International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 5, October 2023, pp. 5314~5332
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i5.pp5314-5332

Development of an intelligent information resource model based on modern natural language processing methods

Zhanna Sadirmekova[1,2], Madina Sambetbayeva[2,3], Sandugash Serikbayeva[3], Gauhar Borankulova[1], Aigerim Yerimbetova[2,4], Aslanbek Murzakhmetov[1]
[1]Department of Information Systems, Faculty of Information Technology, M.Kh. Dulaty Taraz Regional University, Taraz, Kazakhstan
[2]Institute of Information and Computational Technologies, Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan, Almaty, Kazakhstan
[3]Department of Information Systems, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[4]Department of Information Systems, Faculty of Information Technology, Satbayev University, Almaty, Kazakhstan

Article Info

Article history:

Received Jan 18, 2023
Revised Jan 31, 2023
Accepted Feb 10, 2023

Keywords:

Information resources
Methods natural language processing
Natural language processing
Text analysis
Ontology

ABSTRACT

Currently, there is an avalanche-like increase in the need for automatic text processing, respectively, new effective methods and tools for processing texts in natural language are emerging. Although these methods, tools and resources are mostly presented on the internet, many of them remain inaccessible to developers, since they are not systematized, distributed in various directories or on separate sites of both humanitarian and technical orientation. All this greatly complicates their search and practical use in conducting research in computational linguistics and developing applied systems for natural text processing. This paper is aimed at solving the need described above. The paper goal is to develop model of an intelligent information resource based on modern methods of natural language processing (IIR NLP). The main goal of IIR NLP is to render convenient valuable access for specialists in the field of computational linguistics. The originality of our proposed approach is that the developed ontology of the subject area "NLP" will be used to systematize all the above knowledge, data, information resources and organize meaningful access to them, and semantic web standards and technology tools will be used as a software basis.

Corresponding Author:

Zhanna Sadirmekova
Institute of Information and Computational Technologies, Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan
28 Shevchenco str., Almaty, 050010, Kazakhstan
Email: janna_1988@mail.ru

1. INTRODUCTION

At present, ontologies are the most effective instrument to formalizing and also systematizing data and knowledge in scientific subject fields. The scientific newness of research is that an ontology of modern natural language processing (NLP) methods will be developed for the first time, including both classical NLP methods and methods using machine learning. Previously developed ontologies in computational linguistics included mainly the description of classical NLP methods, paying little attention to machine learning methods. In addition, many new NLP methods and models have recently appeared that have not yet been reflected in previously developed ontologies. Developed within the framework of this research will become the conceptual basis of a multilingual intellectual information resource on modern NLP methods, which will ensure the systematization of all information on these methods, its integration into a single information space, convenient navigation through it, as well as meaningful access to it.

To obtain an ontology that would describe NLP quite fully, it is necessary to process a lot scientific papers and information resources from modeling field. To simplify and speed up this process, approaches are being developed to automate replenishment of ontology basing on NLP [1], [2] and modern web resources [3]. For automatic text processing, clustering-based approaches are used, as well as methods for extracting named entities based on neural networks (bidirectional encoder representations from transformers (BERT) [4], generative pre-trained transformer (GPT), TransformerXL, or RoBERTa), and linguistic template-based approaches that use information about vocabulary, syntactic and semantic relationships. However, the first approach requires large corpora of texts in national languages to work well, so methods based on linguistic patterns are more common. The origins of the linguistic approach are the idea proposed in [5] about the possibility of automating the construction of semantic links based on diagnostic contexts presented in the form of lexico-syntactic patterns. This method, known as Hearst patterns, is designed to process unstructured English texts. Research that solves the problem of automatic or semi-automatic replenishment of the ontology is based on patterns of language structures located in texts. These are lexico-syntactic patterns used syntactic information and lexical perceptions [6], [7] or lexico-semantic patterns [8], [9], which combine lexical perceptions with semantic and syntactic information in extraction process. A feature of the proposed approach is that the applied patterns will be built automatically basing on other types of ontology patterns [10], prerequisites of which are presented in [11], [12], and for further term extraction, named entity extraction methods based on machine learning will be used [13], trained on those terms that will be extracted using patterns. It should be noted that the development and application of patterns for the Kazakh language will require serious scientific research, since patterns are highly dependent on the grammatical features of the language. Next, an analysis was made literature data on the modern method NLP.

In the presented review [14] considered modern methods and approaches to text processing and analysis based both on the use of statistical methods and machine learning approaches (including deep learning) and on an effective combination of different approaches. Methods of semantic analysis of texts, representation of words, sentences and documents, including those based on representation in vector space, are presented. The difference in approaches and possibilities for analyzing both individual texts and large arrays of textual information is discussed. Also, in this review given comparison review of modern software tool’s ability for working with natural languages and solutions for applied areas (information-analytical and search systems, corporate document management, and automatic analysis of data from social networks) is given. Typical tasks for computational linguistics and text analysis in various fields of scientific and economic activity are summarized.

Belov *et al.* [15] provides an overview of deep learning methods application in NLP, focusing on several problems where deep learning methods shows a stronger effect. In addition, the paper explores, describes and revises major tools in NLP researching. It also describes the functionality of modern software, their hardware and popular cases.

Lauriola *et al.* [16] present an up-to-date review evaluating NLP researches and applications in construction sector. Thus, basic functions of NLP such as document organization and information extraction, have been significantly improved using various methods. NLP functionality can also be useful for applications to improve efficiency on decision-making systems and management. But additional efforts are needed to continuously improve functionality and develop a framework for NLP application, taking into account trade-off between efficiency and precision. As the authors write, this paper differs in that: unlike other papers looking at artificial intelligence (AI) in a construction system which only enter NLP in general, this paper is first one that provides i) a complex implementation to lower-level methods, basic applications, and subsequent appendices; and ii) a deep discussion of ideas, problems and further directions and implementation. Research results are useful because they provide to project teams with handy links for recognizing advance NLP techniques.

Wu *et al.* [17] describes semantic relationship classification (RC) and named entity recognition (NER) models in information technology scientific texts. Scientific texts contain useful information about advance scientific achievements, however the effective processing of continuously increasing amounts of data is a time-consuming problem. Therefore, this problem always requires improve of information processing automatic methods. Modern models, as a rule, solve the indicated problems quite well using deep learning. To get quality data from specific fields of knowledge, the authors propose to retrain the obtained models on specially prepared corpora. The paper contains a description of the created corpus of Russian texts. The RuSERRC corpus containing 80 marked up and 1,600 unmarked documents with entities and semantic relationships.

Bruches *et al.* [18] describes the problems of the information explosion associated with automatic digital data processing systems, including their recognition, sorting, meaningful processing and presentation in a form acceptable to human perception. A natural solution is to create intelligent systems for extracting knowledge from unstructured information. This paper is a systematic review of international and domestic publications devoted to the leading trend in the field of automatic processing of text information flows, namely text mining (TM). The TM major concepts and problem, its role in field of AI is considered, as well as the difficulties in processing texts in natural language (NLP), due to the weak structure and ambiguity of linguistic information. The stages of pre-processing of texts, their purification and selection of features are described, which, along with the results of morphological, syntactic and semantic analysis, are components of TM. In a general way, a description of the mechanisms of NLP is given: morphological, syntactic, semantic and pragmatic analysis. A comparative analysis of modern TM tools is presented, which allows choosing a platform based on the specifics of the task being solved and the practical skills of the user.

Musaev and Grigoriev [19], authors propose a reasoning network modeling framework and an automatic graph of literature knowledge based on ontology and NLP. His will facilitate the effective study of knowledge from various literature abstracts. In this structure proposed an ontology of representation to describe abstract data of literature on the four elements of knowledge (background, goals, decisions, and conclusions). NLP methods used to automatically extract instances of ontology from literature abstract. A four-dimensional integrated knowledge graph is built on the basis of the representation ontology using NLP technology. To testing structure, a case study is being conducted to analyze the construction management literature.

Chen and Luo [20] proposes an NLP-based semantic framework that implements automatic rule-based conformance building information modeling (BIM) checking at development stage. Semantically extensive data and information can be extracted using NLP techniques, which have been parsed to create conceptual classes and individual ontology elements. BIM data was extracted from Revit into a spreadsheet using the Dynamo tool, and then mapped to an ontology using the CellFie tool. The interoperability of various source data has been significantly improved due to the isomorphism of information within semantic integration, as a result of which the data processed by the semantic web rules language has been transformed from security rules to achieve the goal that automated compliance verification is implemented in the project documentation. The practical and scientific feasibility of the proposed structure was confirmed with a 95.21% recall and 90.63% accuracy of the verification of compliance with the case study in China. In comparison with traditional methods of conformance testing, the proposed framework has high effectiveness, responsiveness, data interaction and interoperability. Despite the existence of the developed NLP methods, a sufficiently complete classification of them has not yet been built; there is no formalized description of these methods; access to ready-made implementations and instructions for their use and embedding is not provided.

Scientific newness of our paper is that for the first time will be developed an ontology of modern methods of automatic text processing, including both classical NLP methods and methods using machine learning. Previously developed ontologies in computational linguistics included mainly the description of classical NLP methods, paying little attention to machine learning methods. In addition, many new NLP methods and models have recently appeared that have not yet been reflected in previously developed ontologies.

Thus, the created unique multilingual intelligent information resource based on modern methods of natural language processing (IIR NLP) will facilitate the search for information on all modern methods of NLP and make it possible to use them in practice in conducting research in the field of computational linguistics and developing applied automatic text processing systems in which government and commercial organizations are interested. This will raise the level of research work and the entire scientific and technical potential, and increase the competitiveness of scientific organizations and their teams. For the first time, a large-scale multilingual resource based on the Kazakh language will be created, within which information will be systematized on classical and modern methods of automatic text processing. Originality of proposed methods is order to systematize all the above knowledge, data and information resources and organize meaningful access to them, the developed ontology of the subject area NLP will be used, and the standards and tools of semantic web technologies will be used as a software basis. Users of such a resource can be both researchers, teachers and students researching, teaching and studying this discipline.

