SEMANTIC AND ONTOLOGICAL RELATIONS IN AIIRE NATURAL LANGUAGE PROCESSOR

Alexey Dobrov

***Abstract**: AIIRE is a free open source natural language processor, developed by a team of researchers in Saint-Petersburg, Russia. AIIRE is an implementation of full-scale NLU process, based on the method of inter-level interaction and rule-based disambiguation. Semantic graphs that are built by AIIRE are based on the involved ontology. The rules that concern correspondence between semantic relations (used in semantic graphs by AIIRE) and conceptual relations (used in ontology) is a matter of discussion. Semantic graphs are evaluated from syntactic trees, and, in general, although word-independent syntactic constituent classes tend to denote rather abstract relations (cf. genitive construction in general), the instances of those classes (specific phrases) in theory may denote any subclasses of those relations. The developed algorithm of choosing a relation subclass in each case is also a matter of discussion.*

***Keywords**: ontological semantics, lexical disambiguation, NLU, conceptual relations, semantic graphs.*

***ACM Classification Keywords**: 1.2.7 – Natural Language Processing: Language Parsing and Understanding, Text Analysis*

Introduction

AIIRE is a free open source natural language processor, developed by a team of researchers in Saint-Petersburg, Russia. The acronym ‘AIIRE’ stands for ‘Artificial Intelligence-based Information Retrieval Engine’, that was the first production-level application of the developed NLU-kernel. The team was formed in 2003-2005, in the Laboratory for Informational Linguistic Technologies of the Institute of Linguistic Studies in Saint-Petersburg, Russia, and continues its work as the RnD department of Geline company.

The method of inter-level interaction

AIIRE natural language processor is an implementation of full-scale NLU process, based on the method of inter-level interaction and rule-based disambiguation. The method of inter-level interaction was first proposed by G.S. Tseitin in 1985 [Tseitin, 1985], but was not implemented since then because of its complexity, and because of lack of well-developed linguistic software and high-performance hardware. The basic idea of the method was to get rid of the artificial separation of levels of linguistic analysis and to analyze morphology, syntax, and semantics in the same time. This way of analysis allows to perform disambiguation on lower levels using the upper-level rules immediately after the ambiguity arises on the lower levels, rather than after the whole text (or sentence) is analyzed on these levels. E.g., morphological disambiguation can be performed using results of syntactic binding immediately after the first two terms of the syntactic tree are bound (or not bound) according to restrictions of the involved grammar. Furthermore, if two combinations of morphological analysis hypotheses still remain after syntactic binding, i.e., if both combinations give grammatical syntactic trees, then these trees are analyzed in terms of semantics immediately after the trees are produced, semantic restrictions acting as filters to reduce ambiguity on the level of syntax.

The idea of inter-level interaction helps to prevent (or, at least, to reduce drastically) combinatorial explosions, which is crucial for NLU performance. Formal grammars produce a plenty of ambiguities during the analysis, especially when ellipsis is allowed. Moreover, if each textual wordform has just two interpretations, then traditional separation of the levels of analysis leads to 1024 combinations having to be considered just to perform syntactic analysis of a 10-word sentence, each combination having a set of hypotheses of syntactic binding. Dozens of thousands of grammatical syntactic trees are produced, and just a few of them remain after the semantic analysis. The idea of inter-level interaction allows to apply the restrictions of the highest levels of analysis much sooner, reducing the maximum amount of combinations to tens or rarely to hundreds.

Rule-based and Machine-learning approaches to disambiguation

Nowadays, statistical heuristics are much more popular than rule-based methods of disambiguation. Corpus-based approaches had a great success in morphological disambiguation (95% quality, cf. [Hajic et al. 2001]), which seemed a breakthrough in the field. Corpus-based (so-called machine-learning, ML) approaches seem to be essentially more simple and objective than rule-based methods, and are quite efficient for many tasks. The problem remains, however, that statistical heuristics never guarantee the absence of false-negative results (i.e., correct hypotheses being merely culled), which can have catastrophic consequences even when morphological analysis is followed only by syntactic parsing, as the whole syntactic trees are lost. Formal grammars are based on strict rules of grammatical coordination and government, of ellipsis and word-order, and loss of a correct hypothesis of morphological analysis often leads to a complete failure yet on the level of syntax. That is why AIIRE does not use any statistical heuristics to perform disambigution, although further research may help to develop some corpus-based mechanisms that guarantee absence of false-negatives. The same is true, however, not only for NLU-systems like AIIRE, but also for efficient implementations of spell-checkers and punctuation checkers (cf. [Petkevič, 2006]), because of the same inadmissibility of false-negatives.

