[7] Androutsopoulos, G. Lampouras, D. Galanis Generating Natural Language Descriptions from OWL Ontologies: the NaturalOWL System. *Journal of Artificial Intelligence Research*, 2013, vol. 48, No. 01, pp. 671–715.

[8] A. Gatt, E. Krahmer Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 2018, vol. 61, pp. 65–170.

[9] K. Hu, X. F. Xi, Z. M. Cui, Y. Y. Zhou, Y. J. Qiu Survey of Deep Learning Table-to-Text Generation. *Journal of Frontiers of Computer Science and Technology*, 2022, vol. 16, No. 11, pp. 2487–2504.

[10] G. Claire, S. Anastasia, N. Shashi The WebNLG Challenge: Generating Text from RDF Data. In Proceedings of the 10th International Conference on Natural Language Generation, Santiago de Compostela, Spain, 2017, pp. 124–133.

[11] V. V. Golenkov, N. A. Gulyakina Proekt otkrytoi semanticheskoi tekhnologii komponentnogo proektirovaniya intellektual'nykh sistem. Chast' 1 Printsipy sozdaniya [Project of open semantic technology of component designing of intelligent systems. Part 1 Principles of creation]. *Ontologiya proektirovaniya [Ontology of designing]*, 2014, No. 1, pp. 42–64 (In Russ.).

[12] M. E. Sadouski The structure of next-generation intelligent computer system interfaces. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2022, pp. 199–208.

[13] D. V. Shunkevich Hybrid problem solvers of intelligent computer systems of a new generation. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2022, pp. 119–144.

[14] L. W. Qian, W. Z. Li Ontological Approach for Generating Natural Language Texts from Knowledge Base. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2021, pp. 159–168.

[15] J. Li, H. Hu, X. Zhang, M. Li, L. Li, L. Xu Light Pre-Trained Chinese Language Model for NLP Tasks. In CCF International Conference on Natural Language Processing and Chinese Computing, Minsk, 14 October 2020, Springer, Cham. zhengzhou, 2020, pp. 567–578.

[16] W. Z. Li, L. W. Qian Development of a problem solver for automatic answer verification in the intelligent tutoring systems. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems]*, Minsk, 2021, pp. 169–178.

[17] Y. H. Tseng, L. H. Lee Chinese Open Relation Extraction for Knowledge Acquisition. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26-30 April 2014, pp. 12–16.

[18] K. Papineni, S. Roukos, T. Ward, W. J. Zhu BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), USA, 2002, pp. 311–318.

[19] C. Y. Lin ROUGE: A Package for Automatic Evaluation of Summaries. In Text Summarization Branches Out, Barcelona Spain, 2004, pp. 74–81.

[20] B. D. Trisedya, J. Z. Qi, R. Zhang, W. Wang GTR-LSTM: A Triple Encoder for Sentence Generation from RDF Data. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne Australia, 2018, pp. 1627–1637.

[21] M. Kale, A. Rastogi Text-to-Text Pre-Training for Data-to-Text Tasks. In Proceedings of the 13th International Conference on Natural Language Generation, Dublin Ireland, 2020, pp. 97–102

Онтологический подход к разработке китайско-языкового интерфейса в интеллектуальных системах

Цянь Лунвэй

В статье рассматриваются существующие подходы к приобретению фактографических знаний из текстов естественного языка и генерации текстов естестфенного языка из фрагментов базы знаний (фактографических знаний), которые рассматриваюся как две основные задачи, решаемые естественно-языковыми интерфейсами интеллектуальнных систем в нашей работе. Был проведен анализ проблем, возникающих при разработке естественно-языкового интерфейса интеллектуальнных систем, а также приобретении фактографических знаний из текстов естественного языка и генерации текстов естественного языка из фрагментов базы знаний в настоящее время.

В рамках технологии OSTIS был предложена разработка единой семантической модели естественноязыкового интерфейса интеллектуальнных систем, которые в основном состоят из sc-модели базы знаний лингвистики и sc-модели соответствующих решателей задач для обработки естественного языка. Среди них в sc-модели базы знаний лингвистики позволяется объединение лингвистических знаний на различных уровнях, в sc-модели соответствующих решателей задач позволяется интеграция моделей на основе правил и моделей нейронных сетей для обработки естественного языка. Более того, на основе единой семантической модели естественно-языкового интерфейса был реальзаван китайско-языковой интерфейс ostis-систем и оценен разработанный китайско-языковой интерфейс по трём аспектам. По сравнению с другими системами, разработанный китайско-языковой интерфейс имеет лучшую эффективность.

