Computer Science and Information Systems 19(2):913–934 https://doi.org/10.2298/CSIS211101009A

An Approach to Email Categorization and Response Generation

Sasa Arsovski[1], Muniru Idris Oladele[2], Adrian David Cheok[2], Velibor Premcevski[3] and Branko Markoski[3,*]

[1] Raffles University, Menara Kotaraya
Menara Kotaraya, Level 9, #09, 01, Jalan Trus, Bandar Johor Bahru,
80000 Johor Bahru, Johor
sasa.arsovski@gmail.com
[2] Imagineering Institute, Johor Malaysia
Anchor 5, Mall of Medini, 4, Lebuh Medini Utara,
79200 Nusajaya, Johor
idris@imagineeringinstitute.org
adrian@imagineeringinstitute.org
[3] University of Novi Sad, Technical Faculty "Mihajlo Pupin",
23000 Zrenjanin, Serbia
velibor.premcevski@tfzr.rs
markoni@uns.ac.rs

Abstract. The creation of automatic e-mail responder systems with human-quality responses is challenging due to the ambiguity of meanings and difficulty in response modeling. In this paper, we present the Personal Email Responder (PER); a novel system for email categorization and semi-automatic response generation. The key novelty presented in this paper is an approach to email categorization that distinguishes query and non-query email messages using Natural Language Processing (NLP) and Neural Network (NN) methods. The second novelty is the use of Artificial Intelligence Markup Language (AIML)-based chatbot for semiautomatic response creation. The proposed methodology was implemented as a prototype mobile application, which was then used to conduct an experiment. Email messages logs collected in the experimental phase are used to evaluate the proposed methodology and estimate the accuracy of the presented system for email categorization and semi-automatic response generation.

Keywords: email responder, deep learning, AIML, chatbot**.**

1. Introduction

Nowadays, there are many forms of digital communication such as Internet messaging, chat, social networking, etc. Despite all those communication forms, email remains the leading form of business communication [1]. With the huge increase in email overload, users find it very challenging to process and respond to all incoming messages [2].

* Corresponding author

According to [1], mobile email has shown rapid growth; currently, 65% of email users worldwide access their email via a mobile device. Authors in [3] stated that these days email overload has new forms. Users receive different types of emails that match multiple aspects of their life. Research findings in [3] indicate that email overload is present both at work and in private life. These findings suggest new opportunities for email overload research.

The main objective of the personal email responder system that will be presented in this paper is to help users to overcome email overload challenges and minimize user efforts in the email answering process. The first challenge toward building a PER is to identify which types of messages can be automatically or semi-automatically processed.

Email categorization and response generation systems deploy different approaches; however, they can be grouped into the three categories: 1) text retrieval approaches, 2) text categorization by machine learning and, 3) statistical text similarity calculation by matching of text patterns and templates [4].

After an initial study based on a log that contains 12,000 personal email messages, we categorize the email messages in to two main categories. The first category is the query email messages. Query email messages are messages that contain some question and require an answer. The second category is declarative and informative messages. Declarative and informative messages are messages that contain declarative sentences which do not require an answer, or the answer can be created automatically.

For the model construction, we will use Deep learning techniques, NLP tools and, AIML chatbot functionalities. Deep learning [5] is a set of algorithms in machine learning that attempt to learn on multiple levels, corresponding to various levels of abstraction. It typically uses artificial Neural Networks (NN) [6, 7].

The key novelty of this work is an approach that is applied for email message categorization. Also, in this paper, we will present novel system based on AIML functionalities for semiautomatic email response generation.

The rest of the paper is organized as follows: the second section contains related works; the third section provides the dataset analysis details and model architecture; results and discussions are presented in the fourth section. Section five contains our conclusion and future work.

2. Related Works

Automating distinctive features of email systems has a long history. One of the good examples is the work presented in [8]. Motivated by the need to lessen the burden of the large volume of emails of many users, the authors suggested a smart environment for email processing. The personal email assistant can prioritize, filter, refile the incoming emails based on classification, search through the emails, and make vacation responses. Although some of these tasks were already included in different email systems, the authors stated that none of them combined all those capabilities. As an email responder, proposed system creates appropriate routine responses to incoming messages. The compilation of these responses is based on the email subject, current schedule, and the content of the email. The personal email assistant proposed in [8] uses personal calendar, personal preferences, email context, and decides about response content.

According to the work in [4], automated email answering will be a text categorization task if all messages in one text category have the same standard answer. In [9], several machine learning algorithms were adopted. The algorithms included the k-NN, Naive Bayes, RIPPER, and SVM, and were implemented for generating automatic answers to 4490 technical support-related email messages. These messages were grouped into 47 categories; each category had no less than 30 messages. All experiments were carried out using tenfold cross validation. Support Vector Machine (SVM) shows the best performance, with accuracy (the share of correctly suggested standard answers) of 56% for a single answer, and for 78% of the email messages, the correct answer was among the top 5 suggested answers.

In [10], the sentence matching approach is adopted, however, this is a more challenging task than document categorization; it involves locating one or several questions in a message and selecting one or several standard answers to the questions. Query emails are analyzed against a database system which comprises several tag-question (FAQ) and standard answer pair. When the new e-mail arrives, the system analyzes the question in an email and provides an answer mapped to a predefined question template. When the system cannot find the appropriate question template, it must determine the similarity between the new questions and the existing question templates. In this case, the system will calculate the similarity between the concepts present in the sentences that are compared. Using the method defined in [11], to calculate overall similarity score between the two sentences, the system compares a new question with pre-defined set of questions templates. When it finds a most similar question template, the system takes an answer to the template question in response to a new question.

Kosseim et al. [12] use information extraction templates to (i) identify the query message the purpose, the sender, etc., (ii) extract named entities from the query message, (iii) extract relations between the main concepts, and (iv) capture domain specific relations. The next step is semantic validation. The system verifies whether the extracted data and the respective templates all together make any sense as an answer. The third step is an analysis of the obtained information and querying some external sources for new data to complete the answer. Finally, the system fills the answer template with the data and generates the answer text.

Text-pattern matching is another minority approach. Sneiders [4] has developed a technique that operates a set of manually crafted text patterns assigned to FAQs. A text pattern resembles a regular expression. It contains stems of words and their synonyms. It can match phrases, stand-alone words, and compound words. Each FAQ has one or several required text patterns (they must match a query) and one or several forbidden text patterns (they must not match the query). Experiments in two languages (Swedish and Latvian) and two domains (insurance and telecom) showed consistent results: if the system did retrieve an answer, the answer was correct in about 90% of the cases. The recall values were 68% and 76% in the respective language and domain.