The main source for disambiguation in AIIRE is its ontology. Grammatical restrictions help to reduce (but not to eliminate) ambiguities on the level of morphology, but grammar just very rarely helps to get rid of homonymy (it happens in a few cases when homonyms have partially different paradigms, e.g., Russian word '*лист*' has plural '*листы*' to denote 'sheets' and '*листья*' to denote 'leaves'), and never helps to choose between word meanings (more precisely, it is a convention in the AIIRE project, that if two different meanings can be distinguished only by means of grammatical context, then these meanings belong to different homonyms of the same word).

AIIRE ontology, semantic dictionaries, thesauri and knowledge bases

The database that contains lexical meanings is called ontology in AIIRE project, but it also can be called (because it provides functionality of) semantic dictionary, as it was in [Leontyeva, 2006], or thesaurus, because it represents main inter-meaning relations, which are normally registered in thesauri, or even a knowledge base, because it provides knowledge not only on conceptual classes, but also on their instances, at least on those registered in Wikipedia. Nevertheless, the term 'ontology' was chosen for the following reasons:

- Semantic dictionaries are principally language-dependent, whereas ontologies are not. AIIRE ontology contains lexical meanings, but it also contains even more concepts that can not be bound to specific lexical entities, sometimes, even to expressions of any languages. These concepts are parts of meanings or superclasses like 'Object that is localizable in any-dimensional linear space, and therefore has dimensions, and therefore has size', which are necessary for restricting semantic valencies of meanings of lexical entities like '*big*', '*small*', '*at*', '*in*', '*within*' and many others to a very abstract class of objects, which contains physical objects, geometric figures, parts of images, of texts, and even of melodies. If object has size, then it can be *small* or *big*, no matter whether this object is physical or virtual, and no matter how many dimensions it has. If object is localizable in any linear space, then it can be located *in* another object in the same space, and they both have size, no matter which kind of space is in question — three (or four-) dimensional physical space or one-dimensional space of text. Abstract concepts like these are never registered in semantic dictionaries, but certainly are registered in ontologies under the code names like 'Localizable' and 'Sized', which are not to be confused with similar lexical entities.

- Thesauri like WordNet (cf. [Miller, 1995], [Fellbaum 1998]) are certainly kinds of semantic dictionaries, although they are restricted to inter-meaning relations like *hyponymy* or *synonymy* and a set of others. Thesauri do not always distinguish between class-to-class and instance-to-class types of inheritance, and therefore sometimes contain encyclopedic knowledge like specific instances (hyponyms) of abstract classes (hypernyms). However, thesauri never contain conceptual relations like 'can perform action' (relation between meanings of nouns and verbs, that reflects ability of instances of class denoted by noun to perform actions that are instances of class denoted by verb). These relations are mostly driven by extralinguistic knowledge (e.g., ability of physical objects to move in physical space, as disability of, e.g., collections of data to do so, can never be deduced merely from language), and, in the same time, are crucial for lexical disambiguation (e.g., sentences like *The table was moved to the corner of the room*, where the word *table* can mean both an object of furniture (a physical object), and a collection of data (e.g., *database table*) can be disambiguated only because of the knowledge that only physical objects can move in physical space).
- Knowledge base is, probably, more proper term to denote AIIRE ontology than dictionary or thesaurus, but still is not specific enough to reflect the fact that the items stored in the ontology are models of concepts, which form not only an inheritance hierarchy, but also an extensive network due to the above mentioned relations. Ontologies are sometimes even treated as kinds of knowledge bases [Knowledge Base, 2014], which, in contrast to other kinds of knowledge bases, have hierarchical structure.