2. CONSTRUCTION A MODEL OF INTEGRATED SUPPORT FOR THE DEVELOPMENT OF IIR NLP

The conducted analytical review showed that there are no universal tools for the development of IIR NLP in the public domain, and existing solutions using different principles and approaches have their own advantages and disadvantages. This problem is devoted to the description and formalization of the proposed model of integrated development support (IDS) of IIR NLP [21]. The main elements of the IDS model are presented in Figure 1 (where: providing meaningful access (PMA)-provision of content access; IIR NLP-intelligent information resource based on modern methods of natural language processing).

To describe the IDS model, consider the formal system $FM=<C, R, A>$, where C is atomic concepts set of the IDS, R is atomic roles set that describe concepts properties and relationship between them, A is the set of axioms that establish a connection between concepts and roles. Axioms can contain descriptions of concepts, composed of atomic concepts and role constraints using constructors of certain logic. The *FM* system under consideration is specified within the description logic SOIN(D) [22]. To describe *FM* elements, additional symbols *(,), =, {,}, [,]* were introduced, the semantics of which, if necessary, is explained as the expressions using them are introduced. The use of additional symbols does not go beyond the logic of SOIN (D), but simply allows you to describe the elements of the IDS model more compactly.

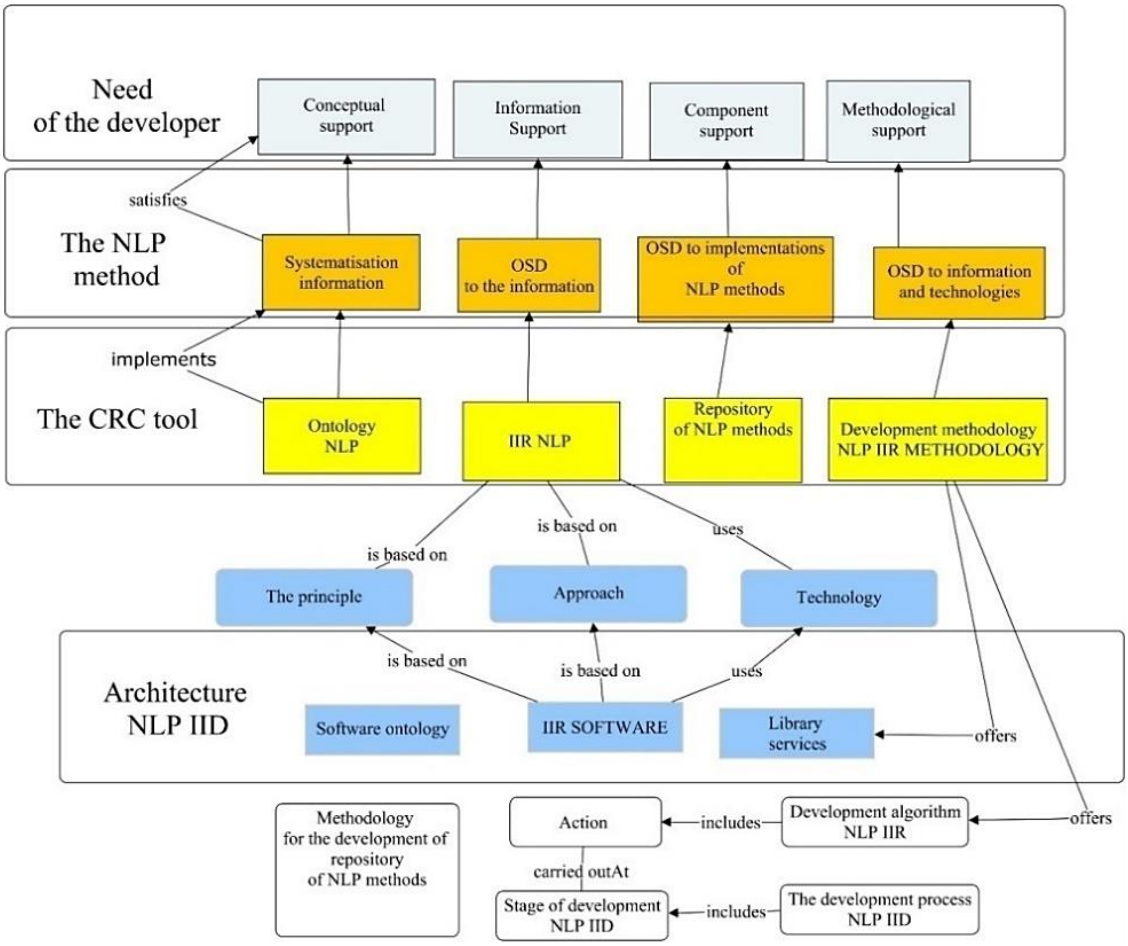

Figure 1. Main elements of the IDS model

Sets *C* and *R* include the following main concepts and roles as shown in Figure 1:

```
C={Developer_IIR NLP, Developer_Need _IIR NLP, Method_IDS, Tool_IDS,
      Development_Technique_IIR NLP, Development_Principle_IIR NLP,
      Approach_to_development_IRNLP, Technology_for_development_IIR NLP,
      Methodology_of_repository_development_of_methods_NLP,
      Architecture_IIR NLP, Development_algorithm_IIR NLP, Action,
      Development_Process_IIR NLP, Stage_development_ IIR NLP
}
R={satisfies, implements, borrows, uses, builds on, proposes, includes}
```

The set *A* includes axioms that define the "general-private" relationship between concepts, define the domains of definition of the left and right arguments of binary relations corresponding to roles, and describe the properties of some concepts and roles. Axioms contain compound concepts and role constraints, which are defined according to SOIN(D) logic constructors.

Consider, as an example, an axiom that describes the properties of the concept "Tool_ IDS":

```
Remedy_IDS⊑∃ based on [Design_principle_IIR NLP ⊔
Approach_to_development_IIR NLP]⊓
∃uses. Technology_for_development_ IIR NLP
```

Interpretation I of a formal system that defines its semantics is represented by a pair $I=<\Delta, f>$, where Δ is the domain, the set of all instances (individuals) of IDS model, f is interpretation function.

$$f: C \cup R \rightarrow (2\Delta) \cup (2\Delta \times 2\Delta).$$

$f(Ci) \subseteq \Delta, Ci \in C$ – each concept is associated with a set of its instances (individuals). $f(Ri) \subseteq \Delta \times \Delta$, $Ri \in R$ – each role is associated with a binary relation, i.e., the set of pairs of individuals connected by it.

The function f is extended to interpret compound concepts according to the inductive rules of SOIN(D) logic. Then the interpretation of the formal system *FM* will define the IDS model as a set of specific objects belonging to certain classes, connected by certain relations and satisfying Axioms A. Note that the term "model" can be considered not only in a logical or algebraic sense, but also as a simplified information representation of a set of objects and processes of the real world [23] that describe the IDS process.

When describing concepts, roles, axioms and their interpretations, the following naming conventions are used [24]: i) Concept names are written with an uppercase letter followed by a lowercase letter. If the name consists of several words, they are connected with an underscore (*Concept_IDS*, *Need_developer_IIR NLP*); ii) Role names are written in lowercase letters. If the name consists of several words, they are "glued" to each other, and each word, starting from the second, is written with a capital letter (based on, performed on the stage); and iii) Object names consist entirely of uppercase letters. If the name consists of several words, they are connected with an underscore (*KNOWLEDGE_ENGINEER*, *CONCEPT_SUPPORT*).

The IDS model considers integrated support as a process of meeting needs of IIR NLP developers, running in parallel with the IIR development process. It suggests methods for satisfying needs and means for implementing these methods. Specialists of various profiles participate in the development of IIR NLP-experts of the subject area (software) for which IIR NLP is being created, knowledge engineers who are specialists in the field of knowledge representation and programmers:

```
f(Developer_IIR NL)={KNOWLEDGE_ENGINEER, EXPERT, PROGRAMMER}
```

In addition to general ideas about the field of knowledge, developers need information about specific methods of automatic text processing of natural language texts and their corresponding software tools and resources, about the classes of problems solved by these methods, about the capabilities and limitations of each of them. In addition, developers should have access to information about the main stages of NLP methods and the methods used in each of them, as well as about the tools that implement these methods. Component support for developers plays important role in IIR NLP implementation stage. Ability to select and test ready-made software components that implement the necessary methods of automatic processing texts or organizing an interface with users can significantly facilitate and speed up the process of creating IIR NLP. Another important need for developers at the implementation stage is methodological support, including a description of the approaches and principles for the development of IIR NLP, development tools and methods for their application:

```
f(Developer_need)={
      CONCEPTUAL_SUPPORT, INFORMATIONAL_SUPPORT,
      COMPONENT_SUPPORT, METHODOLOGICAL_SUPPORT
}
```

According to the proposed concept, the conceptual basis of the IDS is provided by systematizing information about the area of knowledge "Automatic text processing", the result of which is the ontology of this area. Information support is carried out by PMA to knowledge structured on the basis of ontology, information resources and methods related to the field of knowledge "Automatic text processing". The means of such support is a specialized intellectual information resource on modern methods of automatic text processing. Component support of the resilience development initiative (RDI) development process is carried out by providing direct meaningful access to the implementations of NLP methods and is provided by the repository NLP methods. Methodological support is provided by providing information about the approaches, principles, technologies, algorithm for developing IIR NLP, as well as their architecture. Thus, the main methods of IDS are methods of working with information about NLP methods and aspects-their systematization and PMA to them, as well as PMA to implementations of NLP methods and the methodology for developing IIR NLP:

```
f(IDS_Method)={
ORGANIZATION_INFORMATION,
      PMA_TO_INFORMATION_ABOUT_NLP_METHODS,
      PMA_TO_IMPLEMENTATION_NLP_METHODS,
PMA_TO_INFORMATION_ABOUT_APPROACHES_PRINCIPLES_TECHNOLOGIES
}
```