Received 29.03.2023

Tools for Creating and Maintaining a Knowledge Base by Integrating Wolfram Mathematica System and Nevod Package

Valery B. Taranchuk
Department of Computer Applications and Systems
Belarusian State University
Minsk, Belarus
golen@bsu.by

Vladislav A. Savionok
Department of Software for Information Technologies
Belarusian State University of Informatics and Radioelectronicss
Minsk, Belarus
v.savenok@gmail.com

***Abstract*—One of the outcomes of the current review of the state of knowledge base design and analysis technologies, software and hardware platforms for the implementation of semantically compatible intelligent computer systems is the conclusion about the need to formulate the principles of collective design, development, verification of knowledge bases. Accordingly, it is important not only to formulate and justify the theory, to formalize the requirements, but also to develop tools to represent such formal theory in the form (format) of the knowledge base of the corresponding scientific knowledge portal. It is this concept that is the goal of implementation and development of the OSTIS Ecosystem, which, in particular, is intended to solve problems of convergence and merging of functional properties of systems of different classes; range expansion, organizational and technical unification, realization of coordination of software, computing and telecommunication means; unification of intelligent computer systems. A special place in such unification should be given to the solution of problems of integration of OSTIS Ecosystem tools with computer mathematics systems, especially with computer algebra systems.**

This paper presents an example of integration of the local intelligent computer system based on Nevod library with the knowledge base of Wolfram Mathematica computer algebra system, which can be interpreted as an analogue of the integration of knowledge bases of the corporate OSTISsystem into the OSTIS Ecosystem. Examples of the use of tools to analyze the local knowledge base, its transfer from virtual to real status are presented and explained.

***Keywords*—Semantic analysis, OSTIS technology, Wolfram Mathematica, Wolfram Knowledgebase, Entity, temporal markers, Nevod**

I. Introduction

According to the assessment of the current state of the field of artificial intelligence (AI), given in [1], there is an active development of many different areas, such as formal ontologies, the artificial neural of networks, machine learning, multi-agent systems, etc. However, this activity does not bring an aggregate increase in the level of intelligence of modern intelligent computer systems (ICS). This is due to the current isolation between methods and designing tools in each of the areas of AI. The solution of this problem is seen in the construction of a general formal theory of ICS, and designing a comprehensive technology for their development and life cycle support — the OSTIS technology [1]. This will allow to achieve convergence of all areas of AI through their mutual integration and joint development. It is this concept that is the goal of implementation and development of the OSTIS Ecosystem.

Modern design support frameworks in the field of AI are mainly aimed at the development of highly specialized solutions, which can act as individual components of the ICS. In order to obtain guaranteed compatibility of all developed components, it is necessary to transform these tools into a unified technology for comprehensive design and support of the full life cycle of the ICS. Despite the independent development of the ICS, special attention should also be paid to their external interfaces, since the intercommunication of ICS between each other will be required as part of complex systems for the automation of various human activities. In this direction a number of problems are solved, such as document search in local and global networks In other words, there is a need for unification and convergence of next-generation ICS, along with their components. This will open the way to the design of optimized complexes that include all the necessary AI to solve the tasks at hand. It is important to note that in order to meet the optimization requirements when solving certain tasks and achieve maximum performance, it is necessary to organize the effective interaction of the ICS with the connected information resources. The main actions for solving the key methodological problems, which are the reason for the current state of the field of AI, are also given in [1]. Note that similar problems are solved in the field of computer algebra systems: in their design, development, content update and functionality expansion [2], [3].

Methodological and technical solutions for integration of various types of knowledge implemented in the computer algebra system Wolfram Mathematica (WM), Wolfram Language (WL) are described in this paper.

The software solutions implemented in the Wolfram Knowledgebase are marked and illustrated with examples. From the standpoint of the need of unification of next-generation ICS, integration with the Wolfram Knowledgebase is performed on the general basis, meaning that the same method can be applied to integrate the WM with other knowledge systems, including the OSTIS Metasystem. The examples in this paper demonstrate several methods of integration of various tools implemented in Wolfram Mathematica computer algebra system, and by means of independent library Nevod [4].

