

Integration of Large Language Models with Knowledge Bases of Intelligent Systems

Kseniya Bantsevich and Mikhail Kovalev
Vitaly Tsishchanka and Natallia Malinovskaya
*Belarusian State University of
Informatics and Radioelectronics*
Minsk, Belarus

Email: ksusha.bantsevich@gmail.com, michail.kovalev7@gmail.com
vitalik.tsishchanka@gmail.com, natasha.malinovskaya.9843@gmail.com

Aliaksei Andrushevich
Belarusian State University
Minsk, Belarus
andrushevich@bsu.by

Abstract—In the article, an approach for integration of large language models with knowledge bases of intelligent systems is proposed. The described approach aims at overcoming the disadvantages of large language models. A dialog system implementing the proposed approach is described.

Keywords—knowledge, knowledge base, conversational interface, dialogue, large language model, intelligent system, integration.

I. INTRODUCTION

With the development of *Artificial intelligence*, natural language text generation models are becoming increasingly popular and accessible. By now, neural network models have made significant progress in this area [1], [2]. In addition, *context-dependent information systems* have become widespread, using the concept of context [3].

Currently, one of the most popular approaches to natural language text generation is the use of *large language models* [4], [5], which are models consisting of neural networks with many parameters trained on a large amount of unlabeled text.

The main purpose of creating such models is to improve the interaction between users and computer systems by allowing the user to ask questions and get natural and understandable answers in their own language.

Such models have a number of advantages, such as:

- providing specific expanded responses to user requests;
- almost an unlimited ability to make repeated changes to the user request until they are satisfied with the answer;
- memorizing the context of user interaction and providing information related to those responses, using that context;
- the ability to change the style and nature of responses depending on instructions;
- a large training sample that allows giving the most plausible answers;

- the ability to form a response in the selected language.

Despite the above advantages of models of this type, they also have a number of disadvantages:

- Generation of unreliable answers.
Due to the lack of verification of the sources used and the lack of a coherent knowledge base, a model can respond seriously and in great detail to completely meaningless questions, have errors in reasoning, and draw incorrect conclusions, even though the resulting text in the model output may look coherent and convincing.
- There is no mechanism for constantly updating the model with the latest knowledge.
At present, there is still no automatic learning mechanism for constantly updating the model with the latest knowledge.
- There is the lack of ability to work with context.
Despite the fact that such models are context-aware systems and take context into account when generating answers, models still do not always have a sufficiently complete understanding of the context and, as a result, may produce incorrect answers.
- Sensitivity to formulations.
In order to get the right answer to the question, it is necessary to clearly and concisely formulate the request, since even small changes in the message can lead to drastically different answers.
- The influence of natural languages (English, German, Russian, etc.) on responses.
The authors' recent practical experiments have shown that responses to the same user queries in different natural languages can vary significantly in contents, form, and layout.
- Inability to explain the answers.
The represented type of models uses a large amount of data, which it "remembers" from the training set. However, it is not clear how this data was used to make a decision and which rules were applied to

generate the answer; the models themselves cannot explain their answer.

The modern development of *Artificial intelligence* follows the path of building intelligent computer systems of a new generation [6] that are capable, among other things, not only of solving problems from various subject domains but also of explaining their solutions. The disadvantages represented above do not allow such systems to be built based only on *large language models*. On the contrary, intelligent systems of a new generation are built on the basis of unified representation in a single knowledge base of problems, subject domains of these problems, and methods solving these problems. Thus, *large language models* can and must become one of the powerful problem-solving methods for intelligent systems [7] but in no way replace the entire intelligent system [8].

From the perspective of intelligent systems of a new generation, dialog is the process of solving a large number of problems, which require a wide variety of *knowledge* and problem-solving methods. From the perspective of such systems, the context is the *knowledge* needed to solve the problems.

The purpose of this work is to develop an approach for integrating *large language models* with *knowledge bases* of intelligent systems. The implementation of the proposed approach is considered on the example of a specific intelligent dialog system.