Probably, the most significant difference between AIIRE ontology and semantic dictionary, thesaurus or knowledge base of any kind is that relations that constitute this ontology are concepts themselves and form their own inheritance hierarchy. The same is true for some other ontologies, e.g. OpenCyc, but is not an obligatory requirement to the structure of ontology (e.g., SUMO does not follow this practice, cf. [Zouaq et al. 2009], [Sheffczyk et al. 2006]). This peculiarity of AIIRE ontology is, however, just a side-effect of one of the conventions adopted in AIIRE project, which states that every entity used in AIIRE ontology must be a concept defined in this ontology. This convention also leads to lexical entities (even of different languages) being treated as concepts that are connected with their meanings with the so-called 'to denote object'[2] relation. More than that, each concept, except for those corresponding to lexical entities themselves, must be bound to a lexical entity (which may be rather a large expression), even if natural language doesn't have

[2] Relations are stored as meanings of non-idiomatic natural language expressions, that are formed according to the following convention: an infinitive verb phrase that describes the relation + a noun phrase that describes its object class. Subject classes are described in braces within the description of the relation concept itself.

words or idioms that denote this concept. This requirement allows to present ontology to its editors in form of a dictionary, which is much more common form of presentation for linguists.

The above-mentioned hierarchy of conceptual relations is the main source of restrictions for lexical disambiguation. This hierarchy is linked with other hierarchies: each relation is linked with its subject and object classes (there are relations named ‘to have subject’ and ‘to have object’, that mark these links), and with its so-called ‘refraction’ (inverse relation) thru ‘to be refracted with relation’ relation. Relations are a priori coventionally divided into direct and reverse ones, according to the order of evaluations of natural language expressions that may lead to these relations. The choice for each relation is sometimes unobvious, because sometimes both direct relation and its refraction can be expressed in natural language without inversions or passivizations, e.g., both ‘to have a pet’ and ‘to belong as a pet to somebody’ relations can be expressed with genitive constructions: cf. *the dog of Mary* and *the owner of the dog*. In such cases, inheritance hierarchy is involved: in the above-mentioned example, ‘to have a pet’ relation is a subclass of ‘to have an object’, which, in turn, has a subclass that is one of the meanings of preposition *with* (cf. *a girl with a dog*, *a girl with long hair*, *a girl with high IQ* etc.) and therefore is direct. Thus, ‘to have a pet’ is considered also direct, because directness of the superclass is inherited.

Rules of relation inheritance and overriding

In AIIRE ontology, model of concept is a set of attributes, each attribute being a relation-object pair. Only direct relations of concept are stored in the ontology, their refractions being evaluated and presented to users on a par with direct ones. Directness of relations is also used in conceptual graph normalization procedure: reverse relations are converted into direct ones. This is the only reason why relations have to be a priori conventionally divided into direct and reverse: it is obvious from the definition of relation, that both a relation and its refraction can be chosen to be treated as direct or reverse.

It is also necessary to mention, that there are strict rules of inheritance and overriding of relations in AIIRE ontology, that are crucial for the mechanism of semantic restrictions. If a concept inherits another concept (i.e., it has an attribute with ‘to inherit concept’ as relation and the latter concept as object), then every attribute of inherited concept is primarily treated as an attribute of of the inheriting one. In the same time, each attribute of the inherited concept can be overridden by one or more attributes of the inheriting one. One attribute overrides another one if and only if relation of the overriding attribute inherits or coincides with relation of the overridden one, and the same is true for the objects of these relation. E.g., concept ‘span’ has an attribute <‘to have size’, ‘length’> (i.e., size of a span is its length), whereas concept ‘time span’, although it inherits ‘span’, has an attribute <‘to have size’, ‘duration’>, which means that size or length of a time span is duration. Overriding attributes can be acquired thru overridden relation-object pair, but, due to overriding, the range of possible attribute values is restricted on each step of inheritance. As relation between instances and their classes is also a kind of inheritance, overriding rules apply to the instances. Furthermore, there is a significant limitation for instances, which states, that all attributes of an instance must override at least one of the attributes of its class. Thus, the sentence *The size of time span was two minutes* does meet these semantic restrictions, whereas *The size of time span was two meters* violates these restrictions.

Conceptual and semantic relations. What is conceptual relation?