Recently, there are a few works that propose state-of-the-art email auto-responder systems. Those approaches use Deep Learning techniques for creating automatic email responders.

In the [13] authors developed a knowledge-based Question-Answer system. The user has the possibility to ask questions to a Question-Answer system that will retrieve related questions from the mailing list and show the significant sentences that were extracted the related questions and their answers. However, the system is limited only to "how" type questions. Google provides a system that analyzes the incoming email and generates a reply [9]. Then, the user can modify the reply, and based on this modification, the system will update the email analysis mechanism. The system consists of two parts: 1) filter and modeler - for performing language analysis and characterization of the email body, and 2) reply composer - for generating a response to incoming email, based on the library of phrases. This system is based on general concepts and does not reflect the persona of email owner.

In [14], the authors address two main prediction issues arising in email systems. The recommendation system can alert the user when the email needs a reply or attachment. This process is very important for facilitating email response management. Another work, presented in the patent from Google [15], describes a method of eliminating unwanted incoming emails. There is also research for using the NN as an engine for detecting spam messages [16]. However, all these works were directed to automate specific functionalities of various email systems, and none of them is a fully automatic system which can express the personality of the email owner. A similar approach is presented in Mahendra [17]. The author suggested a personified email response generator using a deep neural network (DNN). This system, based on the previous email conversations of the owner, can generate a new message with the same style even for the emails that belong to the invitation or meeting category. The method is based on the Seq2Seq model using Long-Short Term Memory and Gated

Recurrent Units that help to obtain satisfactory results for long sentences. The final purpose of the author is to generate a response that will not require any editing afterward. Hence, right now, the system suggested in this work is the personal email assistant that is not yet able to reply automatically to all incoming emails.

All these works are in a various way starting from question answering systems until semi-automatic replying systems. In addition, almost all related works discuss email answering systems in the closed domain which are based on various technologies and apparatus. Our system extends Busemann work [18] into the next stage.

3. Data Analysis and Model Architecture

In this section, we will present analysis details of the email messages dataset and the architecture of our model for email categorization and response generation.

3.1. Email Dataset Analysis and Message Categorization

For the experiment, we use 12,000 randomly chosen personal Gmail messages. Our approach was based on the following hypothesis:

H_1: If an email body text contains a question, then the user needs to answer to this email - the model generates an automatic or semi-automatic answer. If email body text contains statements without questions, the system will present the user the email.

Our approach to message categorization is similar with the work described in Busemann et al. [18]. Optional input for machine learning was presence or absence of linguistic constructions frequent in questions and problem descriptions:

Table 1. Clause level tags

Tag	Description
S	Simple declarative clause, i.e. one that is not introduced by a (possibly empty) subordinating conjunction or a wh-word and that does not exhibit subject-verb inversion
SBAR	Clause introduced by a (possibly empty) subordinating conjunction
SBARQ	Direct question introduced by a wh-word or a wh-phrase. Indirect questions and relative clauses should be bracketed as SBAR, not SBARQ.
SINV	The inverted declarative sentence, i.e. one in which the subject follows the tensed verb or modal.
SQ	Inverted yes/no question, or main clause of a wh-question, following the wh-phrase in SBARQ.

Table 2. E-mail data analysis

Tag	Number of Emails in dataset
SQ	3169
SQ+SBRQ	8326
All messages	12000

Negation at the sentence or phrase level, yes/no and who-when-what-why-where-which-questions. We analyzed email messages dataset using a Penn Tree Bank clause level tags (Table 1) defined in [19]. For analyzing an email dataset, we use Stanford parser Java library [20] and query the dataset against SQ and SBARQ tags. The analysis results are shown in Table 2.

As shown in Table 2, 69.38 % of email messages in the analyzed dataset contain questions and 26.63 % are SQ questions. Based on those results, we decide to group messages into two main categories. The first category is the query email messages. Query email messages are messages that contain one or more question and require answer. The second category is declarative and informative messages. Declarative and informative email messages are messages that contain declarative sentences which do not require answers or answers can be created automatically. Message categorization of the analyzed dataset is shown in Fig.1.

As shown in Fig.1, In the query email messages category, three types of email messages are identified: Polar question email messages (yes/no questions), Factual question email messages and the combination of those two types.

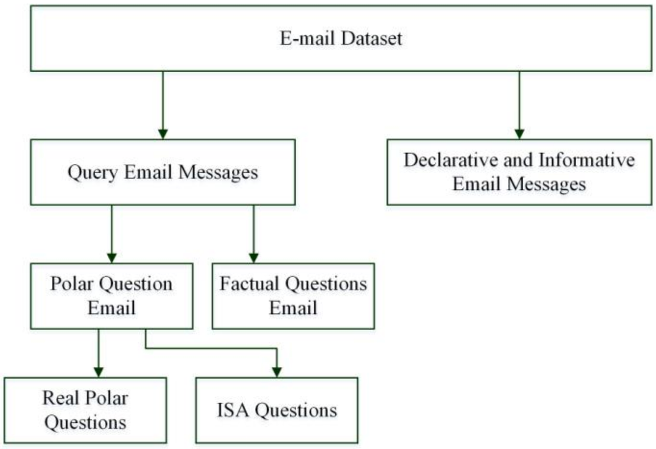

Fig. 1. Email messages categorization

A polar question email messages are messages that contain one or more polar questions. Polar question is a question which has only two possible responses: a” yes” which is an affirmative response or a ”no” which is a negative) response [21]. The example of an email message with multiple polar questions is illustrated below:

Hi John,

Thanks again for the meeting, it was really lovely to meet you. Now, hopefully, you’ve got a bit of insight on what we do too after Mary presentation. Are you still interested in collaboration, and can I place you as a collaborator on my application for the Academy? Thank you!

Best wishes,

Factual question messages are messages that contain one or more Wh-questions. Wh-questions are questions which start with a question-asking word, either a Wh-word (what, when, where, which, who, whose, why) or questions with the word how [22]. The example of Factual question email message type is shown below:

Dear John,

To how many people would be talking about the Conference?

How much is the fee you are offering to Anna? Many thanks.

After initial analysis of the email messages using the Stanford Parser [20], we identified that in the case of Indirect Speech Act (ISA), especially when Factual (Wh) questions are asked in the form of a Polar question, the parser classifies those messages as Polar question emails. According to Yule [23], ISA is defined as an act when the speaker does not explicitly state the intended meaning of the utterance. It is the hearer’s task to analyze the utterance to understand its meaning. An example of an ISA email message:

Dear Prof,

Please, could you send to me date and time of your arrival? Thank you!