IDS means are the ontology of the knowledge area (KA) NLP, intellectual an information resource on NLP, a repository of NLP methods, and a methodology for developing IIR NLP:

```
f(Means_IDS)={
ONTOLOGY_AK_NLP, IIR_NLP, REPOSITORY_METHODS,
METHODOLOGY_DEVELOPMENT_IIR NLP
}
```

IDS methods meet the needs of IIRNLP developers:

```
f(satisfies)={
      (ORGANIZATION_INFORMATION, CONCEPTUAL_SUPPORT),
      (PMA_K_INFORMATION_ABOUT_NLP_METHODS, INFO_SUPPORT),
      (PMA_TO_IMPLEMENTATIONS_NLP_METHODS),
      (COMPONENT_SUPPORT),
      (PMA_K_INFORMATION_ABOUT_APPROACHES_PRINCIPLES_TECHNOLOGIES,
      METHODOLOGICAL_SUPPORT)
}
```

IDS tools implement IDS methods:

```
f(realizes)={
      (ONTOLOGY_AK_NLP),
      (ORGANIZATION_INFORMATION),
      (IIR _NLP, PMA_K_INFORMATION_ABOUT_METHODS),
      (REPOSITORY_METHODS),
      (PMA_TO_IMPLEMENTATION_METHODS),
      (METHODOLOGY_DEVELOPMENT_IIR NLP,
      PMA_K_INFORMATION_ABOUT_APPROACHES_PRINCIPLES_TECHNOLOGIES)
}
```

2.1. Approaches, principles and technologies proposed for IIR NLP development

The NLP model in Figure 2 under consideration includes a number of the most important principles that have proven themselves in practice, as well as approaches and technologies that ensure compliance with these principles, simplify and speed up the development of IIR NLP, determine the properties and functionality of those IIR NLP that will be developed within the framework of this model. Note that the proposed NLP tools can be considered as IIR for the development of IIR NLP. Therefore, they are developed using the same principles, approaches and technologies that will be used for specific IIR NLP.

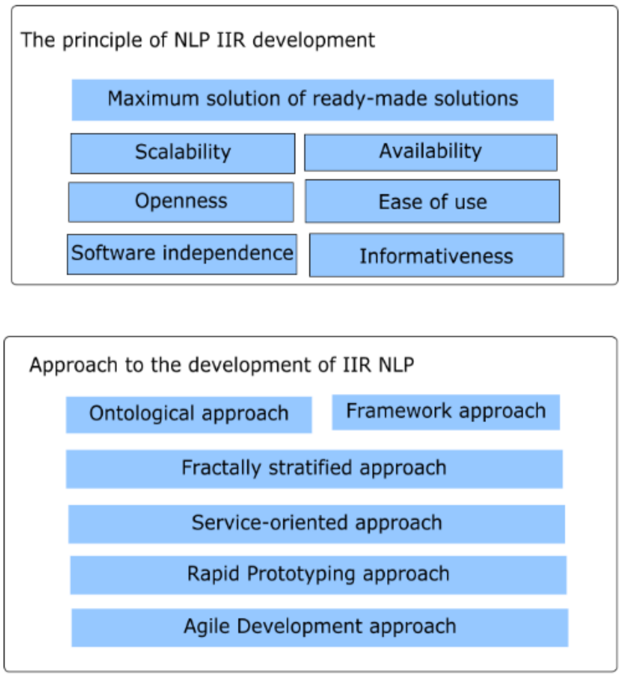

Figure 2. Principles, approaches to the development of IIR NLP

Currently, for scientific and practical purposes, a large number of NLP methods and software products that implement them-libraries, packages, applications-have been developed. And now, in many fields of knowledge, a situation is developing when it is difficult to come up with something new, and what seems to be so is either well-forgotten or not well-known old, or a combination or variation of already known developments. Therefore, when providing comprehensive support, it is very important to provide access to existing implementations and, when developing IIR NLP, ensure maximum use of ready-made solutions. However, life does not stand still. New methods appear, old ones are modified and improved. Another important challenge is to ensure the scalability and relevance of the proposed tools and specific IIR NLP. Compliance with these principles will make it easy to connect new NLP methods and interface components, modify existing ones, solving the problem of obsolescence of the developed systems.

The tools used and the systems being developed should be open-source. Only in this case will their widespread use be ensured. Despite the possibility of choosing the proposed tools for developing IIR NLP and solving specific problems, it is impossible to completely eliminate the need for their modification. The principle of openness will make it painless. To reduce the qualification requirements for developers and users of IIR NLP, the proposed tools should be easy to use and, if possible, independent of the subject area.

Despite the fact that many products are freely available, developers of IIR NLP, instead of using them in whole or at least in part, often "reinvent the wheel". The main reasons for this situation are the lack of a general systematic description of software products in this area, the incompleteness of the author's specifications, and the obsolescence of the platforms on which these products were implemented. Therefore, an important principle for both ensuring the IDS and creating the IIR NLP is the principle of informativeness, i.e., PMA to information about all aspects of the development and operation of IIR NLP:

```
f(Principle_development_IIR NLP)={
      MAXIMUM_USE_of_TOUCH_SOLUTIONS, SCALABILITY, AVAILABILITY,
      OPENNESS, EASY_USE, SOFTWARE_INDEPENDENCE, INFORMATION
}
```

One of the main elements of intelligent systems is knowledge bases (KB), whose role is currently played by ontologies. When developing methods and tools for IDS, the ontological approach was also used to develop knowledge bases. The NLP ontology is the conceptual basis of the IDS. According to the proposed model, IIR NLP should also be developed on the basis of ontology. When developing a KB, in addition to the ontological one, it is proposed to use the fractal-stratified approach (FS-approach), which is based on the fractal approach proposed by Sadirmekova *et al.* [25]. This approach allows us to take into account the similarity of the knowledge system as a whole and its fragments. The FS-approach was supplemented by the author with the possibility of using different types of stratification in FS models.

The spherical bundle proposed by the author of the approach is convenient to use to represent different levels of knowledge detail; line bundle (block, parallelepiped) represents non-embedded pieces of knowledge. The beam, "penetrating" the layered parallelepiped of knowledge, is an invariant, i.e., fragment of knowledge present in all strata. The strata in such a bundle can represent, for example, different aspects of concepts used by different specialists.

Combined stratification-linear-hierarchical, pyramidal, is convenient to use to represent non-nested fragments of knowledge. They have their own hierarchical structure. In combined stratified models, an invariant is a meta ontology that contains the most general concepts of the knowledge space, which are concretized both at different levels of detail and in different strata. Using the FS-approach regulates and simplifies the systematization of knowledge, allows you to reuse previously developed fragments of knowledge.

Service-oriented approach has proven itself well [26]. All the functionality of IIR NLP, the methods that will be used to solve the tasks, as well as the user interface, are proposed to be implemented as services. The use of this approach creates the prerequisites for the feasibility of the basic principles of the IDS mentioned above.

Rapid prototyping, agile development approaches and the framework approach are universal in the development of a wide class of software systems. These approaches are in good agreement with the principles and other approaches included in the IDS model. Their advantages make it possible to successfully apply them to the development of IIR NLP. Thus, for the development of IIR NLP, it is proposed to use the following approaches:

```
f(Approach_to_development_IIR NLP)={
      ONTOLOGICAL_APPROACH,
      FRACTAL_STRATIFIED_APPROACH,
      SERVICE-ORIENTED_APPROACH,
      APPROACH_FAST_PROTOTYPING,
```

```
            FLEXIBLE_DEVELOPMENT APPROACH,
            FRAMEWORK_APPROACH
        }
```

The approaches used ensure compliance with the principles mentioned above:

```
Approach_to_development_ IIR NLP ⊑
       ∃provides. Design_principle_IIR NLP
       The IDS model includes two technologies used to develop IIR NLP (Figure 3):
       f(Technology_for_development_ IIR NLP) = {
             TECHNOLOGY_SEMANTIC_WEB,
             TECHNOLOGY_DEVELOPMENT_ISSEA
       }
```

These technologies contain some set of tools. These tools are sufficient to ensure that the development process is simple, understandable, and satisfies the principles discussed above. Figure 3 shows their main components:

```
Technology_for_development_IIR NLP ⊑
       ∃includes. [Knowledge_representer
       ⊔KB_development tool
       ⊔Information_storage⊔Service
       ⊔output_machine
       ⊔Query_language⊔Shell_ISSEA].
```