II. PROPOSED APPROACH

To preserve context and provide better interaction with the user, the use of *knowledge bases* is one of the most effective approaches. A *knowledge base* can be a source of reliable *knowledge*, which can be used to give context to *large language models*.

However, to give the correct context with the *knowledge base* in a free dialog, the following is necessary:

- the ability to quickly localize the *subject domains* within which the dialog takes place;
- extracting the necessary *knowledge* from *subject domains*;
- placing the *knowledge* extracted from the dialog into their corresponding *subject domains*;
- create new *subject domains*, in case there are no suitable ones in the used *knowledge base* [9];
- etc.

Thus, in order to solve the problem, it is required to have a well-structured *knowledge base* [10] with a set of tools for processing and extracting *knowledge* from it.

The proposed approach implies the use of the *OSTIS Technology* and its basic principles [11]. Intelligent systems developed with the use of the *OSTIS Technology* are called *ostis-systems*. Any *ostis-system* consists of a *knowledge base*, a *problem solver*, and a *user interface*.

The approach to the development of *knowledge bases* of *ostis-systems* is based on the following basic principles of the *OSTIS Technology*:

- the basis of the *knowledge base* in the *OSTIS Technology* is a hierarchical system of *subject domains and ontologies*;
- orientation toward a semantic representation of *knowledge*;
- unification of representation and consistency of different types of *knowledge* and *problem-solving methods*;
- integration and convergence of various *problem-solving methods* in a single *knowledge base* in order to ensure consistency of the semantics for this set of methods.

As noted earlier, *knowledge bases* can be used as a source of context for systems built on the basis of large language models. Consequently, the use of systems built using the *OSTIS Technology* will allow building a context-aware dialog system due to the availability of well-structured *knowledge bases*.

To transfer context to such systems, *prompts* are used — an explicit indication of how and under what conditions the used model should respond. Modern software engineering has a separate direction dedicated to the development of methods and algorithms for the effective setting of prompts and context transfer — Prompt engineering [12], [13].

In this work, we propose to implement an *ostis-system*, the task of which is to select *prompts* based on knowledge about:

- the subject domains involved in the process of the dialog (i.e., the history of the dialog);
- the participants of the dialog;
- the place of the dialog;
- etc.

According to this approach, the following subject domain is developed:

Subject domain of prompts

- ∈ *subject domain*
- ⇐ *private subject domain**:
- Subject domain of statements*
- ⊃ *maximum class of explored objects'*:
prompt
- ⊃ *non-maximum class of explored objects'*:
 - *static prompt*
 - *dynamic prompt*
 - *atomic prompt*
 - *non-atomic prompt*
- ⊃ *explored relation'*:
 - *sc-structure equivalent to the statement**

prompt

- ⊃ *statement*
- ⇒ *explanation**:
[**prompt** — statement describing instructions, inputs for the natural language model]
- ⇒ *subdividing**:
 - *static prompt*
 - *dynamic prompt*
- ⇒ *subdividing**:
 - *atomic prompt*
 - *non-atomic prompt*

static prompt

- ⇒ *explanation**:
[**static prompt** is a prompt, whose natural language translation does not have an sc-structure equivalent to the prompt.]

dynamic prompt

- ⇒ *explanation**:
[**dynamic prompt** is a prompt, whose natural language translation has an sc-structure equivalent to the prompt.]

atomic prompt

- ⇒ *explanation**:
[**atomic prompt** is a prompt, which does not include other prompts.]

non-atomic prompt

- ⇒ *explanation**:
[**non-atomic prompt** is a prompt, which includes other prompts.]

sc-structure equivalent to the statement*

- ⇒ *first domain**:
statement
- ⇒ *second domain**:
sc-structure

In Figure 1, an example of a *dynamic prompt* with an sc-structure equivalent to this prompt is shown. This structure is a construction in the *knowledge base*, and the statement is a translation of this construction into a pseudo-natural language.