Best wishes

3.2. Model Architecture

The main objective of our model is to help users to overcome email overload challenges and minimize user efforts in the email answering process. Our model needs to enable automatic and semi-automatic response generation to the query email messages category. The first step in our approach is to categorize email message. Using the Stanford parser [20], we grouped email body text sentences into main categories (Query email messages, and Declarative and informative email messages), using the methodology defined in the previous paragraph. Depending on the email message category, our system decides how to create email response. Fig.2 depicts workflow of the proposed system for answer generation.

As shown in Fig.2, our system contains two main modules. The first module gathers all unread email messages from a user's Gmail account and parses those messages using Stanford NLP. After parsing email message body texts, the system does the extraction of questions from the email message body text and passes questions and email text to the second module. The second module is a query module. In this module, we will introduce *three-query layers:*

The first layer detects if the email message body text consists of SQ or SBARQ question types. If not, the layer will present the user with an email message and finish processing. If email body text consists of an SQ or SBARQ type of questions layer will pass all information to the second query layer.

The second query layer detects SQ question types. If the message does not have an SQ question, the system will create a custom response template for an SBARQ question type and ask the user to fill the gaps in the recommended response template (template generation will be explained in the following text). If an SQ question type is detected, all questions are passed to the third query layer.

In the third query layer, the system will check if the passed SQ question is a real SQ question or an ISA type of question. For ISA type question detection, we will use a Convolutional Neural Network (CNN) for multi label text categorization. If an ISA question type is detected, the system will present the user with a question and wait for a user action (yes/no) answer. After the user action, the system will pass the question and user action indicator to the AIML chatbot. Depending on the value of the user action indicator, the chatbot will generate *yes* or *no* type of answer and present the user with an answer. The system allows the user to edit the generated answer. If query layer detects that it is a real SQ type of question, the question will be passed directly to the AIML chatbot for automatic answer generation.

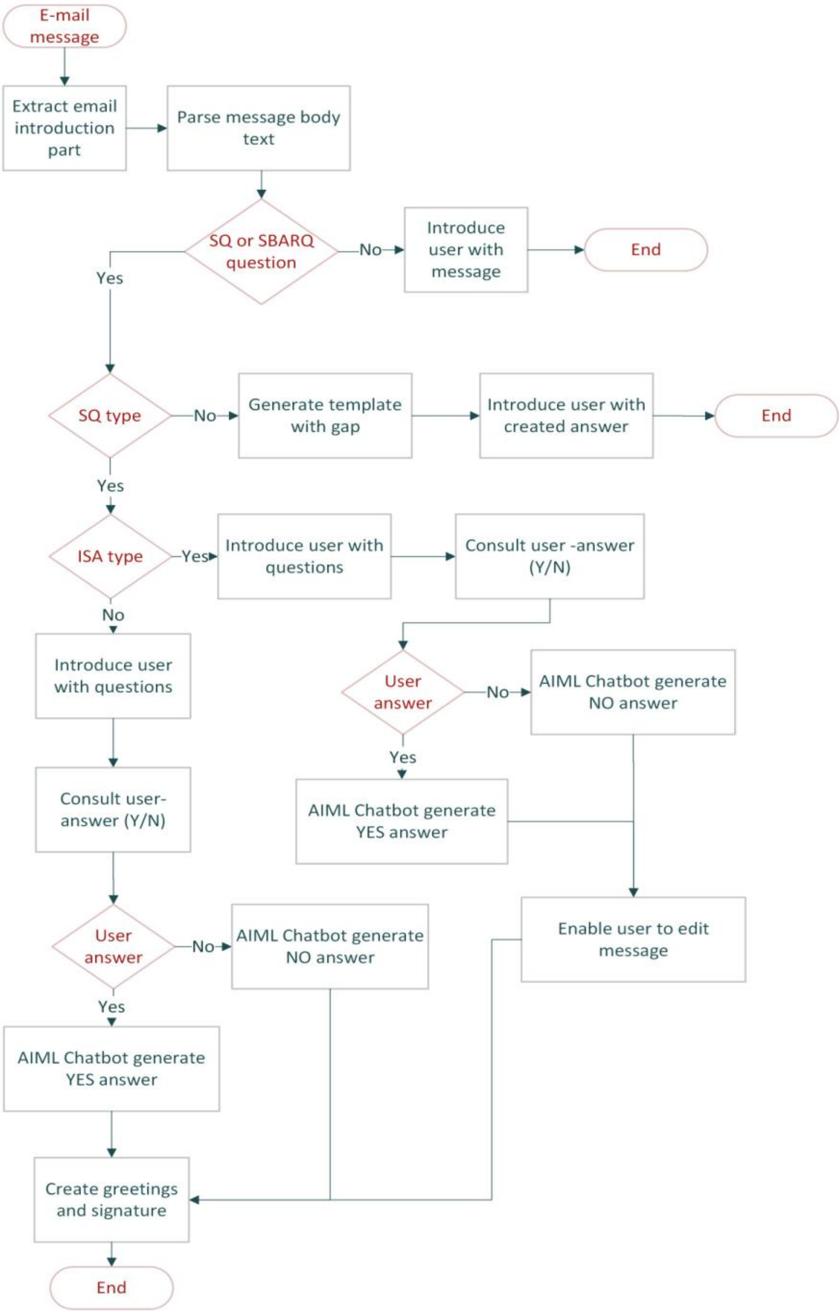

Fig. 2. Email answer generation – Workflow

3.3. Template Generation for the SBARQ Question Type

For the template generation, we proposed the use of LSTM seq2seq NN model as is described in [2]. LSTM is a special kind of RNN. Seq2seq is an NN model that uses one LSTM to read the input sequence encoder, for every time step and the result is a vector. It then uses another LSTM to extract the output sequence from that vector decoder (Fig.3) [24]. The goal of LSTM is to estimate the conditional probability $p(y_1,...,y_{T_0}|x_1,...,x_T)$, where $(x_1,...,x_T)$ is an input sequence (words) and $(y_1,...,y_{T_0})$ is its corresponding output sequence (words) whose length T^0 may differ from T. The LSTM computes this conditional probability by first obtaining the fixed dimensional representation v of the input sequence $(x_1,...,x_T)$ given by the last hidden state of the LSTM, and then computing the probability of $y_1,...,y_{T_0}$ with a standard LSTM-LM formulation whose initial hidden state is set to the representation v of $x_1,...,x_T$:

$$p(y_1,...,y_{T_0}|x_1,...,x_T) = \prod_{t=1}^{T_0} p(y_t|v,y_1,...,y_{t-1}) \qquad (1)$$

In this equation, each $p(y_t|v,y_1,...,y_{t-1})$ distribution is represented with a SoftMax over all the words in the vocabulary. The overall scheme is outlined in Fig.3, where the shown LSTM computes the representation of $A,B,C< EOS >$ (sequences of the input data) and then uses this representation to compute

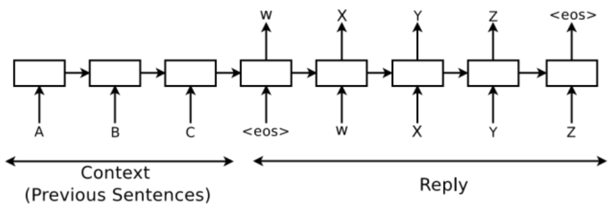

Fig. 3. seq2seq model [24]

the probability of $W, X, Y, Z < EOS >$ (sequences of the output data).