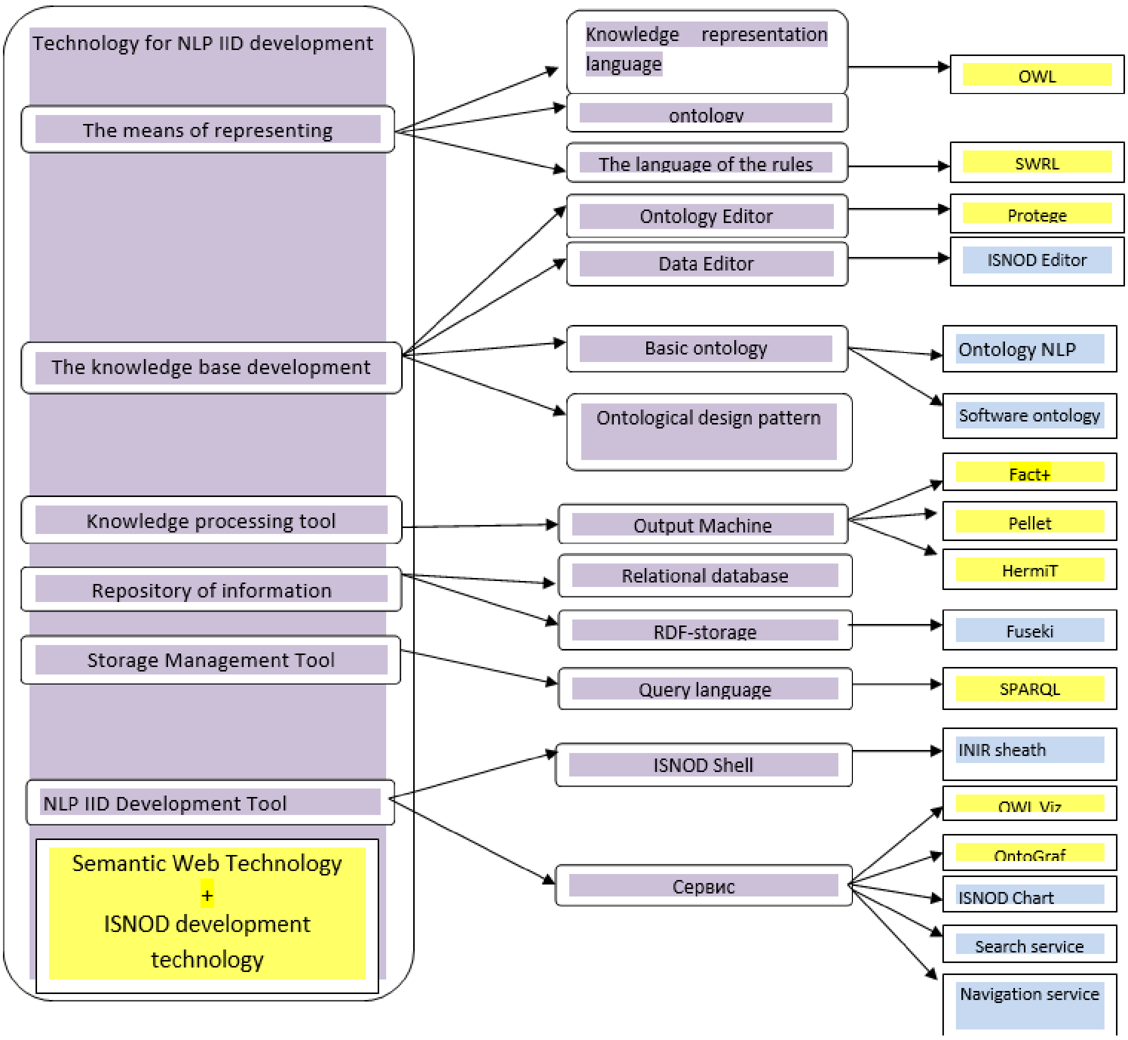

Figure 3. Technologies used to develop IIR NLP

In modern intellectual systems, knowledge, as a rule, is represented by an ontology and inference rules, which allow obtaining new knowledge that is not explicitly contained in the knowledge base, and are described by the corresponding languages. To develop a knowledge base, it is convenient to use ontology editors and data editors. Such editors make it possible to reduce the qualification requirements for developers and involve subject matter experts in the creation of knowledge bases. The use of base ontologies and ontology design patterns can also simplify the development of ontologies. Various types of storages can be used to store information.

The functionality of a particular technology is determined by the set of services it provides. These can be user interface services designed to receive data from the user and display the results of the work of the IIR NLP; analytical services that allow you to search for the necessary information and navigate through the information content of the IIR NLP; graphic services designed to present information to the user in graphical form; services that implement certain NLP methods and allow solving the tasks assigned to IIR NLP.

Inference engines are used to organize logical inference in an ontology based on axioms and rules, as well as to check the ontology for consistency. To search for information in the repository, a query language is used. The IIR development technology provides a shell for the future IIR NLP. The following axioms define the hierarchy of concepts of the IDS model that describe the technologies used:

```
Knowledge_representation tool ⊒Knowledge_representation_language⊔Ontology
⊔Rule_language
KB_development tool ⊒Ontology_editor⊔data_editor⊔base_ontology⊔ontology_design
pattern
Information_storage⊒rdf_store⊔Relational_DB
Service ⊒UI_service⊔Analytical_service⊔
Graphic_service⊔Service_implementing_NLP_method
Analytical_service⊒search_service⊔Service_navigation
```

Figure 3 shows technologies used to develop IIR NLP. Specific elements of semantic web technology are highlighted in yellow. The elements of IIR NLP development technology are highlighted in blue.

```
f(includes)={
      (TECHNOLOGY_ SEMANTIC_WEB,
      (OWL, PROTÉGÉ, ONTOGRAF, OWLVIZ,
      FACT+, PELLET, HERMIT, SPARQL))
}
```

Such an abbreviation means the set of all pairs of the “includes” relation, the first element of which is the semantic web technology, and the second is one of the components of this technology, listed in parentheses. Similarly, instances of this relation are presented for the technology of developing information system supporting scientific and educational activities (ISSEA) [27]:

```
f(includes)={
      (TECHNOLOGY_DEVELOPMENT_ ISSEA,
      (EDITOR_ISSEA, ONTOLOGY_SCIENTIFIC_KNOWLEDGE,
      ONTOLOGY_SCIENTIFIC_ACTIVITY,
      ONTOLOGY_TASK_METHODS,
      ONTOLOGY_INFORMATION_RESOURCES,
      FUSEKI, GRAPHICS_IIR NLP,
      SHELL_ISSEA))
}
```

At the same time, the ISSEA development technology uses the components and tools of semantic web:

```
f(uses)={
(TECHNOLOGY_DEVELOPMENT_ISSNOD, TECHNOLOGY_SEMANTIC_WEB)
}
```

One of the important parts of IIR NLP development is conceptual support. The NLP ontology is a conceptual support tool for the development of IIR NLP [28]. Figure 4 shows the main components of the ontology, as well as approaches, technologies and specific technology components used to develop it. Next, the consideration NLP ontology model described in the framework of the SOIN(D) descriptive logic:

```
Ontology_DSS⊑∃ includes. Ontology_element⊓
       ∃ is built on the basis of.base_ontology
Ontology_element⊒ Concept ⊔Concept_property⊔ Axiom
```

```
        f(usedFor Development)={
              ((FRACTAL_STRATIFIED_APPROACH,
              METHODOLOGY_DEVELOPING_ONTOLOGIES_SCIENTIFIC_FIELDS,
              ONTOLOGY_SCIENTIFIC_KNOWLEDGE,
ONTOLOGY_SCIENTIFIC_ACTIVITY,
              ONTOLOGY_NLP_METHODS,
              ONTOLOGY_INFORMATION_RESOURCES),
              ONTOLOGY_OS_ NLP)
        }
```

Such an abbreviation means the set of all pairs of the "*usedForDevelopment*" relation. The first element of which is the approach, or technology component, that was used in the development of the ontology. The second element is the ontology of the NLP:

```
        f(Base_ontology)={
              ONTOLOGY_SCIENTIFIC_KNOWLEDGE, ONTOLOGY_TASKS_AND_METHODS,
              ONTOLOGY_INFORMATION_RESOURCES,
              ONTOLOGY_SCIENTIFIC_ACTIVITY
        }
```

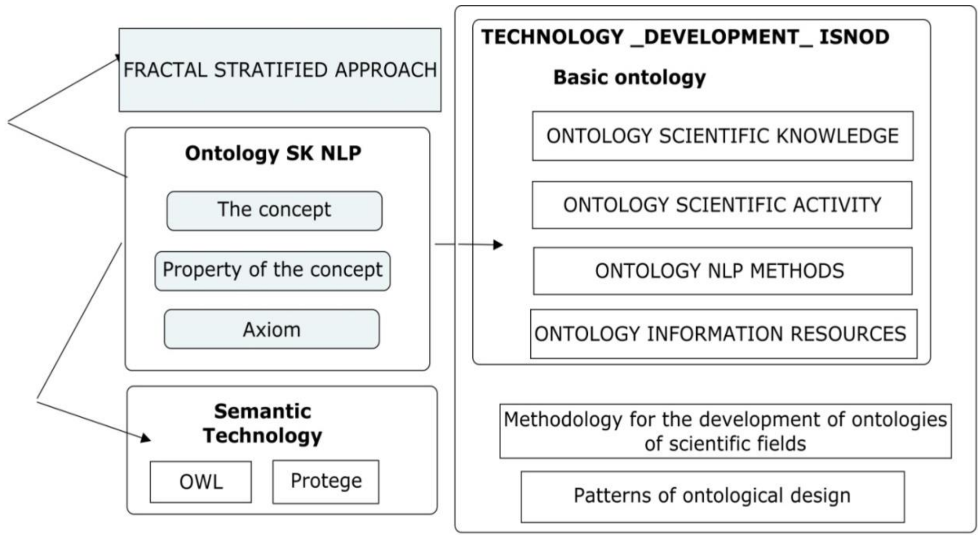

Figure 4. NLP ontology model

2.2. Development of data warehouse

The component support tool for developers is the NLP methods repository. The repository model is shown in Figure 5. Repository includes three major types of components:

```
repository⊑∃includes.repository_component
repository_component⊒Method_NLP⊔info_object⊔
        Software_system
```

The following axioms describe the subclasses and roles of repository components, as well as their relationship to IIR NLP content:

```
Software_system⊒network_object⊔ Non-network_object
Software_system⊑∃has Type.Software_system type ⊓
        ∃implements.Method_NLP
info_object⊑∃describes. [Method_NLP⊔
        Software_system]
Content_IIR⊑∃ contains. [info_object⊔
        Network_object]_
        f(Type of_software_system)={
              DESKTOP_APPLICATION, WEB_APPLICATION, MODULE, PACKAGE, SOFTWARE_LIBRARY,
SERVICE, TECHNOLOGICAL_COMPLEX, FRAMEWORK
        }
```