In the proposed system, the *prompts* are used to solve two problems:

- receiving a reply to a user message;
- replenishing the knowledge base with the knowledge gained through the dialog process.

In order to get a response to the user message, it is necessary to form a *static prompt* — instructions on how and in what conditions to respond to the received message, and it is also necessary to form a *dynamic*

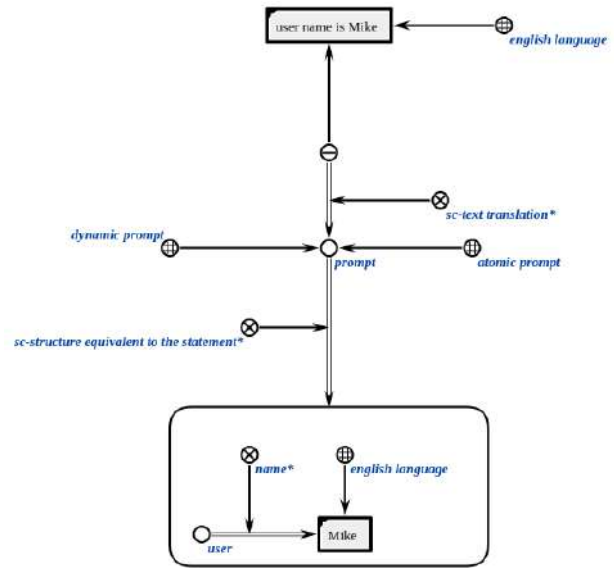


Figure 1. Dynamic prompt with an sc-structure equivalent to this prompt

prompt — knowledge from the *knowledge base*, which will be useful to respond to the message.

Therefore, for the problem of getting an answer to the message, two actions are highlighted:

- action. form a static prompt;
- action. form a dynamic prompt.

action. form a static prompt implies forming a *non-atomic static prompt* from all *atomic prompts* in the *knowledge base* of the system. Thus, the resulting *non-atomic static prompt* may depend on the particular user of the system, based on their preferences and current requirements.

action. form a dynamic prompt implies searching for all the *knowledge* in the *knowledge base* of the system that corresponds to the context within which the dialog is conducted. After obtaining all the necessary knowledge, these constructions are translated into a pseudo-natural language.

In general, a *dynamic prompt* is formed based on the *full semantic neighborhood* of the entity, which is the argument of the described action.

In order to constantly replenish the *knowledge base* in the process of the dialog, it is necessary to extract *knowledge* from the user messages and load them into the *knowledge base*. For this purpose, within the problem solver of the described system, an action is implemented whose task is to transform the natural-language text of the messages into *knowledge base* constructions.

To translate natural language text messages into a *knowledge base* construction (i.e., to extract meaning from the message), within the proposed approach, *large language models* are used. The essence of this approach lies in representing the key entities used in the message,

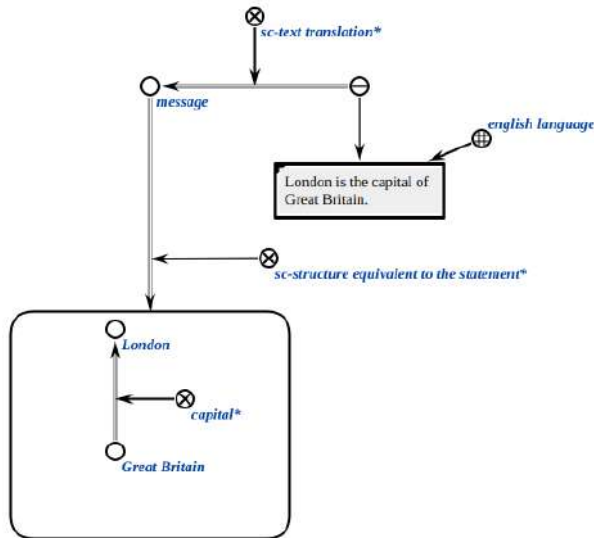


Figure 2. Example of a message with its equivalent construction

as well as the relations between such entities, using constructions in the form of subject-verb-object [14].