When the system passes the SBARQ question to the NN model, it will create an automatic answer. Using a Stanford NLP named entity recognition, the system will recognize and remove all named entities from the generated answer and present the user with a template. The system will enable to user to fill the gaps made in the named entities extraction process. The proposed PER functionality is not implemented at the time of writing this paper since we were unable to collect enough personal messages to train the proposed NN model.

3.4. ISA Question Type Categorization

For ISA question detection, we use CNN model for multi-label text categorization. Convolutional Neural Networks (CNN) are used for multi-label text classification tasks. CNN is a category of NN that has been proven very effective in areas such as image recognition and classification [25]. CNNs have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self-driving cars [26]. CNNs are a type of feedforward artificial NNs in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex [25, 27]. CNNs exploit spatially local correlation by enforcing a local connectivity pattern between neurons of adjacent layers as shown in the Fig.4.

A feature map is obtained by repeated application of a function across subregions of the entire image, in other words, by convolution of the input image with a linear filter, adding a bias term and then, applying a nonlinear function. If we denote the k-th feature map at a given layer as h^k, whose filters are determined by the weights W^k and bias b_k, then the feature map h^k is obtained as follows (for *tanh* non-linearity's):

$$h^k_{ij} = (W^k * x)_{ij} + b_k \quad (2)$$

For training the NN model we use a dataset of 1000 human detected ISA type questions. We train a model using batched stochastic gradient descent (SGD), which is a standard choice for CNNs. SGD is a stochastic approximation of the gradient

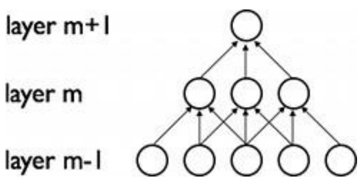

Fig. 4. A simple illustration of the CNN

descent optimization method for minimizing an objective function that is written as a sum of differentiable functions [28]. We randomly select batches of data from the training set (stochastic gradient descent). Tuning the batch size is one of the aspects of getting training right - if a batch size is too small, then there will be a lot of variance within a batch. If a batch size too large, the model will run out of memory, or training will progress too slow. Epoch is defined as one pass through the training data. There are multiple batches in each epoch. The learning rates (LR) is directly connected with the batch size. In CNN, every network layer acts as a detection filter for the presence of specific features or patterns present in the original data [5].

We define training parameters as suggested in Sneiders [4]. The model has the following parameters: num. epochs= 1, batch size=37, num. filters=32, filter sizes= 3,4,5, embedding dimension=200 and drop out = 0.5. By running prediction on the test dataset, we achieved 84.02% accuracy in ISA question type classification. Examples of the email categorization will be shown in the fourth section of this paper.

3.5. AIML-based Chatbot for Semi-Automatic Response Creation

The key novelty presented in this paper is semi-automatic email response generation powered by the Artificial Intelligence Markup Language (AIML). AIML is an XML compliant language for authoring Questioning-Answering agent [29]. We leveraged on an existing implementation through Program AB; a Java programming language interpreter for AIML [30].

Functionally, an AIML category defines a question-answer pair which is represented with the category tags denoted as < category >< /category >. In the category tags are the pattern, which stores the expected input from the users and template tags, which contains the automatic response for the respective input. A simple AIML category is illustrated in Listing 1.

Listing 1- A simple AIML category

```
<category >
<pattern >Hello * <pattern >
<template >I am fine </ template >
</category >
```

According to the presented categorization of the emails, the email dataset comprises the polar and non-polar questions [31]; however, our focus was on the polar questions. Polar questions are generally called yes/no questions and they are usually defined by a “tensed auxiliary clause-initially” [32]. A tensed auxiliary clause-initially is formed by transforming an indicative-mood statement to an interrogative-mood statement, generally by positional transformation of the subject-(auxiliary verb) order in the indicative-mood statement [33], for instance, “I can” to “Can you ”. Auxiliary verbs (AV) are also called the “helping verbs”. Though they do not introduce new semantic content in a clause, they express tenses, moods, voices or modality. According to Kies [34], AV can be grouped into Table 3:

Table 3. An illustration of the mood in sentences

	Indicative-mood state	Interrogative-mood statement
Word order	{*S*| *Av*| *Mv*}	{*Av*| *S*| *Mv*}
Example	I can bring the letter	Can you bring the letter
	I am ready to sign the papers on his behalf	Are you ready to sign the papers on his behalf

**S, Av and Mv* are the subject, the auxiliary verb and the main verb of the clause respectively

the modal auxiliaries (can, could, should, may, would), perfect auxiliaries (has, have), progressive auxiliaries (is, are), passive auxiliary (were) and auxiliary support (do). An indicative-mood statement expresses facts in an honest, direct, relevant way while an interrogative-mood statement depicts a desire for information. Table 3. illustrates two examples of such transformations; for instance, by changing the order of the words, I (subject) and can (modal auxiliary verb) in the sentence “I can come with the letter”, the polar question, “Can I come with the letter” is formed. The basic structure for PQ is illustrated below:

PQ = {Av + S + P}, where PQ is Polar Question, Av is

Auxiliary verb, S is the Subject and P, the Predicate (this includes the main verb and all other details that describe what is going on in the question).