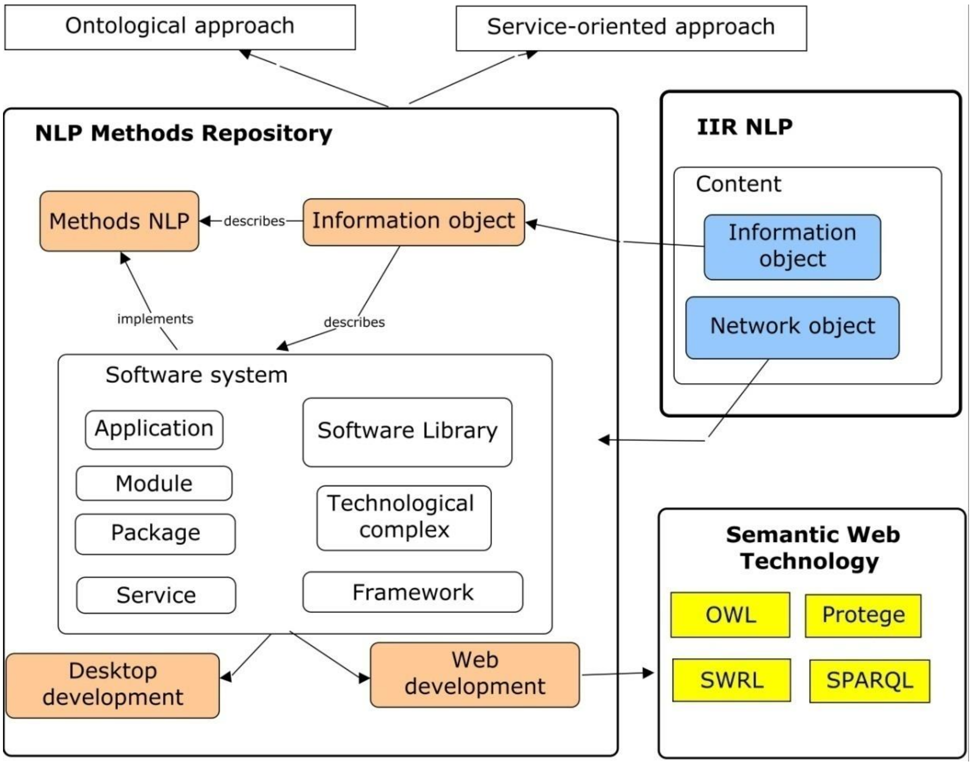

Figure 5. NLP method repository model

Moment, the repository includes a representative set of methods and software systems that implement them, which, if necessary, can be supplemented:

```
f(Method_NLP)={
      QUESTIONNAIRE, PROBABILISTIC_MODELING,
COGNITIVE_MODELING, METHOD OF UNDETERMINATED_COMPUTIONS,
      ONTOLOGICAL_MODELING,
      REASONING_BASED ON_CASEDENTS,
      RULES-BASED REASONING,
      EVENT_MODELING, EXPERT_ASSESSMENT
}
f(Software_system)={
      CMAP_TOOLS, COLIBRI, GOOGLE_FORMS,
      SEMP_TAO, NETICA, NEMO+, INTEGRA,
      PROTEGE, CLIPS, CO_MOD,
      GRM_ONTO_MAP,
      GRM_COG_MAP,
      GRM_EVENT_MAP, MEO,
      SEMP_NEMO,
      UNICALC, DATA_ARRIGE_PROCESSING_SERVICE,
      SERVICE_INTERACTION_WITH_EXTERNAL_DATA
}
f(implements)={
      ((CMAP_TOOLS, PROTEGE, GRM_ONTO_MAP),
      ONTOLOGICAL_MODELING),
      (COLIBRI, REASONING_FROM_CASEDENTS),
      ((SEMP_TAO, CLIPS), RULES_BASED_REASONING),
      (NETICA, PROBABILITY_MODELING),
      ((GRM_COG_MAP, CO_MOD),
      COGNITIVE_MODELING),
      (GRM _EVENT_MAP, EVENT_MAPING),
      ((NEMO+, INTEGRA,
      SEMP_NEMO, UNICALC),
      UNDERDETERMINATED CALCULATION METHOD),
      (SYSTEM _EXPERT_ASSESSMENT),
      (EXPERT_ASSESSMENT),
      (GOOGLE_FORMS, QUESTIONNAIRE)
}
```

Following approaches and technologies were used to develop the repository:

```
f(usedFor Development)={
      ((ONTOLOGICAL_APPROACH, SERVICE_ORIENTED_APPROACH,
      TECHNOLOGY_SEMANTIC_WEB), REPOSITORY)
}
```

Access to the repository, to the description of the methods and the software systems that implement them, is provided through the IIR NLP:

```
f(providesAccess)={(IIR_NLP, REPOSITORY)}
```

The development of repository consists in development of new NLP methods, services/software systems implementing them and the integration of these methods and systems, as well as freely available implemented methods with IIR NLP. The integration of each implemented method, in turn, consists in creating an information object in the content of the IIR NLP that describes the software development that implements it, its type, connections with other objects, with instructions for its use, the format of the data used, the IP address at which you can launch the service or download the software system. Repository implementation is a collection of NLP information objects and systems that describe the methods, the tasks they solve and the software systems that implement them. If the software system has an implementation in the form of a service or is freely available and available for download, then it can also become part of the repository. Access to the repository is carried out by means of IIR NLP. By selecting the appropriate method on the resource, the user receives a link by which he can either launch the service or download the appropriate software system.

```
{TECHNIQUE_DEVELOPING_REPOSITORY_NLP_METHODS_DEVELOPMENT} ⊑
       ∃offers.{ALGORITHM_DEVELOPMENT_REPOSITORY} ⊓
       ∃ based on. [Principle_of_development_IIRNLP⊔
                             {SERVICE-ORIENTED_APPROACH}] ⊓
       ∃uses.Service_implementation_technology
{ALGORITHM FOR_DEVELOPING_REPOSITORY} ⊑
       ∃includesAction. Action
Action ⊑Development_of_NLP_method⊔Service_development⊔
       Integration_method_with_IIR NLP
```

According to the proposed model, the main component of the developed IIR NLP, its framework, on which interface and functional components will be "strung" in the form of services, is the thematic IIR of the selected software. It is being developed using the same technology as IIR NLP, using the ISSEA shell it offers. The knowledge base is built on ontology software basis. Services included in IIR NLP architecture are borrowed from the NLP methods repository, and the information objects that describe them are taken from the NLP ontology and IIR NLP content.

The architectural components of a typical IIR NLP are shown below (highlighted in bold outline) and the scheme of their interaction:

```
f(includes)={
      (IIR NLP, (ONTOLOGY_SOFTWARE,
      IIR_SOFTWARE, REPOSITORY_NLP_METHODS)),
      (IIR_software, (ONTOLOGY_software,
      CONTENT_IIR_software))
}
f(borrowFrom)={
      (ONTOLOGY _PO, ONTOLOGY_ NLP),
      (CONTENT_IIR_SOFTWARE),
      (CONTENT_IIR_NLP),
      (LIBRARY_SERVICES),
      (REPOSITORY)
}
```

The IIR NLP development methodology offers an architecture and algorithm as shown in Figure 6, is based on the principles and approaches discussed above and uses the above technologies. PMA to information about approaches and principles of IIR NLP development, as well as to the technologies used, this methodology is a means of methodological support. Compliance with these principles will make it easy to connect new NLP methods and interface components, modify existing ones, solving the problem of obsolescence of the systems being developed.

```
{METHODOLOGY_DEVELOPMENT_IIR NLP} ⊑
       ∃offers. [{ARCHITECTURE_IIR NLP} ⊔
              {ALGORITHM_DEVELOPMENT_ IIR NLP}] ⊓
       ∃ based on. [Design_principle_IIR NLP⊔
              Approach_to_development_IIR NLP]⊓
       ∃uses. Technology_for_development_ IIR NLP⊓
       ∃implements. Method_IDS
```