As an NLU system, AIIRE produces semantic graphs as final representations of textual semantics. These graphs are completely built from the concepts defined in the ontology. The term ‘semantic graph’ is used here in a very wide sense, not as a synonym of ‘conceptual graph’ J. Sowa had proposed in [Sowa 1999]. It is maybe even better not to call this representation a graph, because its edges can act as vertices and have their own edges, which is, however, not forbidden in terms of mathematics. Both vertices and edges of semantic graphs are either concepts from the ontology, or, more frequently, textual concepts that must be subclasses of the ontological ones and have relations with each other. Textual concepts must follow inheritance and overriding rules of ontology. Edges of semantic graphs must be relations, vertices can be concepts of any kind. Not all ontological relations are involved in semantic graphs. For this particular reason, it seems expedient to distinguish between conceptual relations (any relations between concepts) and semantic relations (conceptual relations that can be edges in semantic graphs). In order to define the term 'conceptual relation' more precisely, a mathematical definition of relation can be used: relation is a subset of direct product of several sets. In the case of AIIRE ontology, relations connect conceptual classes or instances, and are binary, thus, it is better to define conceptual relations between classes as classes of pairs of instances of two classes, relation instances being particular pairs. Some relations can be deduced from other relations as their superpositions, in which cases they are called implicational. It is also possible to bind relations with predicates, and thru predicates even with algorithms that evaluate these predicates (e.g., the relation named 'to occur before situation' between two situations can be deduced from time points of these situations and from comparison of their numeric representations in seconds since epoch), but this way of modeling conceptual relations is still in the earliest stages of development.

Inheritance hierarchy of relations. Relations of concept 'concept'

As hierarchy of relations is linked with their subjects and objects, relations can be classified by their participants. Some relations are linked with the root of the ontology (concept 'concept') as subject, which means that any concept can be subject of these relations. E.g., there are two existence-marking relations between any concept and any reality ('to exist in a reality' and 'not to exist in a reality'), two kinds of equivalence relations between any concepts ('to be / not to be equal to concept' and 'to be like / unlike concept'), quantifier-binding relation 'to be bound to quantifier', the refractions of some previously mentioned relations ('to be / not to be subject of relation', 'to be / not to be object of relation'), and markers of their implications ('to be / not to be subject of situation', 'to be / not to be object of situation'). These relations are rather abstract, and in the majority of cases only their subclasses are actually involved into evaluation of semantics, e.g., 'to be subject of action', 'to be object of relation' etc. These relations are necessary for disambiguation of subject-predicate and verb-object constructions, prepositional phrases and their combinations with nominal and verbal phrases, and some other kinds of constructions.

Variable and constant objects and their relations

Some relations are linked with a bit less abstract concept 'variable object' as possible subject. This class corresponds to any object that can change in time (and, therefore, appear or disappear), in contrast with constant objects. Variable objects can have states and are linked with their states with corresponding relations ('to have a historical state', 'to have a contemporary state', etc.), they can be created (invented (but not dicovered), produced, built, born etc.) by someone ('to be created by someone'), at some time ('to be created at year/century/millenium'), with some instrument ('to be created using object'), etc. The majority of nominal meanings correspond to variable objects, nevertheless, in some cases it is important to distinguish between constants and variables to perform lexical disambiguation. E.g., in the phrase *The square was built in eighteenth century* the word *square* is ambiguous: it can denote both an area in an inhabited locality and an equilateral quadrangle. The latter is an abstract geometric figure which is not variable (quadrangle is a geometric set of points, and sets are constants), and thus can not be created or built. Therefore, the ambiguity is resolved only by means of the constant-variable distinction. It is worth mentioning, however, that in a very special context the word *square* can mean not the quadrangle itself, but an image of quadrangle depicted by someone; images are variables and therefore can be created. Furthermore, the verb *to invent* can denote (again, in a very special context) 'to discover'. These contextual meanings are produced by the mechanism of metonymy, which is enabled in AIIRE only in cases when none of the word meanings meet the semantic requirements of the context. This mechanism is described lower; it is based on the so-called 'backbone relations', e.g., geometric figures have a backbone relation 'to be depicted on a picture' with physical pictures, so that, e.g., the sentence *The triangle was built in eighteenth century* has an interpretation which assumes that the word *triangle* metonymically means a picture of triangle. Practically each constant concept has backbone relations with variable concepts, therefore constant-variable distinction rarely reduces ambiguity, but helps choosing between immanent word meanings and meanings of their metonyms.