Generally, responses to PQ are type-conforming; they usually conform with the constraints set by the grammatical structure of the question [35]. In other words, the responses usually agree with the function of the PQ, since a confirmation (positive or negative) rather than a new piece of information is provided Selting and Couper-Kuhlen [35], for instance, the question “Can you bring the letter” only requires a confirmatory answer (yes or no). Based on this premise and manual examination of our dataset, we crafted syntactic structures or frames to represent patterns that capture the variations in the dataset; of which the auxiliary verb, the subject, root (main) verb and pronouns were key components. However, special attention was given to the pronouns; some assume different forms in question-and-answer mode. For instance, a comprehensive answer to the PQ “Can you give me the letter?” is “Yes, I can give you the letter”; the subjective pronoun (you) changes to a quote while the objective pronoun (me) becomes “you”. Table 3. shows our crafted syntactic frames and examples of e-mails that fit match with them.

Furthermore, as mentioned earlier the answers to polar questions are type-conforming, the successful crafting of syntactic frames for the email dataset paved the way for easy response generation using an *AIML* parser.

The syntactic frames we crafted were logically transformed into AIML categories and implemented using the Java powered AIML interpreter (Program AB). The interpreter was run as a web service and four different implementations were deployed. They are:

Yes: used to provide a positive response

No: used to provide negative response

Short Yes: for short positive response

Typically, the AIML interpreter comprises the knowledge base (a collection of categories) and an input-response pattern matching algorithm. When the AIML interpreter gets an input (PQ), it searches through its knowledge base to determine which *category* has a *pattern* that the input exhibits. If a match is found, the value of the respective *template* tag will be fetched and sent as the output. Special tags (set, map, star) and wildcard characters (ˆ, *) are generally used as variable placeholders in the *pattern* “or/and” *template* tags to capture differing possibilities as explained in [30]. Table 5 shows an example of questions and respective responses generated by our AIML-based chatbot.

Table 4. Syntactic frames for the email dataset along with examples

Syntactic Frames	Original E-mail Sample	Framed E-mail Sample
(SF) + AV + {NP | PN } + (SF) + RVB + (SF) + (PN) + (SF) + (PN) + (SF) One pronoun or noun phrase before the root verb Zero or more pronouns after root verb	Adrian and Lonce, would you be in town on 4 May for this? Dear Sir, can I get another 3 days for leave? Are you free at the night? Dear Sir, will Karthik be subjected to the same grading system as an intern for payment? Has the international jury set up yet? As virtual experiences of extreme thrills become more ever-present, will we be satisfied with mere simulating danger? Hi Adrian, can you send to us the detailed proposal for QAFU pls? Did you book your ticket to Sri Lanka? Do you think they will be available? Can you kindly check this? Dear Sir, would you like our bio interns to start the DNA project immediately after their exams or when term starts. Will websites and virtual spaces continue to make it easier for us to connect with like-minded people and create strong, active, supportive communities.	{Adrian and Lonce}SF , + {would}AV + {you}PN + {be}RVB + {in town on 4 May for this}SF {Dear Sir ,}SF + {Can}AV + {I}PN + {get}RVB + {another 3 days for leave}SF {Are}AV + {you}PN + {free}RVB + {at the night}SF {Dear Sir}SF + {Will}AV + {Karthik}NP +{be subjected}RVB + {to the same grading system as an intern for payment}SF {Has}AV + {the international jury}NP + {set up}RVB + {yet}SF {As virtual experiences of extreme thrills become more ever-present}SF + {will}AV + {we}PN + {be satisfied}RVB + {with mere simulating danger}SF {Hi Adrian,}SF + {Can}AV + {you}PN + {send}RVB + {to}SF + {us} PN + {the detailed proposal for QAFU} + {pls}SF {Did}AV + {you}PN + {book}RVB + {your}PN + {ticket to Sri Lanka?}SF {Do}AV + {you}PN + {think}RVB + {they}PN + {will be available}SF {Can}AV + {you}PN+{kindly}SF + {check}RVB + {this}PN {Dear Sir, }SF + {Would}AV + {you}PN + {like}RVB + {our}PN + {bio interns to start the DNA project immediately after}SF + {their}PN + {exams or when term starts}SF {will}AV + {websites and virtual spaces}NP + {continue to make}RVB + {it}PN + {easier for}SF + {us}PN + {to connect with like-minded people and create strong, active, supportive communities}SF
(SF) + AV + {NP | PN} + SF + PN + (SF) + RVB + (SF) + PN + (SF) + (PN) + (SF) A noun phrase and a pronoun or two pronouns before the root verb Zero or more than one pronoun after the root verb	Could you and your laboratory accept my students Is it possible for me to arrange a meeting sometime this week, for all of you to discuss this exploratory trip together pls? Could you also let me know by return whether she has any allergies or medical issues?	{Could}AV + {you}PN + {and}SF + {your}PN + {laboratory}SF + {accept}RVB + {my}PN + {students}SF {Is}AV + {it}PN + {possible for}SF + {me}PN + {to arrange}RVB + {a meeting sometime this week, for all of}SF + {you}PN + {to discuss this exploratory trip together pls}SF {Could}AV + {you}PN + {also}SF + {let}RVB + {me}PN + {know by return whether}SF + {she}PN + {has any allergies or medical issues}SF

SF is a sentence fragment; Av is an auxiliary verb; NP is a noun Phrase; PN is Pronoun, RVB is the root verb in the question and | implies ”OR”. Note that an item enclosed within a bracket can either be empty or not. *Short No*: for short negative response.

Table 5. Sample questions and their corresponding AIML-generated responses

Question	Yes	No	Short Yes	Short No
Would you be able to come to Portugal	Yes I would be able to come to Portugal	No I would not be able to come to Portugal	Yes I would	No I would not
Is it possible to stick a screen somewhere for them for that purpose	Yes it is possible to stick a screen somewhere for them for that purpose	No it is not possible to stick a screen somewhere for them for that purpose	Yes it is possible	No it is not possible.
As one of the organizers, would you be able to send me summary of the reports	Yes I would be able to send you summary of the reports	No I would not be able to send you summary of the re-ports	Yes I would	No I would not
Has the international jury set up yet	Yes the international jury has set up yet	No the international jury has not set up yet	Yes the international jury has	No the international jury has not.
Dear Sir, would you like our bio interns to start the DNA project immediately after their exams or when term starts	Yes I would like your bio interns to start the DNA project immediately after their exams or when term starts	No I would not like your bio interns to start the DNA project immediately after their exams or when term starts	Yes I would	No I would not
Will websites and virtual spaces continue to make it easier for us to connect with like-minded people and create strong, active, supportive communities	Yes websites and virtual spaces will continue to make it easier for you to connect with like-minded people and create strong, active, supportive communities	No websites and virtual spaces will not continue to make it easier for you to connect with likeminded people and create strong, active, supportive communities	Yes websites and virtual spaces will	No websites and virtual spaces will not
Could you also let me know by return whether she has any allergies or medical issues?	Yes I could let you know by return whether she has any allergies or medical issues?	No I could not let you know by return whether she has any allergies or medical issues?	Yes I could	No I could not
Dear Sir, Will Karthik be subjected to the same grading system as an intern for payment?	Yes Karthik will be subjected to the same grading system as an intern for payment	No Karthik will not be subjected to the same grading system as an intern for payment	Yes Karthik will	No Karthik will not
Are you available on the morning of Thursday 12th June for the filming?	Yes I am available on the morning of Thursday 12th June for the filming	No I am not available on the morning of Thursday 12th June for the filming	Yes I am	No I am not
Am I right in assuming this is wrong and the correct line-up should be Adrian and Buchanan?	Yes you are right in assuming this is wrong and the correct line-up should be Adrian and Buchanan	No you are not right in assuming this is wrong and the correct line-up should be Adrian and Buchanan	Yes you are	No you are not