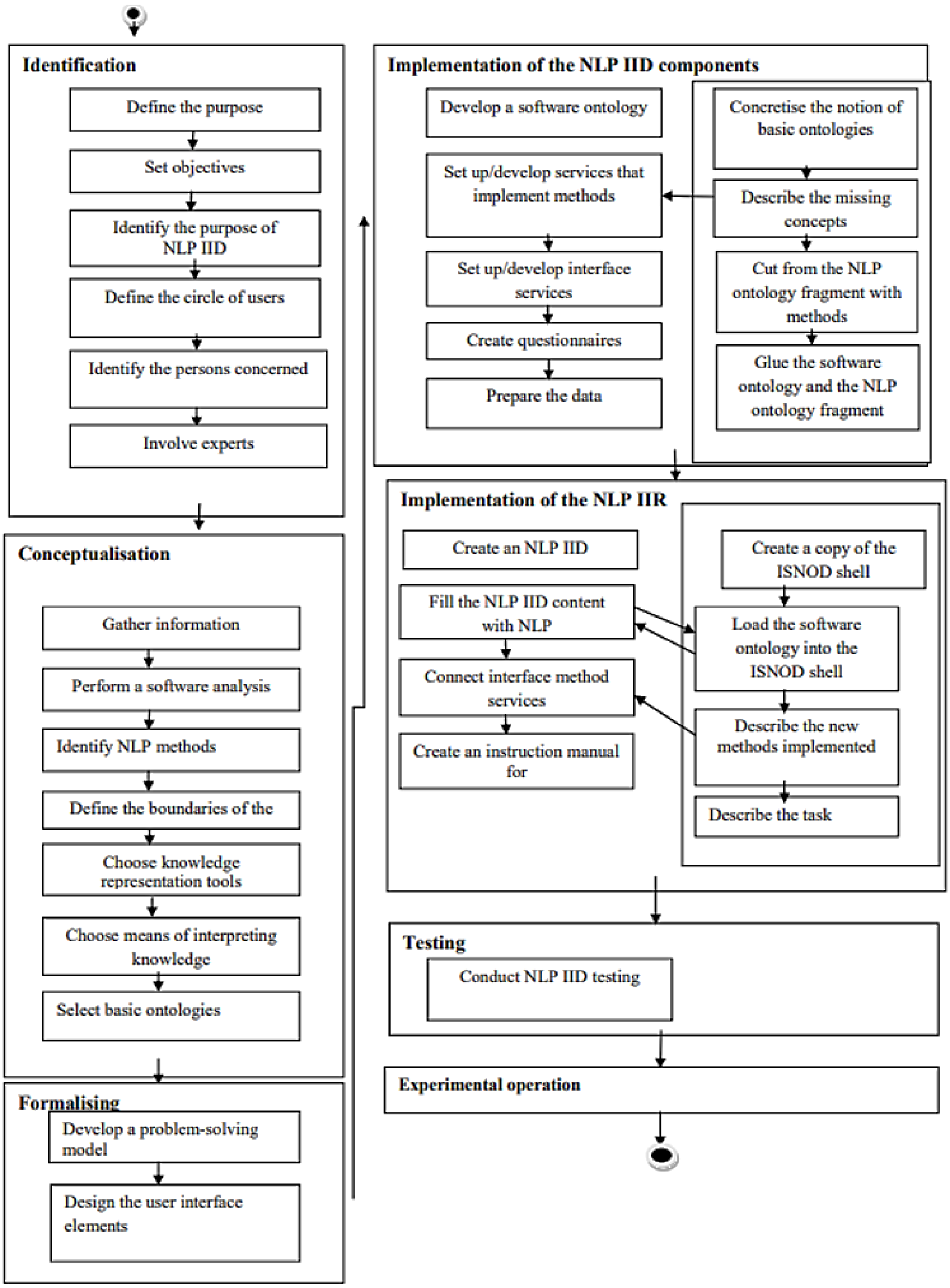

Figure 6. Stages and algorithm of IIR NLP development

The method under consideration proposes an algorithm that regulates the entire process of developing IIR NLP and includes a sequence of actions:

```
{ALGORITHM _DEVELOPMENT_IIR NLP} ⊑
       ∃ regulates. Development_Process_ IIR NLP⊓
∃includesAction. Action
```

Next, we consider the stages and algorithms of developing IIR NLP in SOIN(D):

```
       f(Action)={
             DEFINE_GOAL, SET_TASKS,
             DEFINE_PURPOSE_IIR NLP,
             DETERMINE_CIRCLE_USERS,
             IDENTIFY_STAKEHOLDERS, INVOLVE_EXPERTS,
             COLLECT_INFORMATION, PERFORM_SOFTWARE_ANALYSIS,
             DETERMINE_METHODS_SOLVING_PROBLEM,
             DETERMINE_BORDERS_SOFTWARE,
             CHOOSE_KNOWLEDGE_REPRESENTATION TOOLS,
             CHOOSE_MEANS_INTERPRETATION_KNOWLEDGE,
             SELECT_BASE_ONTOLOGIES,
             DEVELOP_MODEL_SOLUTION_PROBLEM,
             DESIGN_USER_INTERFACE_ELEMENTS,
             DEVELOP_ONTOLOGY_SOFTWARE,
             SPECIFY_CONCEPTS_OF_BASE_ONTOLOGIES,
             DESCRIBE_MISSING_CONCEPTS,
             CUT_FROM_ONTOLOGY_NLP_FRAGMENT_WITH_METHODS,
             GLUE_ONTOLOGY_BY_AND_FRAGMENT_ONTOLOGY_NLP,
             CUSTOMIZE_DEVELOP_SERVICES_IMPLEMENTING_METHODS,
             CUSTOMIZE_DEVELOP_SERVICES_INTERFACE,
             CREATE_QUESTIONNAIRE_QUESTIONNAIRES, PREPARE_DATA,
             CREATE_IIR_SOFTWARE, CREATE_COPY_SHELL_ISSEA,
             LOAD_ONTOLOGY_SOFTWARE_INTO_SHELL_ ISSEA,
             FILL OUT_CONTENT_IIR NLP,
             DESCRIPTION_NEW_IMPLEMENTED_METHODS,
             DESCRIBE_MODEL_TASKS,
             CONNECT_SERVICES_INTERFACE_METHODS,
             CREATE_INSTRUCTIONS_FOR_USERS,
             PERFORM_TESTING_ IIRNLP
          }
```

The development process of IIR NLP includes the same stages as the development of any intelligent systems:

```
{PROCESS OF_DEVELOPMENT_OF IIR NLP} ⊑
       ∃includesStage.Stage_of development_IIR NLP

       f(Stage_development_ IIR NLP)={
             IDENTIFICATION, CONCEPTUALIZATION, FORMALIZATION,
             IMPLEMENTATION, TESTING, PILOT_OPERATION
}
```

On the set of all actions, the relation "≤" is defined, which specifies the order of the algorithm actions, as well as the relation that matches the stages of IIR NLP development with a set of actions performed at these stages, and the relation that matches the action with the type of developer responsible for performing this action:

```
       Action ⊑∃≤. Action
       Action ⊑∃ executed at the stage. Stage_of development_IIR NLP
       Developer ⊑∃ responsible for. Action
```

3. IMPLEMENTATION OF MODEL

Ontology, which is the basis of conceptual support, in turn, is a formal system that defines the concepts, roles and axioms of NLP. Figure 7 shows the main concepts (notions) of this area. The NLP ontology, according to the methodology, is built on basis of scientific knowledge ontologies basics, scientific activities, methods and problems, scientific information resources by supplementing and concretizing the concepts contained in them, as well as ontological design patterns. Basic ontologies are provided by the ISSEA development technology. Let us consider the description of the NLP ontology using the SHOIN(D) description logic formalism, which was also used to describe the IDS model.

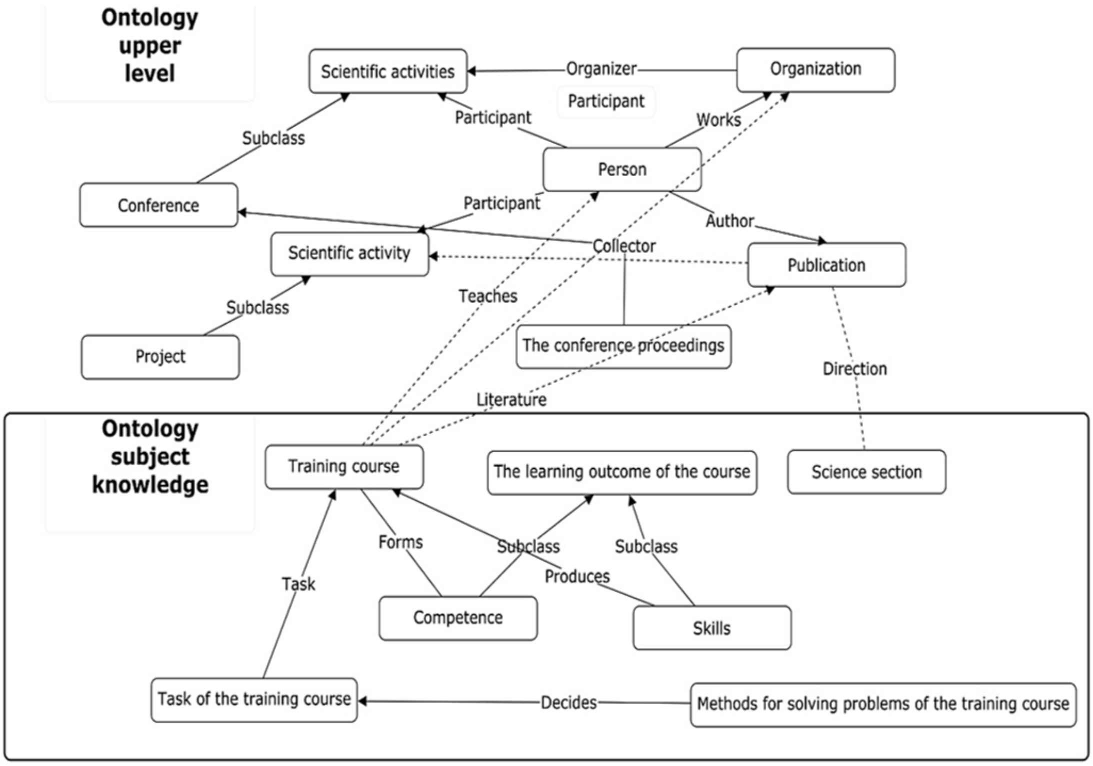

Figure 7. The main concepts (notions) of this area

Let *FO=<OC, OR, OA>* be a formal system, where OC is the set of concepts of the SP OR; OR is atomic roles set describes the concepts properties and relationship between them; OA is the set of axioms that establish connections between concepts and roles. Let us consider concretizations of the concepts of the basic ontology and some roles and axioms of the formal FO system.

NLP ontology was developed in the OWL language using the Protégé editor [29]. To represent the vertical-horizontal organization of the ontology, a structural pattern was developed, implemented using the OWL language and the Protégé editor. This structural pattern shown in Figure 8.

