Abstract: The aim of this study is to develop an algorithm for automatic representation of the text of natural language as a formal system for the subsequent automatic extraction as reasonable answers to profound questions in the context of the text, and the deep logical consequences of the text and related areas of knowledge to which the text refers. The most universal method of constructing algorithms of automatic treatment of text for a particular purpose is a representation of knowledge in the form of a graph expressing the semantic values of the text. The paper presents an algorithm of automatic presentation of text and its associated knowledge as a formal logic programming theory for sufficiently strict texts, such as legal texts. This representation is a semantic-syntactic as the causal-investigatory relationships between the various parts are both logical and semantic. This representation of the text allows to resolve the issues of causal-investigatory relationships of present concepts, as methods of the theory and practice of logic programming and methods of model theory as well. In particular, these means of classical branches of mathematics can be used to address such issues as the definition and determination of consequences and questions of consistency of the theory.

Keywords: Text, sentence, syntax, semantics, consequently, cause, diagram, theory, logic programming, automation, formalism

1 Introduction

You can bring a variety of requirements for computer processing of text in natural language. Simple systems must be able to present the facts and statements contained in the text. More sophisticated systems must be able to explain the meaning of the concepts contained in the text or accompanying domain of knowledge of the text and detect connections with such concepts. Even more sophisticated systems should be able to do all that less complex systems can, and to be able to find the logical consequences arising from the content of the text and from the domain of knowledge that is related to the text and implied. The most important requirements to the most sophisticated system of this kind is the ability to solve problems of consistency, solvability and completeness. Ideally, such a system could be the theory of predicate calculus, corresponding to text and the accompanying domain of knowledge, and it can actually be created but in the general case, it remains ineffective in practice, because of the ineffectiveness of the implementation of the logical inference. Thus, the only way to use the features of the text, while using the features of the text, namely, the degree of complexity of statements and the method of using undetermined objects, i.e. ways of using quantifiers and variables, it is possible to hope for the success of the construction of a complex system.

We propose a way to automate the in-depth understanding of some texts in natural language. In-depth understanding of the text $T$ is the identification of causal-investigatory relationships of the concepts in $T$ with knowledge area to which refers the $T$. Thus, for the implementation of this understanding, it is necessary to imagine an automatic way the text and the accompanying domain (area) of knowledge in the form of a formal system $P$, such that it contains mechanisms to extract the consequences and finding reasons (antecedents), the provisions under consideration. The system $P$ can be developed from start to finish as it had been done in some basic work on artificial intelligence [1,2], other approaches can be viewed in [3]. In this paper, we will continue to develop the approach [4]. We propose an algorithm of automatic conversion of some
natural language texts to formalism of logic programming $P_0$.

2 Algorithm of automation of knowledge representation in the form $P_0$

2.1 Texts and conversion

We consider the texts of certain kind. The proposed $\tau$ of text $T$ has one of the following types.

1. $\tau$ – in the positive or negative sentence, expressing one main action or positive of the actions that can be expressed in $\tau$ a verb-predicate (action), or a group of words. Furthermore, the sentence $\tau$ contains an object-subject that performs the action (agent), expressed by a word -subject or a group of words. The remaining words are called the result (result) of the actions of the agent. The result can have an independent sense or gain a sense in combination with the action and the agent. Example: "I’m going". The sentence $\tau$ may contain indefinite object, then it means that it is any instance of a certain class of objects referred to. Example: "Human life is the highest value." It is meant by the undefined person of human class of any specimen of human class.

2. $\tau$ has one of the following types: $\tau_1$ and ... and $\tau_n$; $\tau_0$ if $\tau_1$ and ... and $\tau_n$, where $\tau_i$ has type 1.

It should be noted that Type1 sentence $\tau$ may contain sentences from $T$ Type1 in a group of agent or result.

We also note that the described structure of the sentences from $T$ can be built automatically. The text $T$ can be represented as a set of expressions $P \leftarrow P_1 \ldots P_n$, where $P, P_i, n \geq 0$ are sentences of Type 1. We guess that when $n = 0$ – that is a structure $P$, and Type 2 sentences – $\tau_1$ and ... and $\tau_n$ are represented by a set $P_1, \ldots, P_n$. Since all undefined objects in $\tau$ are treated as arbitrary instances from instances of class, the text $T$ can be presented in the form of logic programming theory. As an illustration of this approach, we take the text of the Constitution of the Republic of Kazakhstan as $T$. We assume that $T$ has the properties 1 and 2 due to the fact that the laws in this document are first universal, and second are structured as a set of sentences of type 1 or a set of conditions and a conclusion, where each of the conditions and the conclusion is a Type 1 sentence, those have the structure of the implication of Horn type. Here we suggesting that constitutional laws can not have an alternate conclusions that representation of the law as the implication is not the conclusion of the disjunction implication statements. On this basis, set out below is a universal representation for the given class of text.

2.2 $P_0$-scheme K.1.1.1.

Structuring of sentence, as it was described above, can be represented by $P_0$-scheme. As an example we will look at $P_0$-scheme (Figure 1) of the first point of the first article of the first section of Republic of Kazakhstan.