4. Implementation and Evaluation

4.1. Model Implementation

The presented model of the system is implemented as a mobile application. The application consists of two modules. The first is a backend module; a web service that queries a user's Gmail inbox and gathers all unread email messages. Then, it processes the gathered messages (parses email message text, recognize PQ, extract recognized questions) and writes results in the database. This process is repeated every two minutes.

The frontend part of our system is a mobile application. This application reads messages processed by the backend module and presents the user with the results. As shown in Fig.5, the user can swipe right to answer *yes* to a request or email question, or swipe left to *decline (No)* to a request or email question. Upon capture of the user action by the mobile application, the question is passed to the respective AIML chatbot and receives an answer.

4.2. Model Evaluation

In our experiment, a group of 40-people installed a mobile application and registered their Gmail account into our system. We analyzed 424 email messages and system generated answers to evaluate model accuracy. Categorization of the obtained dataset gives us results that are shown in Table 6.

Table 6. E-mail categories and number of questions

Tag	Number of categorized emails	Number of questions in categorized emails
SQ	94	234
SQ+SBRQ	95	342
All messages	424	

Analyzing results of the email categorization that are shown in Table 6, we can conclude that our "real life dataset" is similar to the initially analyzed dataset and have 22% of question emails. Also, we can conclude that most of the email messages consist of more than one question. Responses generated by our system are shown in Table 7. Our system can answer semiautomatically to 63.3% of the tested polar questions according to the human evaluation (1 is an acceptable answer 0 is an unacceptable answer). Further, proposed system recognizes ISA and enables the user to edit a generated message and significantly decrease the time of response creation. In 13 % of the cases, answers that are generated by AIML chatbot can be used with a good accuracy

We further carried out a human evaluation of the PER responses. The human evaluator analyzed 94 emails and provided responses to them. These responses were compared with responses generated by the PER using the BLEU score [36]. The BLEU score algorithm compares the N-grams of two text fragments and counts the number of matches, the similarity score of these texts is a function of the number of matches [37]. The BLEU score ranges from 0 to 1; values between 0.5 and 1 reveals a high similarity. We categorized BLEU score results in to 10 categories as is shown in Table 8. Fig.6 shows a plot of the BLEU score categories and number of emails in the respective category BLEU scores of at least 40% (0.4) while 46 have scores of at least 50% (0.5). This implies that the responses from the PER and the human evaluator have high level of similarity.

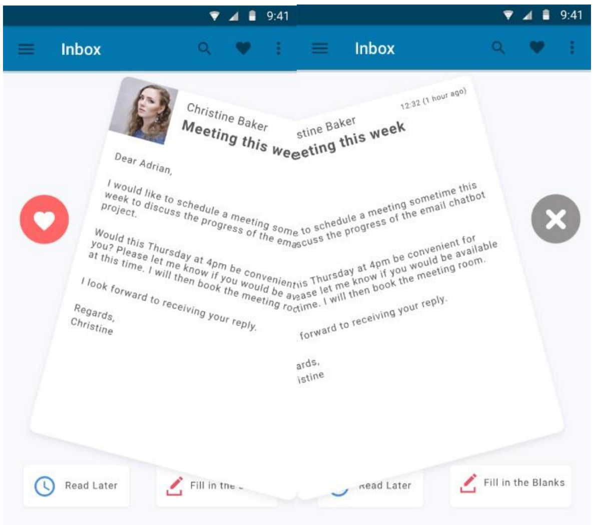

Fig. 5. Screen shot of the mobile app with the Yes/No functionality

5. Conclusion

In this paper, we present a novel system for email categorization and semi-automatic response generation, Automatic Personal Email Responder (PER). The main objective of the Automatic Personal Email Responder system which is presented in this paper, is to help users overcome email overload challenges and minimize user efforts in the email answering process. Our approach was based on the following hypothesis:

H_1: If an email body text contains a question then the user needs to answer to this email - the model generates an automatic or semi-automatic answer. If email body text contains statements without questions, the system will present the user the email.

Follow the hypothesis H_1, the key novelties that are presented in this paper are:

Email categorization approach that distinguishes query and non-query email messages using Natural Language Processing (NLP) and Neural Network (NN) methods,

Implementation of the Artificial Intelligence Markup Language (AIML) - based chatbot for semi-automatic response creation,

Application of template structured database for question answering (AIML) in email answering systems.

Table 7. System responses to questions in e-mails

Processed email text	Extracted polar question	Categorized question		Answers generated by the AIML chatbot
		SQ	ISA	
Lord Vader, Did the dark side reach a conclusion about my visa application?? People keep asking me if I am coming	Did the dark side reach a conclusion about my visa application??	Did the dark side reach a conclusion about my visa application??	-	Yes the dark side did reach a conclusion about your visa application
Dear ***** organizing committee I am ***** in ***** University, Japan. I will present in ***** poster session (submission no.70). Then, I have two questions about facilities and equipment in poster session. Is there a *desk* to put PC and our prototype devices in presentation space? I want to use a desk to explain our system actually. Best regards.	Is there a * desk * to put PC and our prototype devices in presentation space	Is there a * desk * to put PC and our prototype devices in presentation space	-	Yes ther is a desk to put PC and your prototyped devices in presentation space
Interested in your product. Could you send me prices and how to get?	Could you send me prices and how to get?	-	Could you send me prices and how to get?	Yes I could send you prices and how to get
Hi, I thought it would be great for us to sit down and chat. I am free Tuesday and Wednesday. Can you do either of those days?	Can you do either of those days?	Can you do either of those days?	-	Yes I can do either of those days
Why are we sending this? We take security very seriously and we want to keep you in the loop on important actions in your account. We were unable to determine whether you have used this browser or device with your account before. This can happen when you sign in for the first time on a new computer, phone or browser, when you use your browser's incognito or private browsing mode or clear your cookies, or when somebody else is accessing your account. The Google Accounts team	Why are we sending this	-	-	-