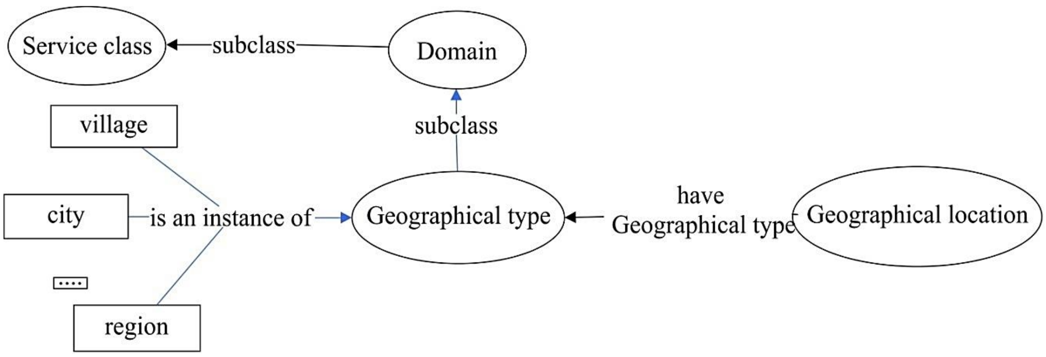

Figure 8. Structural pattern for representing the range of valid values and an example of its use

To implement this pattern, the annotation property of the OWL annotation language was used property, which can be assigned to any ontology element or object. The list of built-in annotation properties has been supplemented with a new type designer property with the values *{DM, KNOWLEDGE ENGINEER, PROGRAMMER}*. This property, along with the built-in label property, is assigned to all ontology elements and objects. During the work of the IIR NLP, the user can choose which type of developer he belongs to, and depending on this, only classes, objects and properties of this type will be shown to him.

To create the IIR NLP, we used ISSEA development technology [30]. This technology provides a resource shell, a methodology for building an ontology and resource content, a data editor, as well as tools for extracting scientific information from the internet [31]. To implement the IIR NLP, a copy of the ISSEA shell was created and user interface elements were configured. The NLP ontology was loaded into the shell. Figure 9 shows the IIR NLP page, on the left side of which shows ontology concepts organized in a public-private hierarchy. The architecture of the ISSEA platform includes three levels: i) level of information presentation; ii) level of information processing; iii) level of storage and access to information. The basic functionality includes navigation in the ISSEA information space (knowledge and data networks), searching for information in ISSEA content (including advanced-in terms of ontology) and editing ISSEA content. The widespread Jena Fuseki repository is used to store the ontology and data [32], however, another RDF repository can also be used.

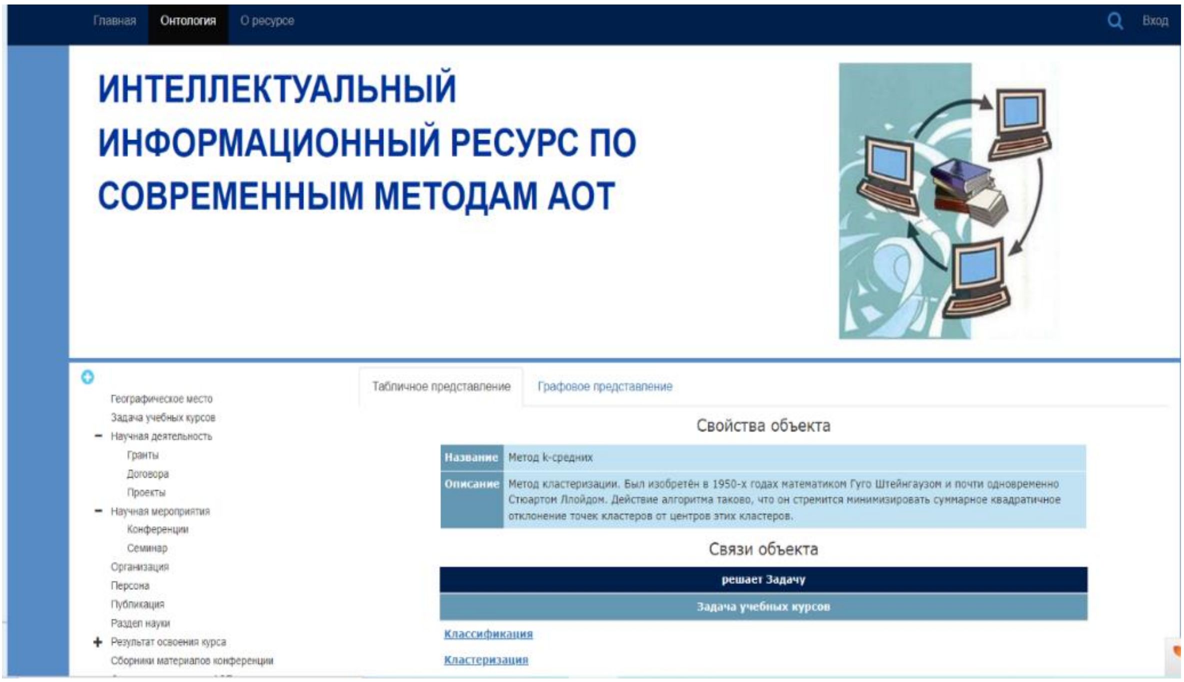

Figure 9. IIR NLP page

4. RESULTS AND DISCUSSION

As a result of the research, we propose a methodology to development of IIR NLP, it offers the architecture and algorithm for the development of IIR NLP. The principles and approaches underlying the methodology determine the following main features: i) focus on semi-formalized software; ii) independence from software; iii) focus on the maximum use of ready-made developments (both copyright and third-party); iv) use of semantic web technologies and service-oriented approach, ISSEA development technologies; v) use of the ISSEA shell as a framework for the future IIR NLP; and vi) openness and scalability of the proposed tools; convenience and low entry threshold for the use of the proposed funds.

The research scientific novelty is that in first time was developed an ontology of modern NLP, including both classical NLP methods and methods using machine learning. Previously developed ontologies in computational linguistics included mainly the description of classical NLP methods, paying little attention to machine learning methods. At the moment, there is a machine learning ontology, which contains a small set of NLP methods based on machine learning. Developed within the framework of this research has become the conceptual basis of a multilingual intellectual information resource on modern methods of automatic text processing, which ensures the systematization of all information using these methods, its integration into a single information space, convenient navigation through it, as well as meaningful access to it. Thus, the use of ontologies as an information model of IIR NLP allows to describe in more detail the catalog of resources on a given topic and their interrelationships, as well as systematize input and output parameters, which significantly increases the performance of the system.

5. CONCLUSION

The model of integrated support for the development of IIR NLP is a set of methods, tools, principles, approaches, technologies, algorithms designed to meet the needs of developers, which are offered to them at all stages of creating systems of this class. The main methods of IDS are methods of working with information about the methods and aspects of NLP-their systematization and PMA to them, as well as implementations of NLP methods and the methodology for developing IIR. The means that implement the methods of conceptual, informational, component and methodological support are, respectively, the NLP ontology, the IIR NLP, the NLP methods repository and the IIR NLP development methodology.

Repository development methodology is based on approach of service-oriented and principle of maximum use of ready-made solutions. To include the NLP method in the repository, it is necessary to find a ready-made one, or develop a software system that implements it, create information objects in the IIR NLP content that describe the method and software system, as well as all aspects of their interaction and use.

A conceptual model of a multilingual intellectual information resource based on modern NLP methods based on ontology was developed. The ontology systematizes information about the NLP knowledge area and provides IIR NLP developers with a single conceptual basis. The description logic language SOIN(D) was used to formalize the model. An intelligent information resource based on the NLP ontology is a means of meaningful access both to information about a given PO and to a repository of NLP methods.

ACKNOWLEDGEMENTS

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP14972834).

REFERENCES

[1] J. Braga, J. L. R. Dias, and F. Regateiro, "A machine learning ontology," *A Preprint*, pp. 1–9, 2021.

[2] G. Petasis, V. Karkaletsis, G. Paliouras, A. Krithara, and E. Zavitsanos, "Ontology population and enrichment: state of the art," in *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, 2011, pp. 134–166. doi: 10.1007/978-3-642-20795-2_6.

[3] G. Ganino, D. Lembo, M. Mecella, and F. Scafoglieri, "Ontology population for open-source intelligence: a GATE-based solution," *Software: Practice and Experience*, vol. 48, no. 12, pp. 2302–2330, Dec. 2018, doi: 10.1002/spe.2640.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Prepr. arXiv.1810.04805*, Oct. 2018.

[5] H. Alani *et al.*, "Automatic ontology-based knowledge extraction from web documents," *IEEE Intelligent Systems*, vol. 18, no. 1, pp. 14–21, Jan. 2003, doi: 10.1109/MIS.2003.1179189.

[6] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," *The 14th International Conference on Computational Linguistics*, vol. 2, 1992.

[7] D. Maynard, A. Funk, W. Peters, and R. Court, "Using lexico-syntactic ontology design patterns for ontology creation and population," in *Proceedings of the 2009 International Conference on Ontology Patterns*, 2009, pp. 39–52.

[8] J. M. Ruiz-Martínez, R. Valencia-García, R. Martínez-Béjar, and A. Hoffmann, "BioOntoVerb: a top level ontology based framework to populate biomedical ontologies from texts," *Knowledge-Based Systems*, vol. 36, pp. 68–80, Dec. 2012, doi: 10.1016/j.knosys.2012.06.002.

[9] W. IJntema, J. Sangers, F. Hogenboom, and F. Frasincar, "A lexico-semantic pattern language for learning ontology instances from text," *Journal of Web Semantics*, vol. 15, pp. 37–50, Sep. 2012, doi: 10.1016/j.websem.2012.01.002.

[10] Y. Zagorulko, G. Zagorulko, and O. Borovikova, "Pattern-based methodology for building the ontologies of scientific subject domains," in *New Trends in Intelligent Software Methodologies, Tools and Techniques*, vol. 303, 2021, pp. 529–542. doi: 10.3233/978-1-61499-900-3-529.

[11] E. Blomqvist, K. Hammar, and V. Presutti, "Engineering ontologies with patterns: The eXtreme design methodology," *Ontology Engineering with Ontology Design Patterns, Series: Studies on the Semantic Web*, vol. 25, pp. 23–50, 2016.

[12] K. Ovchinnikova, I. Kononenko, and E. Sidorova, "Development of lexico-syntactic ontology design patterns for information extraction of scientific data," in *Proceedings of the XXIII International Conference on Data Analytics and Management in Data Intensive Domains*, 2021, vol. 3036, pp. 349–361.