Physical objects, physical places, and their relations

One of the most significant base classes of AIIRE ontology is 'physical object'. Physical objects are classified as variable triple-dimensional bodies (bodies are non-empty localizable objects) and inherit relations of these superclasses, but these relations are overridden by immanent attributes of physical objects themselves. In particular, physical objects can be parts of other physical objects ('to be / not to be part of physical object'), they can move in physical space in some directions ('to move in physical space in direction'); physical objects can be located or situated in physical space ('to be located in physical place'), inside (outside, behind etc.) other physical objects; physical objects have physical attributes, such as taste, odor, color, form, temperature, viscosity, toughness, weight etc. All these attributes are essential for lexical disambiguation, because lexical ambiguity very often appears as distinction of physical and non-physical meanings. E.g., the famous phrase *colorless green ideas* can be interpreted only as boring (*colorless*) ideas of environmentalists (*green*), whereas the words *colorless* and *green* certainly have meanings that correspond to physical characteristics of color, but the word *ideas* has no meanings that correspond to any physical object. The same mechanism helps to perform disambiguation in genitive constructions or (in languages without inflectional genitive) in constructions with prepositions like *of* (cf. *the foundation of the theory / the foundation of the building*), in constructions with locative prepositions (cf. *a chair in the room / a chair in physics*), and in any constructions referring to motion (cf. *oncoming train / oncoming event*), etc.

Physical objects are not to be confused with physical places, although physical objects have backbone relations with the places where they are located and therefore produce metonymy. Like physical objects, physical places are triple-dimensional and localizable, but they are not necessarily variable, and, what is the most important, they are not bodies. Places themselves do not have physical characteristics such as color or weight, but due to reverse backbone relations with physical objects contained in them, they are sort of pretending to have temperature, color and some other physical attributes (cf. *warm green place*, which is to be interpreted as 'a place where the majority of physical objects are warm and green'). Countries and other inhabited localities are variable places, which have (at each moment of time) backbone relations to constant places (e.g., Germany is a variable place, which can change its borders, but its contemporary borders correspond to a constant place, which had existed even before Germany arose as a state and even before mankind arose as a biological species). Inhabited localities also have backbone relations with their societies (these relations are likely to come from more abstract backbone relations with contained physical objects mentioned above, but it is not necessarily true), so that metonymy also appears in phrases like *Germany has voted for Angela Merkel*. Physical places have a few immanent relations, most of them overriding those of abstract places (parts of any-dimensional spaces) and more abstract concepts. Physical places can be situated near other physical places (but not, e.g., near places in melodies), they can be scenes of some events or processes, they may have other physical places as parts (but not physical objects), and therefore can be parts of other physical places themselves.

Processes, actions, states, activities, subject domains and their relations

Another base-class concept which is very important for semantic analysis is 'process'. Processes are treated as contiguous sequences of situations. As sequences, processes have beginnings and endings, and therefore can begin or end. Because processes are contiguous, processes can last or break. Sequences are subclasses of aggregates, and, because inclusion relation 'to include concept' of aggregates are backbone, processes always produce metonymy with the situations that constitute them. Situations, in turn, can not last like processes (so that the expression *situation lasts* is interpreted only thru reverse metonymy with process), but they are linked with time spans (again with a backbone relation), with their participants (subjects and objects) and with relations (states or actions) between them. Situations can be atomic, in which case they are constants, they can be real or hypothetical, and they are not immanently localizable in space, but have an implicational relation 'to take place' with physical places, which correspond to localizations of situation participants. As the relation 'to correspond to relation' of situations is backbone, situations also produce metonymy with their relations, which allows expressions like *signing of the contract by the committee on Monday in Warsaw*. Situations may have reasons and purposes, and the same applies to processes because of metonymy. Processes may correspond to activities ('to correspond to activity', 'to have an activity as a purpose'); because of metonymy with time spans, processes can take place before, after or simultaneously with other processes; because processes are sequences, they can be parts of other processes and have subprocesses; furthermore, they can correspond (or not correspond) to processual patterns (so that they can bind with adjectival meanings like 'correct' or 'usual' and their opposites).