Dear *****, How are you today. Could you please confirm if you will be available for our meeting tomorrow	Could you please confirm if you will be available for our meeting tomorrow.	Could you please confirm if you will be available for our meeting tomorrow.	-	Yes I could confirm if I will be available for your meeting tomorrow
Dear Tita, Any progress about our order? Have you receive the payment?	-	-	-	
Hi Rasyidah, Could I follow up on this order. We haven't receive payment for this order.	Could I follow up on this order.	Could I follow up on this order.	-	Yes you could follow up on this order
Dearest Norbert Sorry for all the inconvenience, my staff are working like crazy to make this happen. Would you be available around mid to end of November? We will make it a very important Academic Advisor and Examiner meeting. So your trip will be very serious.	Would you be available around mid to end of November?	Would you be available around mid to end of November?	-	Yes I would be available around mid to end of November
Dear Azrina, Hope all is well. Is there a date for the board meeting in September? Thank you, all the best,	Is there a date for the board meeting in September	Is there a date for the board meeting in September	-	Yes there is a date for the board meeting in September
Dear Adrian, dear Azrina, I hope everything is fine with you. Can you please let me know if you agree with the dates of my visit that I suggested last Friday? They are within the time slot (second half of November) you proposed before. We can talk about the details of the agenda schedule later if you don't have time now. Looking forward to your confirmation. Thank you and best regards -	Can you please let me know if you agree with the dates of my visit that I suggested last Friday?	Can you please let me know if you agree with the dates of my visit that I suggested last Friday?	-	Yes I can let you know if I agree with the dates of your visit that you suggested last Friday

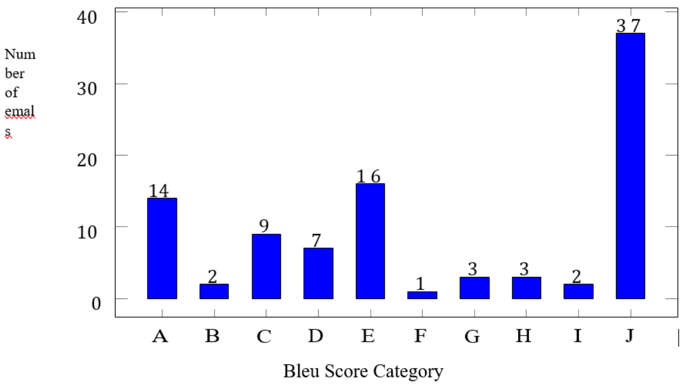

Fig. 6. Summary of BLEU score evaluation of 94 emails

Table 8. BLEU score Categorization

$BlueScore$	$Category.Notation$
$0 - 0.1$	A
$> 0.1\ and <= 0.2$	B
$> 0.2\ and <= 0.3$	C
$> 0.3\ and <= 0.4$	D
$> 0.4\ and <= 0.5$	E
$> 0.5\ and <= 0.6$	F
$> 0.6\ and <= 0.7$	G
$> 0.7\ and <= 0.8$	H
$> 0.8\ and <= 0.9$	I
> 0.9	J

Model evaluation results show that the proposed system can answer to 63.3% question emails correctly from the testing dataset. To increase accuracy of the presented system, we need to create a new NN model for ISA question classification using a bigger dataset. In future works, we will extend the functionality of our system to provide automatic responses to factual questions. We propose implementation of an LSTM seq2seq NN model to create answer templates as responses to the factual questions. Also, we will extend the scope of our system to analyze declarative and informative email messages.

References

1. “Email Statistics Report, 2016-2020,” http://www. radicati.com/wp/wp-content/uploads/2016/01/Email Statistics Report 2016-2020 Executive Summary.pdf, 2016.
2. A. Kannan, K. Kurach, S. Ravi, T. Kaufmann,
3. A. Tomkins, B. Miklos, G. Corrado, L. Lukacs,´ M. Ganea, P. Young et al., “Smart reply: Automated response suggestion for email,” arXiv preprint arXiv:1606.04870, 2016.
4. C. Grevet, D. Choi, D. Kumar, and E. Gilbert, “Overload is overloaded: email in the age of gmail,” in Proceedings
5. of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2014, pp. 793–802.
6. E. Sneiders, “Review of the main approaches to automated email answering.” in WorldCIST (1), 2016, pp. 135–144.
7. L. Deng, D. Yu et al., “Deep learning: methods and applications,” Foundations and Trends R in Signal Processing, vol. 7, no. 3–4, pp. 197–387, 2014.
8. J. M. Ben´ıtez, J. L. Castro, and I. Requena, “Are artificial neural networks black boxes?” IEEE Transactions on neural networks, vol. 8, no. 5, pp. 1156–1164, 1997.
9. G. P. Zhang, “Neural networks for classification: a survey,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 4, pp. 451–462, 2000.
10. R. Bergman, M. Griss, and C. Staelin, “A personal email assistant,” Software Technology Laboratory, HP Laboratories, Palo Alto, 2002.
11. V. S. Ayyadurai, “System and method for contentsensitive automatic reply message generation for text-
12. based asynchronous communications,” Apr. 6 2004, uS Patent 6,718,368.
13. R. Malik, L. V. Subramaniam, and S. Kaushik, “Automatically selecting answer templates to respond to customer emails.” in IJCAI, vol. 7, no. 1659, 2007, p. 3015.