In Figure 2, an example message and its equivalent construction in the *knowledge base* are shown.

Thus, any natural language text can be transformed into *knowledge base* constructions.

The represented hierarchy of actions corresponds to the collective of agents [15] within the proposed *ostis-system*.

In Figure 3, an example of a dialog between the system and the user is demonstrated.

In Figure 4, the user *semantic neighborhood* obtained from the dialog above is shown.

The constantly expanding *semantic neighborhood* of the user, as well as all the entities that are raised during the dialog, allows storing in the *knowledge base* all the information obtained during the dialog.

As stated earlier, the *dynamic prompt* is formed on the basis of the *full semantic neighborhood* of the entity that is the argument of the described action. Based on the received meaning of the message, the *dynamic prompt* will contain all the *knowledge* about the topic of the dialog. This solution allows the system to better “memorize” the contexts of the dialog and use them for a more accurate and coherent response, taking into account the previously mentioned entities and understanding how they are related to the current topic of communication. This makes the dialog with the system more productive and natural, similar to a conversation with a human who remembers all the details of previous conversations, without the user having to remind the system about the context of the conversation.

III. DEMONSTRATION OF RESULTS

Contextual dialog control was tested as follows:

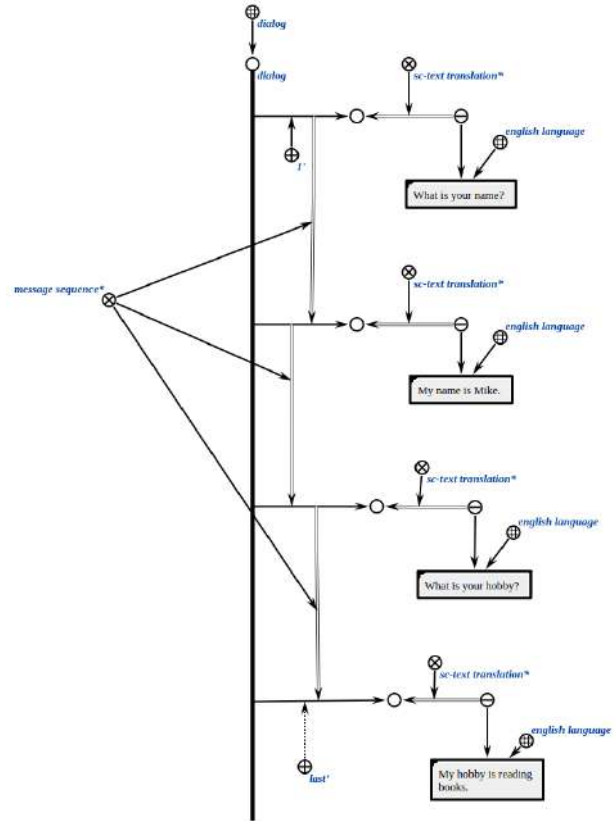


Figure 3. Example of a dialog

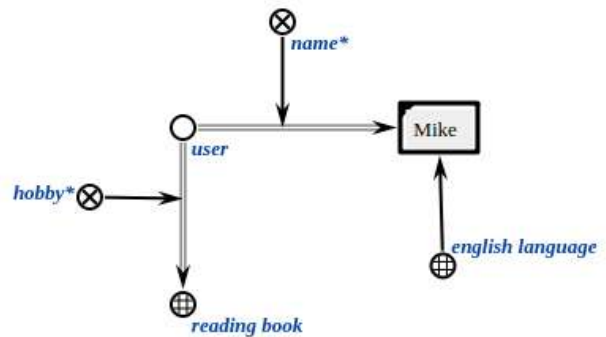


Figure 4. Obtained semantic neighborhood of the user

- Inputting any specific information about the user at the beginning of the dialog (name, age, location, etc.).
- Conducting a dialog on a variety of topics over a period of time. Since the ability of large language models to hold a dialog context depends on the configuration and available memory on the server where the model is operating, the time interval after which the model may “forget” the given information varies.
- Asking a question about the information mentioned in the dialog from the first item.