The hierarchy of processes in linked with the hierarchy of subclasses of the concept 'to perform action / to be in a state', which is subdivided into imperfective and perfective sub hierarchies (but only for actions), so that there are three (for actions —four) isomorphic hierarchies bound to nominal and verbal meanings. Russian verbs have aspect category (each verb can be either perfective, or imperfective), and English verbs do not, so that Russian verbal meanings always correspond to subclasses of concepts that are denoted by English verb. E.g., English verb *to draw* (a picture) does not have any direct equivalent in Russian, but can be translated either with *рисовать* (to be in the process of drawing), or with *нарисовать* (to have drawn). In AIRE ontology this fact is treated so that English verbal meaning should correspond to superclass ('to draw'), and Russian verbal meanings correspond to its subclasses. Perfective meanings are linked with imperfective meanings with 'to finalize action' relation, and possibility (or impossibility) to express perfection of action or state without any additional components of meaning (in Russian) is one of the criteria to choose whether a verb means action or state. E.g., Russian verb *спать* (to be in the process of sleeping) does have perfective derivates (*поспать* (to have been in the process of sleeping for some time), *проспать* (to have been in the process of oversleeping), *доспать* (to have brought the process of sleeping to some point), etc.), but none of these verbs expresses precisely the meaning 'to have slept', which means that 'to sleep' is not an action, but rather a state. It is worth saying that even in English there is a strong difference between *to have slept* and *to have drawn*: e.g., it is much better to say *I have slept for hours* (121000 literal results in Google, which is synonymic to *I was sleeping for hours* — 76000 results) than *I have drawn for hours* (120 literal results in Google; *I have drawn* is not synonymic to *I was drawing*; *I was drawing for hours* has 14000 literal results in Google). The main difference between actions and states is that actions are processes that obligatorily lead to some changes, whereas states do not necessarily have any results or consequences. Each action has a backbone relation with a state of performing this action; in Russian this distinction is expressed lexically: cf. *идти* (to be in the process of going somewhere) and *ходить* (to be in the state of going, usually 'hither and thither' or multiple times to the same place).

All four above-mentioned hierarchies are classified according to characteristics of compatibility. Thus, transitive verb meanings and meanings of their nominal derivates correspond to subclasses of concepts 'to perform a directional action / to be in a directional state' and 'directional action / state', whereas intransitive verbs and their nominal derivates denote subclasses of nondirectional actions and states. Directional states are aliased as attitudes. Directional actions and states have objects which do not coincide with subjects, whereas nondirectional actions and states are limited to their subjects and therefore are expressed in natural languages as intransitive, or even reflexive verbs. Directional actions and states are classified according to the classes of their objects: e.g., the verb to *approve* means a state (an attitude), which is directed at a thought or idea (or, metonymically, an action of expressing this attitude), whereas the verb to *move* means an action, which is directed to a localizable (usually physical) object, but it can also mean a nondirectional action (e.g., *We moved to another apartment*). Subclasses of attitudes and directional actions are linked to the classes of their objects, and these links follow the rules of attribute overriding, so that lexical disambiguation can be performed. E.g., the verb *to take* can mean an action, which is directed at a physical object (*she took my hand*), and also (among other meanings) an action, which is directed at time span (*she took two hours to find me*).

The same classifications are done according to subject, addressee, and instrument classes of actions and states, regardless of whether they are directional or not. These classifications are vital for lexical disambiguation when it comes to verbs and processual nouns, but, unfortunately, they just rarely help to disambiguate the surrounding context. The reason is that verbs and processual nouns are very often polysemic and produce a significant amount of metonymy, so that if subject, object, instrument or addressee is denoted by polysemic nominal phrases, their meanings combine well with side-meanings of the verb or processual noun. E.g., the above-mentioned verb *to take* is highly polysemic, so the expression *I took the medicine* can be interpreted both as 'I swallowed the pill' and, e.g., as 'I approved the proposed way of healthcare'. That is why it is a very important task (which is still not fulfilled) to deduce as much verbal meanings as possible from other (basic) meanings as metonymic ones, so that they are enabled only in case the basic meanings are exhausted. E.g., the meaning 'to approve' of the verb *to take* seems to be metonymic (maybe, partially, metaphoric), whereas the meaning 'to hold' seems to be one of the basic ones.