Article K.1.1.1. "The Republic of Kazakhstan proclaims itself a democratic, secular, legal and social state whose highest values are an individual, his life, rights and freedom".

Designations: "K" = Republic of Kazakhstan = Kazakhstan; the article says about the man, his life, rights and freedoms. An individual is non-specific and therefore he can be any instance of class of Individual. Therefore, in the diagram it is denoted by a variable, for example, by $X$. In $P_0$-scheme variables are denoted by single capital letters without quotation marks.

$P_0$-scheme K.1.1.1.

2.3 Linear representation of sentence K.1.1.1.

Representation of a scheme easily recorded linearly:

- \begin{itemize}
  - action ([claims]), agent (["K"])
  - result ([democratic state "K", secular state "K", legal state "K", social state "K", action ([are]),
  - agent ([individual $X$, life of individual $X$, rights of individual $X$, freedom of individual $X$])
\end{itemize}
3 The semantic and syntactic representation K.1.1.1.

To deepen study K.1.1.1., we can specify cause-consequence relationships (semantic and syntactic) for \( P'_0 \)-rules \( A_i \) and \( B_j \).

3.1 \( A_1 \) - a consequence of knowledge

Referring to the \( A_1 \). The text \( T \) does not define a democratic state, but in the domain of knowledge jurisprudence, to which the text refers there is a common statement - "state is democratic if the only source of state power in it is the people." Note the word state; here it refers to an instance of any state concept. This sentence belongs \( T \). Furthermore, Article K.1.3.1. says - "the only source of power is the people." The context \( T \) indicates that the sentence in its entirety - the only source of state power in the "K" is the people of the "K". \( P'_0 \)-are the rules of these sentences.

\[ A_1 \] if

\[ \begin{align*}
&[ \text{action} ([\text{is}]), \text{agent} ([\text{people X}]),] \\
&[ \text{result} ([\text{the only source of state power of the X}])], \\
&[ \text{result} ([\text{the only source of state power "K"}])].
\end{align*} \]

In (3), \( X \) plays the role of a variable, and designates any State, i.e. variable \( X \) is connected by universal quantifier, so according to the rules of logic \( A_1 \) is a consequence of (3) and (4).

3.2 The semantic and syntactic representations relating to the \( A_1 \)

Let's represent semantic and syntactic expression "the only source of state power 'K'" in the form of \( P'_0 \)-rules

\[ \begin{align*}
&[ \text{action} ([\text{is}]), \text{agent} ([\text{people "K"}])], \\
&[ \text{result} ([\text{source of state power "K"}])] \text{ if } (4) \\
&[ \text{result} ([\text{source of state power "K"}])] \text{ and } \\
&[ \text{result} ([\text{the only source of state power "K"}])]
\end{align*} \]

Thus, the expression "the only source of state power" appears in the form of its consequences. Note that (5) and (6) are \( P'_0 \)-rules.
3.2.1 Semantics representation the concept 'source'

Next, imagine the semantic concept of "source of state power."

\[
\text{[action ([can produce]), agent ([X]), result ([Y])]} \quad \text{if}
\]
\[
\text{[action ([is]), agent ([the X]), result ([source Y])].}
\]

This \(P'_0\)-rule is the concept of "source Y" as antecedent, but this concept can be represented as a consequent

\[
\text{[action ([is]), agent ([the X]), result ([source Y])]} \quad \text{if}
\]
\[
\text{[action ([can produce]), agent ([X]), result ([Y])].}
\]

3.3 \(P'_0\)-rule of representation \(A_2\)

Referring to \(P'_0\)-rule \(A_2\):

\[
\text{[action ([claims]), agent (["K"]), result ([secular "K" state])].}
\]

In the text \(T\) there is no definition of a secular state, but in the field of knowledge jurisprudence, to which the text belongs, has common statement - "Is secular state if religious associations and the state does not interfere in each other's affairs". Pay attention to the concept of the state and religious associations, are each of them is any instance of corresponding concept. In this definition, two parts of implications are equivalent, so it really is an equivalence and therefore sufficient to establish how the statement "religious associations and the state does not interfere in each other's affairs" will automatically be extracted from \(T\).

3.3.1 \(P'_0\)-rule and semantics of non-interference

\(P'_0\)-rule of statement "religious associations and the state does not interfere in each other's affairs".

\[
\text{[action ([not interferes]), agent ([religious associations of the state X and state X]), result ([in each other's affairs])].}
\]

Next, consider the need to provide for the deduction of the previous statement

Article K.1.5.2. Public associations shall be equal before the law. \text{Illegal interference of the state in the affairs of public associations and of public associations in the affairs of the state}, imposing the functions of state institutions on public associations shall not be permitted.