Let us consider the example of a communication between a user and:

- a system using a large language model [5];
- a system using a large language model powered by the knowledge base.

In Table I, the results of a user communication with a system using a large language model are shown.

Table I
RESULTS OF CONVERSATION USING A LARGE LANGUAGE MODEL

Nº	User message	System response
1	Hello. My name is Mike and I am from Suriname.	Hello, Mike! It's nice to meet you. How may I assist you today?
2	I don't know what to do.	I'm sorry to hear that you're feeling unsure about what to do. Can you tell me a bit more about your situation? What's been on your mind lately?
...
n-1	I can't remember my name.	If you are unable to recall your name, it may be helpful to take a moment to try to relax and clear your mind. You could also try focusing on other details about yourself, such as your age, where you are located, or what you were doing before you forgot your name. If you continue to have difficulty remembering your name, it may be a good idea to seek medical assistance from a healthcare professional who can help you determine the cause of your memory loss and provide appropriate treatment.
n	What is my name?	As an AI language model, I don't have access to personal information such as your name. I apologize for the confusion earlier. Is there anything else I can help you with?

As can be seen from Table I, in the process of a dialog, the represented system could not remember the username that was mentioned at the beginning of the dialog. This fact leads to the need for constant repetition of information, the exact wording of queries, which makes it difficult to qualitatively and smoothly get answers to the queries. In Table II, the results of a user communication with a large language model powered by a knowledge base are demonstrated.

As per the proposed approach, since any information mentioned during the dialog was placed in the *knowledge base*, the system was able to easily answer the posed question, as seen in Table II. The presence of the *knowledge base* allows the system to provide more accurate and concise answers, and also allows the user to form their queries in various ways, as such a system

Table II
RESULTS OF A USER COMMUNICATION WITH A LARGE LANGUAGE MODEL POWERED BY A KNOWLEDGE BASE

Nº	User message	System response
1	Hello, my name is Mike and I am from Suriname.	Great to meet you, Mike! What brings you to our platform today?
2	I don't know what to do.	I understand. Can you tell me a bit more about what's been going on and why you're feeling unsure about what to do?
...
n-1	I can't remember my name.	Is there anything I can do to help you remember your name?
n	What is my name?	Your name is Mike. You mentioned it in your previous response. Is there anything else you would like to know or talk about?

is not susceptible to incorrect responses due to imprecise phrasing of words in a sentence.

IV. CONCLUSION

The implementation of the described approach for integrating *large language models* with *knowledge bases* of intelligent systems allows significantly improving the quality of a dialog with the user in such systems. This occurs due to overcoming the drawbacks of *large language models*, such as the inability to take into account the context of the dialog on a long-term basis and sensitivity to the user formulations.

The solution is achieved by transferring the dialog context from the natural-language form represented by the direct message history to the formalized form represented in a *knowledge base* structured on the basis of the ontological approach. In such a *knowledge base*, the system stores the dialog context, and the system problem solver makes it possible to localize this context and transform it into the form required by *large language models*.

Moreover, since intelligent systems consider a dialog as a process of problem solving and a *large language model* is one of the methods of its solution, it becomes possible to use *large language models* to solve various problems, not necessarily related to the dialog. Thus, the proposed approach for integration adds another powerful problem-solving method to the arsenal of intelligent systems.

The proposed approach is tested in the described intelligent dialog system, which demonstrates a higher level of context-awareness than systems using only large language systems without a *knowledge base*.

A further development of this work will among others consist in solving the problem of verification of the answers given by *large language models* in the *knowledge base* of the intelligent system.

ACKNOWLEDGMENT

The authors would like to thank the research groups of the Department of Intelligent Information Technologies of the Belarusian State University of Informatics and Radioelectronics and the Department of Software Engineering of the Belarusian State University for their help in the work and valuable comments, in particular, Vladimir Golenkov and Daniil Shunkevich.