Sometimes, however, lexical disambiguation works efficiently enough to reduce or even to get rid of ambiguity, especially when verbal meanings are restricted to animate subjects or other participants (more precisely, subclasses of 'someone' concept, because in terms of semantic restrictions, e.g., plants are much less animate than, e.g. computers or robots ore any other human-like objects). 'Someone' is rather a wide class, as its subclasses are not only human-beings, other animals, and human-like subjects, but also societies and organizations that can act as a single person. Organizations have many attributes, some of them coming from metonymy with their members (persons), and some of them being immanent for organizations. Both persons and organizations can possess objects as private property, which is marked with the relation 'to possess object'. This relation overrides more abstract relation 'to have object', so (for persons) other types of possession (also overriding the abstract relation) have to be introduced: e.g., 'to have a relative', 'to have a part of body', etc. Unlike persons, organizations can belong as property to other organizations and persons (because organizations are treated as ownership resources, and persons are not), and be parts of other organizations (as they are aggregates). Organizations can be registered in political units of an inhabited locality, and there are different ways of their creation, marked with relations that override basic variable object creation: organizations can be established, they can be formed or recruited, etc. Persons, as they are organisms with sexual reproduction, have three other possibilities of creation: they can be born by female persons, created by God or created by bioengineering. Both persons and organizations can perform different type of activities that correspond (again, with a backbone relation) to subject domains.

Subject domains are not to be confused neither with topics, nor with activities. Formally, in set theory, subject domain is domain of definition of a predicate, i.e., a union of sets in which a predicate is specified. In AIIRE ontology, subject domains are treated as aggregates of all real and possible situations and their participants, that belong to a specific class of someone's activity. There is a specific relation called 'to belong to subject domain' that links concepts with their subject domains. All classes of actions that correspond to activities that form subject domains can be performed only by subclasses of 'someone'. E.g., there is subject domain named 'science', which is formed by activity named 'cognition', which corresponds to action named 'to cognize'. This action can be performed only by a subclass of 'someone'. Subject domain hierarchy reflects the hierarchies of actions and activities, and thus is built according to very strict criteria, unlike hierarchies of topics that are usually created for the purposes of text classification.

Classification on the basis of subject domains and conceptual relations

Verbal and prepositional meanings are classified not only according to their semantic compatibility (classes of possible subjects and objects), but also according to their subject domains. E.g., there is a class of verbal meanings named 'to perform or be in state of physical motion', that is a superclass for all verbal meanings that somehow refer to physical motion. This class is further subdivided into subclasses (the general compatibility-based classification is reproduced), but it has its immanent attributes, as any motion has a direction and source point. Relations 'to be directed at physical place' and 'to start from physical place' provide compatibility of verbal meanings with corresponding meanings of prepositions. Some other spatial prepositional meanings are also linked to verbs of motion with some special relations (e.g., meanings of prepositions *through* and *past*). As for the locative prepositional meanings that do not refer to directions, these meanings are linked to all verbal meanings on the upper level, because every situation can be located in physical space.

Typical representatives, relation narrowing, and semantic analysis

It is, unfortunately, impossible to outline all the relations used and defined in AIIRE ontology in this paper, but it seems to be important to mention some large classes of relations or frequently used relations that were not mentioned before. Conceptual classes can have typical representatives (relation is called 'to have typical representative'), which are subclasses that act as metonyms. E.g., conceptual class 'place' has subclass 'physical place', which is a typical representative of place, therefore, the word *place* can denote 'physical place', although its meaning is much wider (places can be in texts, images, melodies, etc.) The same is true for the word *animal*, which tends to denote 'non-human animal'. Typical representatives often have specific compatibility because of their own relations (e.g., as it was previously mentioned, physical places have backbone relations with physical objects and, therefore, can be warm, green, etc.)

Relations, as other concepts, are created and refined in AIIRE ontology rather often, so it is quite difficult and, maybe, unnecessary to describe them all. At the moment, 312 different concepts are used as relations in the ontology, some of them being very specific, e.g., there is relation named 'to be given to employee' which binds any resource of professional occupation (including salary, other employees, vacations, etc.) with employees that receive these resources from their employers. Relations like this one are never directly denoted by constituent classes, but are rather often involved in semantic graphs, because of the underlying algorithm of semantic analysis.