\(P'_0\)-rule of selected part.

\[
[\text{action ([not allowed]), agent ([Illegal interference of the state "K"]), result ([in the affairs of public associations "K"])}].
\]

From K.1.5.2 and statement of the "Religious associations are public associations of the state," it follows that "not allowed the illegal interference of the state in the affairs of religionzyh associations and illegal interference of the religious associations in the affairs of the state." Here give the \(P'_0\)-rules to implement this conclusion.

\[
\text{[action ([not allowed]), agent ([Illegal interference of the state "K"), result ([in the affairs of religious associations "K"])]}
\]

\[
\text{[action ([not allowed]), agent ([Illegal interference of the state "K"), result ([in the affairs of public associations "K"])]}
\]

\[
\text{[action ([not allowed]), agent ([Illegal interference of the religious associations "K"), result ([in the affairs of the state "K"])]}
\]

Similarly appears the statement to the contrary of illegal interference.

\[
\text{[action ([not allowed]), agent ([Illegal interference of the religious associations "K"), result ([in the affairs of the state "K"])]}
\]

\[
\text{[action ([not allowed]), agent ([Illegal interference of the public associations "K"), result ([in the affairs of the state "K"])]}
\]

3.3.2 Conditions of non-interference of the two subjects in each other's affairs

The statement "religious association and the state does not interfere in each other's affairs" contains two subjects for which not allowed illegal interference in each other's affairs. On the other hand, about the inadmissibility of illegal interference in the affairs of each other for certain subjects referred to in K.1.5.2.. So we take the conventional
equivalent to the phrase "Two subjects does not interfere in each other’s affairs, if not allowed to illegal interference with one of them in another case". We will deduction the statement "not allowed the illegal interference of one of them in another case" for the subjects the religious associations in the "K" and "K" state based on the text $T$ and associated concepts.

$P_0'$-rule "two subjects does not interfere in each other’s affairs, if not allowed illegal interference of each of them in the affairs of another"

\[
\text{action} ([\text{not interferes}]), \quad \text{agent} ([X \text{ and } Y]), \\
\text{result} ([\text{in each other’s affairs}])
\]

if

\[
\text{action} ([\text{not allowed}]), \\
\text{agent} ([\text{Illegal interference } X]), \\
\text{result} ([\text{in the affairs of } Y])
\]

and

\[
\text{action} ([\text{not allowed}]), \\
\text{agent} ([\text{Illegal interference } Y]), \\
\text{result} ([\text{in the affairs of } X])
\]

3.3.3 The semantic and syntactic representation of statement "Not allowed the illegal interference the $X$ in the affairs the $Y$"

For the practical determination of the truth of the statement "not allowed the illegal interference in the affairs of the subject $X$ in in the affairs of $Y$", it is necessary to adequately express the semantics of this statement. Obviously, the best way to be the next semantics disclosure. If in a fixed manner to the subject $X$ create a list SSDX consisting of lists of legal subjects for each of the legal acts, in which the $X$ is legal subjects, then lists SSDX and SSDY and have no common elements.

$P_0'$-rule list DY of all acts of law in which $Y$ is legal subject and in any of the acts of law from DY not allowed illegal interference the subject $X$.

\[
\text{action} ([\text{not allowed}]), \\
\text{agent} ([\text{Illegal interference } X]), \\
\text{result} ([\text{in the affairs of } Y])
\]

if

\[
\text{action} ([\text{have}]), \quad \text{agent} ([Y]), \\
\text{result} ([\text{list DY of acts of law with legal subject } Y])
\]

and

\[
\text{action} ([\text{have not}]), \quad \text{agent} ([\text{SSD X and SSDY}]), \\
\text{result} ([\text{common elements}])
\]

3.3.4 The semantic and syntactic representation statement "Public associations shall be equal before the law"

In terms of the lists are built $P_0'$ –rules statement "Public associations are equal before the law."

Here, as the above we will build $P_0'$ –rules equivalent to the statement "public association $X$ and public association $Y$ are equal before the law."

\[
\text{action} ([\text{equal}]), \quad \text{agent} ([X \text{ and } Y]), \\
\text{result} ([\text{before the law}])
\]

if

\[
\text{action} ([\text{have}]), \quad \text{agent} ([Y]), \\
\text{result} ([\text{list DY of acts of law with legal subject } Y])
\]

and

\[
\text{action} ([\text{have}]), \quad \text{agent} ([DY]), \\
\text{result} ([\text{list SSDY of lists of subjects of legal subjects in the affairs of DY, written out sequentially for each affair from DY}])
\]

and

\[
\text{action} ([\text{have not}]), \quad \text{agent} ([\text{SSD X and SSDY}]), \\
\text{result} ([\text{common elements}])
\]
subject X])
and
[action ([ have]), agent ([ DX],
result ([ list SSDX of lists of subjects
of legal subjects in the affairs of DX,
written out sequentially for each
affair from DX])]
and
[action ([ built]), agent ([ SSDY],
result ([ from SSDX of replacing X by Y and
Y on X])].

4 Conclusion

Semantic and syntactic representations are constructed similarly in the form of $P_0$–rules for $A_i, B_j$. And altogether these rules are the semantic and syntactic representations of K.1.1.1, K.1.3.1. and related to them sentences $T$ and associated concepts.

References