II. Creating a thematic block for temporal markers analysis

A. Temporal markers analysis

One of the main directions in the field of text processing is the extraction of their semantic component — semantic analysis. In this direction a number of problems are solved, such as document search in local and global networks, automatic annotation and abstracting, classification and clustering of documents, synthesis of texts and machine translation, text tone analysis, and fact extraction (publications mentioned in [5]).

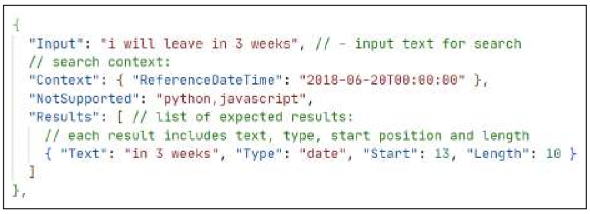
An integral part of the task of extraction of facts and determination of relations between objects is localization in time of the event corresponding to the fact. The information, allowing to localize the event on a time axis, is transferred by means of various in the form and content textual expressions — temporal markers (pointers). The final result of the extraction of temporal markers from the text is their representation and interpretation within the framework of the formal model set in the process of semantic analysis [6].

To solve the problem of extracting temporal references from text the toolkit of one of the leaders in the field of entity recognition Microsoft.Recognizers.Text [7] is widely used.

B. Preparation of data for the thematic block of temporal markers analysis

The MS Recognizers Text (MRT) library provides the ability to recognize entities in texts languages and is widely used in Microsoft products, for example: in predefined templates for LUIS (Language Understanding Intelligent Service), in the platform for creating dialog bots Power Virtual Agents [8], in cognitive language services in Azure cloud infrastructure — NER (Named Entity Recognition). The exact matching means that the identifiers of each extracted element. The information, allowing to localize the event on a time axis, is transferred by means of various.The library is distributed under an open source and free software license from MIT; along with the source code in the repository on GitHub [7] test dataset for different languages are published.

The Microsoft.Recognizers.Text.DateTime module, and in particular its BaseDateExtractor component, is used in MRT to search for markers in the text. This is component corresponds to a test dataset represented in JSON format – the DateExtractor.json file [9]. The dataset contains 143 elements that include absolute and relative dates in different forms, as well as metainformation, which is used to check the correctness of the extraction results. In addition to the various type of linguistic knowledge. A search context, a reference date that indicates the point in time used to translate relative temporal markers into absolute ones, can be attached to the dataset element. An simple example of a test dataset element with comments is shown in Fig. 1.



```
{
  "Input": "i will leave in 3 weeks", // - input text for search
  // search context:
  "Context": { "ReferenceDateTime": "2018-06-20T00:00:00" },
  "NotSupported": "python,javascript",
  "Results": [ // list of expected results:
    // each result includes text, type, start position and length
    { "Text": "in 3 weeks", "Type": "date", "Start": 13, "Length": 10 }
  ]
},
```

Figure 1: Example of a test dataset element.

In [5] the results of comparing the capabilities in temporal pointers extraction of MRT and Nevod library [4], which implements the search method in the text [10], are described. For this purpose two software modules were developed: mMRT and mNevod, which provide search and extraction of temporal markers from text. The exact matching means that the identifiers of each extracted element. Comparative testing of the software modules was performed on the described test dataset using the means of the computer algebra system Wolfram Mathematica to analyze the results.

Input data for mNevod and mMRT modules is Input string. The results of temporal pointer extraction modules are compared with the Results dictionary, which contains the expected position in the text, length and contents of the extracted temporal pointer. The DateExtractor test dataset is used to confirm the functional completeness of the libraries that extract temporary pointers from text.

When checking and tuning fact extraction tools, in particular temporal pointers, an important position to evaluate is the focus on recognition rather than unambiguous identification of entities in the text. The explanes and summaries are deleted. The original DateExtractor test dataset of the MRT library does not allow to fully analyze the functionality of corresponding tools of this type — it covers most variants of dates writing in English, includes common abbreviations, but does not take into account the possibility of distortion of the input text. It seems advisable to compile a new test dataset that takes this aspect into account when evaluating fact extraction tools. The methodology for forming a representative test dataset is outlined in [5].

C. Using Wolfram Mathematica to form a test dataset

Focusing on the tools for extracting temporlal markers in the text, using fragments from DateExtractor, a new test dataset was prepared. In the resulting dataset of 141 items, distortions (errors) most typical for manual typing