This algorithm simply repeats the rules of relation inheritance and overriding, so that, e.g., genitive constructions (or similar constructions with preposition *of* in languages without morphological genitive) are initially provided with a pattern of semantic graph, which consists of two positions to fulfill (these come from syntactic head and specifier meanings) and an abstract relation 'to belong to object' between them. If any concept that is denoted by syntactic constituents is not able to participate in this relation according to relation-overriding rules, then the whole construction is treated as semantically uninterpretable and is culled. If the subject position is fulfilled, then the relation can be substituted with any of its subclasses that override this relation within the subject concept. The same is true for the object. Furthermore, relation substitution can cause both subject and object concepts substitution, if its backward links to subject or object are overridden. E.g., during evaluation of the expression Peter's vacations the upper-mentioned relation 'to belong to object' is substituted with 'to belong as resource to a person' (because vacations are a resource), then with 'to be given to employee', because vacations are resource of professional occupation, and then *Peter*, who was treated as an instance of 'person', is substituted with an isomorphic instance of 'employee', because 'to be given to employee' has 'person' as subject overridden by 'employee'. This is the basic scheme implemented in AIIRE, the real algorithm is much more complex, as it has to consider typical representatives, backbone relations, polysemy of each syntactic constituent and many other upper-mentioned peculiarities of the ontology.

It is worth emphasizing, that there are many problems concerning the ontology now. Semantic restrictions are very strict, so very often it is the case that they lead to impossibility of binding and need to be weakened somehow. Sometimes, but much more rarely, they are too weak, and ambiguity remains unresolved. However, ontology provides an interface to manipulate semantic restrictions directly, and, as the whole system is free and open-source, and, more than that, available for download from CentOS and NauLinux repositories, the project develops rather rapidly. It seems also worth mentioning, that the author of this paper has defended a PhD thesis in computational linguistics, devoted to automatic classification of news messages using syntactic semantics analysis performed by AIIRE

Bibliography

[Fellbaum, 1998] Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

[Hajic et al., 2001] J. Hajic, P. Krbec, P. Kveton, K. Oliva, V. Petkevic (2001). Serial Combination of Rules and Statistics: A Case Study in Czech Tagging. In Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL 2001), — Toulouse, France, 2001

[Knowledege base, 2014] Knowledge base. (2014, June 12). In Wikipedia, The Free Encyclopedia. Retrieved 21:23, July 3, 2014, from

http://en.wikipedia.org/w/index.php?title=Knowledge_base&oldid=612589599

[Leontyeva, 2006] Н.Н. Леонтьева (2006). Автоматическое понимание текстов. Системы, модели, ресурсы. Москва: Academia, 2006 — 303 с.

[Miller, 1995] George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

[Petkevič, 2006] Petkevič V. (2006): Reliable Morphological Disambiguation of Czech: Rule-Based Approach is Necessary. In Insight into the Slovak and Czech Corpus Linguistics (Šimková M. ed.). Veda (Publishing House of the Slovak Academy of Sciences & Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences), Bratislava, pp. 26–44, ISBN 80-224-0880-8.

[Sheffczyk et al. 2006] Scheffczyk J., Pease A., Ellsworth M. (2006). Linking FrameNet to the Suggested Upper Merged Ontology. Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS 2006), Baltimore, Maryland, pp. 289-300 — November, 2006

[Sowa, 1999] J.F. Sowa (1999). Conceptual Graphs: Draft Proposed American National Standard. In International Conference on Conceptual Structures ICCS-99, Lecture Notes in Artificial Intelligence 1640 — Berlin, New York: Springer Verlag, 1999 — pp. 1-65

[Tseitin, 1985] Г.С. Цейтин (1985) Программирование на ассоциативных сетях // ЭВМ в проектировании и производстве. -Л.: Машиностроение, 1985. Вып. 2.

[Zouaq et al. 2009] A. Zouaq, M. Gagnon, B. Ozell (2009). A SUMO-based Semantic Analysis for Knowledge Extraction. In Proceedings of the 4th Language & Technology Conference — Poznań, 2009

Authors' Information

Alexey Dobrov – Saint-Petersburg State University, Assistant Professor; Saint-Petersburg National University of Informational Technologies, Mechanics and Optics, tutor; Geline Company, head of the RnD dept.; Ramax International, software developer; P.O. Box: 199226, Novosmolenskaya emb., 4, apt. 71, St. Petersburg, Russia; e-mail: adobrov@aiire.org

Major Fields of Scientific Research: Artificial Intelligence, Natural Language Understanding, Syntax, Semantics, Ontologies