14. R. Mihalcea, C. Corley, C. Strapparava et al., “Corpusbased and knowledge-based measures of text semantic similarity,” in AAAI, vol. 6, 2006, pp. 775–780.
15. L. Kosseim, S. Beauregard, and G. Lapalme, “Using information extraction and natural language generation to answer e-mail,” Data & Knowledge Engineering, vol. 38,
16. no. 1, pp. 85–100, 2001.
17. Y. Watanabe, R. Nishimura, and Y. Okada, “A question answer system based on confirmed knowledge acquired from a mailing list,” Internet Research, vol. 18, no. 2, pp. 165–176, 2008.
18. M. Dredze, T. Brooks, J. Carroll, J. Magarick, J. Blitzer, and F. Pereira, “Intelligent email: Reply and attachment prediction,” in Proceedings of the 13th international conference on Intelligent user interfaces. ACM, 2008, pp. 321–324.
19. J. Martinson, “System and method for eliminating unsolicited junk or spam electronic mail,” Aug. 29 2003, uS Patent App. 10/652,922.
20. A. C. Rothwell, L. D. Jagger, W. R. Dennis, and D. R. Clarke, “Intelligent spam detection system using an updateable neural analysis engine,” Jul. 27 2004, uS Patent 6,769,016.
21. A. Mahendra, “Personified autoresponder.”
22. S. Busemann, S. Schmeier, and R. G. Arens, “Message classification in the call center,” in Proceedings of the sixth conference on Applied natural language processing. Association for Computational Linguistics, 2000, pp. 158–165.
23. A. Bies, M. Ferguson, K. Katz, R. MacIntyre, V. Tredinnick, G. Kim, M. A. Marcinkiewicz, and B. Schasberger, “Bracketing guidelines for treebank ii style penn treebank project,” University of Pennsylvania, vol. 97, p. 100, 1995.
24. “The Stanford Natural Language Processing Group,” https://nlp.stanford.edu/software/lex-parser.shtml/, 2016.
25. “Polar-question dictionary definition — polar-question defined,” http://www.yourdictionary.com/polar-question/.
26. Q. Dictionary, “Questions: wh- questions - English Grammar Today - Cambridge
27. Dictionary,” http://dictionary.cambridge.org/grammar/ british-grammar/questions-and-negative-sentences/ questions-wh-questions/.
28. G. Yule, “Pragmatics/yule g,” Oxford Un. Press, vol. 138,
29. p. 49, 1996.
30. I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in Advances in neural information processing systems, 2014, pp. 3104– 3112.
31. B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, “Learning deep features for scene recognition using places database,” in Advances in neural information processing systems, 2014, pp. 487–495.
32. S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, “Face recognition: A convolutional neural-network approach,” IEEE transactions on neural networks, vol. 8, no. 1, pp. 98–113, 1997.
33. P. Y. Simard, D. Steinkraus, J. C. Platt et al., “Best practices for convolutional neural networks applied to visual document analysis.” in ICDAR, vol. 3, 2003, pp. 958–962.
34. L. Bottou, “Large-scale machine learning with stochastic gradient descent,” in Proceedings of COMPSTAT’2010. Springer, 2010, pp. 177–186.
35. M. N. Perez, F. J. A. Mata, V. M. Z. Rodriguez, and
36. S. Zhang, “Pervasive healthcare monitoring system,” in Ubiquitous Intelligence and Computing and 2015 IEEE
37. 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UICATC-ScalCom), 2015 IEEE 12th Intl Conf on. IEEE, 2015, pp. 1712–1716.
38. A. Foundation, “Aiml - the artificial intelligence markup language,” http://www.alicebot.org/aiml.html.

39. R. Huddleston, “The contrast between interrogatives and questions,” Journal of Linguistics, vol. 30, no. 2, pp. 411– 439, 1994.
40. L. Bailey, “Question particles: Thai, japanese and english,” Linguistica Atlantica, vol. 32, pp. 34–51, 2013.
41. L. Bailey, “Sentential word order and the syntax of question particles,” Newcastle Working Papers in Linguistics, vol. 16, no. 1, pp. 23–43, 2010.
42. D. Kies, “Modern english grammar,” 2003.
43. M. Selting and E. Couper-Kuhlen, Studies in interactional linguistics. John Benjamins Publishing, 2001, vol. 10.
44. K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002, pp. 311–318.
45. C. Corley and R. Mihalcea, “Measuring the semantic similarity of texts,” in Proceedings of the ACL workshop on empirical modeling of semantic equivalence and entailment. Association for Computational Linguistics, 2005, pp. 13–18.

Sasa Arsovski is Program Director of AI and Robotic and Associate Professor at Raffles University. Dr. Sasa Arsovski is a highly recognized AI researcher in the field of computer vision in industrial quality control, object detection, image understanding, and genetic algorithms in text processing for neural network chatbots. With more than twenty years of experience as an AI and IT developer, Dr Sasa Arsovski has received a Ph.D. from the Department of Applied Computer Science and Informatics, Faculty of Technical Science, University Novi Sad, Serbia. Since 2004 he has been working as an independent associate for the development of ICT in the Guarantee Fund of AP Vojvodina, as a visiting professor at multiple universities (inc. International University in Novi Pazar, Serbia and University Novi Sad, Serbia). He was also the Lab Vice Director and AI researcher at the Imagineering Institute in Johor.

Adrian David Cheok is an Australian electrical engineer and a professor at iUniversity, Tokyo, Japan. He is among the faculty of Ducere Business School in Prahran, Victoria and is a visiting professor at the University of Novi Sad in Serbia. He is also the director of the Imagineering Institute in Malaysia, the Mixed Reality Lab in Singapore and the CEO of Nikola Tesla Technologies Corporation in Malaysia. Until 2020 Cheok was on the organizing committee of the Love and Sex with Robots conference.

Muniru Idris Oladele is Data Scientist at Classy Advertising, LDA, Idris is a technology enthusiast who strives to learn and understand key functionalities in technological innovations and inventions in order to simplify their applicability for everyday use. Trainer, a R&D enthusiast and programmer with expertise in Artificial Intelligence (AI), Human-Computer Interaction (HCI), data analytics and anticipatory user design (AUD).

Velibor Premcevski (M’94) received the B.S. degree in information technology from University of Novi Sad, Technical faculty “Mihajlo Pupin” Serbia, in 2017 and the M.S. degree in information technology from University of Novi Sad, Technical faculty “Mihajlo Pupin” Serbia, in 2019. He is currently pursuing the Ph.D. degree in computer science at University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia.

From 2018 to 2020, he was a Teaching Associate on Technical faculty “Mihajlo Pupin”, Serbia. His research interest includes the development of application software solutions in various fields as financial, medical, entertainment. From 2020 to now he is Teaching Assistant on Technical faculty “Mihajlo Pupin”, Serbia. Member of the organizing committee of the Applied Internet and Information Technologies International conference AIIT 2019, 2020, 2021.

Branko Markoski is an full professor within the, Techical faculty ‘’M. Pupin’’, in Zrenjanin, University of Novi Sad, Republic of Serbia since 2019. He received his B.Sc. (Dipl. Ing.) degree in 1994, M.Sc. (Magister) degree in 2000 and Ph.D. degree in 2007, all in Computer Science from the University of Novi Sad, Republic of Serbia , Faculty of Technical Sciences. His is author of then more two hundred papers published in journal and different conferences proceedings and participated in more than 20 projects. He has published five books. His research interests include cyber security, Low level programming, system engineering, software architectures and context-aware computing.

Received: November 01, 2021; Accepted: March 01, 2022.