[13] V. Yadav and S. Bethard, "A survey on recent advances in named entity recognition from deep learning models," *Prepr. arXiv.1910.11470*, Oct. 2019.

[14] A. Gangemi and V. Presutti, "Ontology design patterns," in *Handbook on Ontologies*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 221–243. doi: 10.1007/978-3-540-92673-3_10.

[15] S. Belov, D. Zrelova, P. Zrelov, and V. Korenkov, "Overview of methods for automatic natural language text processing," *System Analysis in Science and Education*, vol. 3, pp. 8–22, Sep. 2020, doi: 10.37005/2071-9612-2020-3-8-22.

[16] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to Deep learning in natural language processing: models, techniques, and tools," *Neurocomputing*, vol. 470, pp. 443–456, Jan. 2022, doi: 10.1016/j.neucom.2021.05.103.

[17] C. Wu *et al.*, "Natural language processing for smart construction: Current status and future directions," *Automation in Construction*, vol. 134, Feb. 2022, doi: 10.1016/j.autcon.2021.104059.

[18] E. P. Bruches, A. E. Pauls, T. V. Batura, V. V. Isachenko, and D. R. Shsherbatov, "Semantic analysis of scientific texts: experience in creating a corpus and building language pattern," *Software and Systems*, vol. 34, no. 1, pp. 132–144, Mar. 2021, doi: 10.15827/0236-235X.133.132-144.

[19] A. A. Musaev and D. A. Grigoriev, "Extracting knowledge from text messages: overview and state-of-the-art," *Computer Research and Modeling*, vol. 13, no. 6, pp. 1291–1315, Dec. 2021, doi: 10.20537/2076-7633-2021-13-6-1291-1315.

[20] H. Chen and X. Luo, "An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing," *Advanced Engineering Informatics*, vol. 42, Oct. 2019, doi: 10.1016/j.aei.2019.100959.
[21] Y. Zhou, J. She, Y. Huang, L. Li, L. Zhang, and J. Zhang, "A design for safety (DFS) semantic framework development based on natural language processing (NLP) for automated compliance checking using BIM: the case of China," *Buildings*, vol. 12, no. 6, Jun. 2022, doi: 10.3390/buildings12060780.
[22] A. G. Batyrkhanov, Z. B. Sadirmekova, M. A. Sambetbayeva, A. N. Nurgulzhanova, Z. S. Ismagulova, and A. S. Yerimbetova, "Development of methods and technologies for creating intelligent scientific and educational internet resources," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 11, no. 5, pp. 2968–2977, Oct. 2022, doi: 10.11591/eei.v11i5.3075.
[23] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, *The description logic handbook*. Cambridge University Press, 2007. doi: 10.1017/CBO9780511711787.
[24] R. Axelrod, *The structure of decision: Cognitive maps of political elites*. Princeton University Press, 1976.
[25] Z. B. Sadirmekova, J. A. Tussupov, M. A. Sambetbaveva, A. S. Yerimbetova, and Y. A. Zaeorulko, "Features of the development of intelligent scientific and educational internet resources," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*, Sep. 2021, pp. 389–394. doi: 10.1109/UBMK52708.2021.9558999.
[26] B. Z. Sadirmekova, A. J. Tussupov, A. M. Sambetbayeva, and T. Z. Altynbekova, "Development of integrated information systems to support scientific activity," in *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Apr. 2021, pp. 1–6. doi: 10.1109/SIST50301.2021.9465964.
[27] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, "A survey of ontology learning techniques and applications," *Database*, vol. 2018, Jan. 2018, doi: 10.1093/database/bay101.
[28] E. Sidorova, "Ontology-based approach to modeling the process of extracting information from text," *Ontology of Designing*, vol. 8, no. 1, pp. 134–151, 2018, doi: 10.18287/2223-9537-2018-8-1-134-151.
[29] Protégé, "Protégé." 2020. Accessed: Feb. 23, 2021. [Online]. Available: https://protege.stanford.edu/
[30] L. Braginskaya, V. Kovalevsky, A. Grigoryuk, and G. Zagorulko, "Ontological approach to information support of investigations in active seismology," in *2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC)*, Sep. 2017, pp. 25–29. doi: 10.1109/RPC.2017.8168060.
[31] S. S. Kurmanbekovna, B. A. Gabitovich, S. M. Aralbaevna, S. Z. Bakirbaevna, and Y. A. Sembekovna, "Development of technology to support large information storage and organization of reduced user access to this information," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021, doi: 10.14569/IJACSA.2021.0120757.
[32] Apache Jena, "Apache Jena Fuseki," 2020. https://jena.apache.org/documentation/fuseki2/ (accessed Feb. 25, 2021).

BIOGRAPHIES OF AUTHORS

Zhanna Sadirmekova holds a Ph.D. Information Systems from L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. Successfully defended her thesis on "Development of technology for integration of information systems to support scientific and educational activities based on metadata and ontological model of the subject area". She is currently pursuing postgraduate studies at the Federal State Autonomous Educational Institution of Higher Education "Novosibirsk National Research State University" in the field of 03.06.01-Physics and Astronomy. Currently, she is associate professor of Information Systems Department at M.Kh. DulatyTaraz Regional University, Taraz, Kazakhstan. Research interests: information technology in education, artificial intelligence, search, digital library, ontology. She has more than 30 publications, including: 4 papers in Scopus base journals, 2 papers in Web of Science base, 7 papers in journals of Higher Attestation Commission of the Republic of Kazakhstan; h-index-3. She can be contacted at email: janna_1988@mail.ru.

Madina Sambetbayeva holds a Ph.D. Information Systems from L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. She is candidate of technical sciences, Associate professor of Khoja AkhmetYassawi International Kazakh-Turkish university, Department computer engineering, Faculty Engineering, Turkestan, Kazakhstan. Her research interests are related to information systems, information retrieval systems, information retrieval processes, decision support systems, linguistic support of information systems. Currently, she is working on research and development of the fundamentals of the technology for creating models and software for building distributed information systems to support scientific and educational activities, taking into account the morphology of the Kazakh language, identified in the form of electronic libraries that are consistent with international standards and trends in the development of national and international information infrastructure. Project Manager of the grant financing project no. AP05132046 "Development of the fundamentals of the technology for creating models and software for building distributed information systems to support scientific and educational activities (taking into account the morphology of the Kazakh language)". She is also involved in interstate projects, working together with scientists from Novosibirsk State University and the Institute of Computational Technologies SB RAS. She has over 50 publications, of which: 1 educational-methodical manual; 11 papers in Scopus, 16 papers in the journals of the list of Higher Attestation Commission of the Republic of Kazakhstan and the Russian Federation, 5 certificates of official registration of computer programs used in pedagogical and scientific practice. H-index-3. (1 wage-rateas LRA). She can be contacted at email: madina_j@mail.ru.

Sandugash Serikbayeva She accomplished her Ph.D. degreein specialty Information Systems at L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. Dissertation theme is “Creation of models and technologies for building distributed information systems to support scientific and educational activities”. Scientific interests: distributed information system, thesaurus, information retrieval, digital library, ontology. Has more than 30 publications, including: 1 academic book; 8 papers in SCOPUS base journals, 3 papers in Web of Science base, 6 papers in the journals of Higher Attestation Commission of the Republic of Kazakhstan, and the Higher Attestation Commission of the Russian Federation. Scopus H-index-3, Web of Science H-index-1. She can be contacted at email: Inf_8585@mail.ru.

Gauhar Borankulova Candidate of technical sciences, Associate professor and Head of Information Systems Department at M.Kh. DulatyTaraz Regional University. She has more than 60 scientific papers, including 7 works in the rating publications Web of Science and Scopus, h-index-2. Research interests: fiber-optic technologies, microprocessor systems, information systems. Department of Information Systems, Faculty of Information Technology. Taraz, Kazakhstan. She can be contacted at email: b.gau@mail.ru.

Aigerim Yerimbetova She is a Ph.D., Candidate in Technical Sciences, associate professor, Institute of Information and Computational Technologies CS MES RK, Almaty, Kazakhstan; and Satbayev University, Almaty, Kazakhstan. Currently, she is engaged in the research and creation of linguistic algorithms for semantic search and retrieval of information using Link grammar, the Link Grammar Parser software package based on it, and mathematical and logical methods. She is the author of more than 70 papers on computational linguistics and the development of information systems: 1 textbook, 2 monographs, 19 papers in the journal SCOPUS, 14 papers in the journals of the CQAES list; 1 patent, 4 certificates of entry in the State Register of Copyright Rights. She can be contacted at email: aigerian@mail.ru.

Aslanbek Murzakhmetov received Ph.D. degree in 2022 from Department of Information Systems, al-Farabi Kazakh National University in specialty Information Systems. Currently, she is associate professor of Information Systems Department at M.Kh. DulatyTaraz Regional University, Taraz, Kazakhstan. He has more than 20 scientific papers, including 4 works in the rating publications Web of Science and Scopus, a certificate of deposit of intellectual property objects, as well as certificates “On entering information into the state register of rights to objects protected by copyright”. He was a junior researcher of the project “Construction, research and substantiation of non-classical neural network models for solving recognition problems with standard information and their applications for building intelligent systems” of al-FarabiKazNU Research Institute of Mathematics and Mechanics. He was the local coordinator of the LMPI 573901-EPP project-1-2016-1-IT-EPPKA2-CBHE-JP “Bachelor’s degree and Professional Master’s degree for the development, administration, management and protection of computer networks in enterprises” within the framework of the Erasmus+program. Conducted research work with leading scientists of the University of South Florida (Tampa, Florida, USA) and Novosibirsk State University (Novosibirsk, Russia). Research interests: development of mathematical models and algorithms for pattern recognition and classification; optimization systems, big data processing, multi-agent systems, stochastic programming methods, methods of operations research. He can be contacted at email aslanmurzakhmet@gmail.com.