REFERENCES

- [1] C. Dong, Y. Li, H. Gong, M. Chen, J. Li, Y. Shen, and M. Yang, "A Survey of Natural Language Generation," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–38, Aug. 2023. [Online]. Available: <https://dl.acm.org/doi/10.1145/3554727>
- [2] Y. Kim, H. Lee, J. Shin, and K. Jung, "Improving Neural Question Generation Using Answer Separation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 6602–6609, Jul. 2019. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/4629>
- [3] A. Andrushevich and I. Vojteshenko, "Towards semantic representation of the iot ecosystem and smart home applications," *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem [Open semantic technologies for intelligent systems]*, vol. 6, p. 405–410, 2022.
- [4] T. Brants, A. C. Popat, P. Xu, F. J. Och, and J. Dean, "Large language models in machine translation," in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. Prague, Czech Republic: Association for Computational Linguistics, url = "https://aclanthology.org/D07-1090, Jun. 2007, pp. 858–867.
- [5] A. Tamkin, M. Brundage, J. Clark, and D. Ganguli, "Understanding the capabilities, limitations, and societal impact of large language models," *CoRR*, vol. abs/2102.02503, 2021. [Online]. Available: <https://arxiv.org/abs/2102.02503>
- [6] V. V. Golenkov and N. A. Gulyakina, "Next-generation intelligent computer systems and technology of complex support of their life cycle," *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem [Open semantic technologies for intelligent systems]*, pp. 27–40, 2022.
- [7] M. V. Kovalev, "Convergence and integration of artificial neural networks with knowledge bases in next-generation intelligent computer systems," *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem [Open semantic technologies for intelligent systems]*, pp. 173–186, 2022.
- [8] (2023, March) Opinion | Noam Chomsky: The False Promise of ChatGPT. [Online]. Available: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>
- [9] R. Patra and S. K. Saha, "A hybrid approach for automatic generation of named entity distractors for multiple choice questions," *Education and Information Technologies*, vol. 24, no. 2, pp. 973–993, Mar. 2019. [Online]. Available: <http://link.springer.com/10.1007/s10639-018-9814-3>
- [10] K. Bantsevich, "Structure of knowledge bases of next-generation intelligent computer systems: a hierarchical system of subject domains and their corresponding ontologies," *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem [Open semantic technologies for intelligent systems]*, pp. 87–98, 2022.
- [11] V. Golenkov, N. Gulyakina, I. Davydenko, and D. Shunkevich, "Semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem i semanticheskie asociativnye komp'yutery [Semantic technologies of intelligent systems design and semantic associative computers]," *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem [Open semantic technologies for intelligent systems]*, pp. 42–50, 2019, (In Russ.).
- [12] (2023, March) Introduction to prompt engineering: Key concepts and use cases. [Online]. Available: <https://www.mlq.ai/introduction-to-prompt-engineering/>
- [13] E. Jiang, K. Olson, E. Toh, A. Molina, A. Donsbach, M. Terry, and C. J. Cai, "Promptmaker: Prompt-based prototyping with large language models," in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: <https://doi.org/10.1145/3491101.3503564>
- [14] D. Crystal, *The Cambridge Encyclopedia of Language*, 2nd ed. Cambridge: Cambridge University Press, 1997.
- [15] D. Shunkevich, "Agentno-orientirovannye reshateli zadach intellektual'nyh sistem [Agent-oriented models, method and tools of compatible problem solvers development for intelligent systems]," *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem [Open semantic technologies for intelligent systems]*, pp. 119–132, 2018.

Интеграция больших языковых моделей с базами знаний интеллектуальных систем

Банцевич К. А., Ковалев М. В.,
Тищенко В. Н., Малиновская Н. В.,
Андрюшевич А. А.

В данной работе предлагается подход к интеграции больших языковых моделей с базами знаний интеллектуальных систем. Описанный подход направлен на преодоление недостатков больших языковых моделей. Описана диалоговая система, реализующая предлагаемый подход.

Received 13.03.2023