and editing, Paulina Hatlas-Sowinska and Leszek Misztal.; visualization, Paulina Hatlas-Sowinska and Leszek Misztal.; supervision, Paulina Hatlas-Sowinska and Leszek Misztal.; project administration, Leszek Misztal.; funding acquisition, Paulina Hatlas-Sowinska. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by subsidy of the Ministry of Science and Higher Education for statutory activities of the Maritime University of Szczecin, research project No. 1/S/ITM/16.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Set of research data is published in open access DOI resource and is available on: https://doi.org/10.17882/104517.

Acknowledgments: This research outcome has been achieved under the research project No. 1/S/ITM/16 financed from a subsidy of the Ministry of Science and Higher Education for statutory activities of the Maritime University of Szczecin.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ROC	Receiver operating characteristic
VHF	Very High Frequency
PRP	Personal pronoun
NN	Noun
DT	Determiner
JJ	Adjective
VB	Verb
WP	Pronoun
RB	Adverb

References

1. Reason J. Human Error. *Cambridge University Press* 1990. 324 s.
2. Harrald JR, Mazzuchi TA, Spahn J, Van Dorp R, Merrick J, Shrestha S, i in. Using system simulation to model the impact of human error in a maritime system. *Safety Science* October 1998;30(1–2):235–47.
3. Gucma L. Wytyczne do zarządzania ryzykiem morskim. *Wydawnictwo Naukowe Akademii Morskiej w Szczecinie*. 2009.
4. Haugen S. Probabilistic Evaluation of Frequency of Collision Between Ships and Offshore Platforms., *University of w Trondheim*; 1991.
5. Technica. The Risk of Ship/Platform Collision in the Area of United Kingdom Continental Shelf. *London*; 1986. Report No.: Raport OTH 86 217.
6. GUS. *stat.gov.pl*. Rocznik Statystyczny Gospodarki Morskiej 2015. Available on: https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-statystyczny-gospodarki-morskiej-2015,11,8.html (accessed on 13 Nov 2018)
7. webmaster. Accident Investigation Publications [Internet]. Available on: https://www.emsa.europa.eu/accident-investigation-publications/annual-overview.html (accessed on 10 DEC 2024)
8. III [Internet]. [cytowane 10 grudzień 2024]. III. Dostępne na: https://www.iii.org/
9. webmaster. European Maritime Safety Agency (EMSA) - Quality Shipping, Safer Seas, Cleaner Oceans [Internet].. Available on: https://www.emsa.europa.eu/ (accessed on 10 DEC 2024)
10. Hatłas-Sowińska P, Misztal L. Application of a text mining method in navigation and communication for enhancing maritime safety. PLOS ONE. 22 MAR 2024;19(3):e0299582.
11. Arreman J. Review of Ontology and Metaontology: A Contemporary Guide by Francesco Berto & Matteo Plebani. Philosophy in Review, Vol 37, No 2., 2017. 1 APR 2017;37:53–5.
12. Berto F, Plebani M. Ontology and Metaontology: A Contemporary Guide. *Bloomsbury Publishing* 2015. 265 s.
13. Vasilecas O, Eder J, Caplinskas A, redaktorzy. Databases and information systems IV: selected papers from the Seventh International Baltic Conference, DB&IS'2006. Amsterdam ; Washington, DC: *IOS Press* 2007. 305 s. (Frontiers in artificial intelligence and applications).
14. Hofweber T. Ontology and Metaphysics. W: Hofweber T, redaktor. Ontology and the Ambitions of Metaphysics [Internet]. *Oxford University Press* 2016. Available on: https://doi.org/10.1093/acprof:oso/9780198769835.003.0001 [accessed on 18 SEP 2024]
15. Hatlas-Sowinska P, Wielgosz M. Ontology based approach in solving collision situations at sea. *Ocean Engineering*. 15 SEP 2022;260:111941.
16. SciSpace - Paper [Internet]. IEEE; 2013. s. 1–3 Natural language processing future. Available on: https://typeset.io/papers/natural-language-processing-future-3pcnn7ffka (accessed on 10 DEC 2024)
17. SciSpace - Paper [Internet]. 2012. A Research-oriented Survey and Current Status on Feature Extraction, Ontology Construction towards Natural Language Processing. Available on: https://typeset.io/papers/a-research-oriented-survey-and-current-status-on-feature-24vbgzp5h5 (accessed on 10 DEC 2024)

 doi:10.20944/preprints202502.0016.v1

18. SciSpace - Paper [Internet]. Routledge; 2017. s. 325–38 Natural Language Ontology. Available on: https://typeset.io/papers/natural-language-ontology-25fq8n6l3g (accessed on: 10 DEC 2024)
19. Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques second edition, *Morgan Kaufmann Publishers* 2006
20. C.Vercellis Business intelligence – data mining and optimization for decision making, *Wiley* 2009
21. D. Lewis W.Gale Training text classifiers by uncertainty sampling, *ACM SIGIR Conference* 1994
22. Mark Lutz, David Ascher. Learning Python. *O'Reilly*, second edition 2003
23. Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda. Applied Text Analysis with Python. *O'Reilly* 2018
24. Moore Alan D.. Python GUI Programming with Tkinter, *O'Reilly* 2018
25. Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python. *O'Reilly* 2009
26. Harshit Surana, Bodhisattwa Majumder, Anuj Gupta, Sowmya Vajjala. Practical Natural Language Processing. *O'Reilly* 2020
27. Michael W.Berry, Matu Castellanos, Survey of text mining II clustering, classification and retrival, *Springer* 2008
28. Nintin Indurkhya, Fred J. Damerau, editors, Handbook of Natural Language Processing, second edition, *Morgan and Claypool* 2010
29. Christopher Manning, Hinrich Schutze, Foundations of statistical natural language processing, *MIT Press*, Cambridge 1999
30. Olson et al., 2008. Olson D., Delen D., Advanced Data Mining Techniques, *Springer* 2008
31. Triantaphyllou et al., 2006. Triantaphyllou E., Felici G.: Data Mining & Knowledge Discovery based in Rule Induction, *Springer Science* 2006
32. Witten et al., 2005. Ian H.Witten, Eibe Frank: Data Mining Practical Machine Learning Tools and Techniques, *Morgan Kaufmann* 2005
33. Mark Lutz. Programming Python. 3rd Edition. *O'Reilly Media* 2006
34. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly, 2018.
35. Inuiguchi M., Generalizations of Rough Sets and Rule Extraction, *Springer* 2005
36. Pawlak Z., Rough Sets – Theoretical Aspects of Reasoning about Data. *Kluwer Academic Publishers* 1991
37. Duntsch I., Gediga G.. Rough set data analysis. *Methodos Publisher* 2000
38. Bazan J., Synak P., Wrobleski J., Rough Set Algorithms in Classification Problem, *Springer* 2000
39. Drzymala-Busse J., Rough Set Strategies to Data with Missing Attribute Values, *Springer* 2006
40. Oded Maimon, Lior Rokach, Data Mining and Knowledge Discovery Handbook, *Springer Science Business Media* 2005
41. Ramez Elmasri, Shamkant B. Navathe. Fundamentals of Database Systems 4th Edition, *Addison-Wesley* 2004