

СЛОВАРНЫЙ ПОДХОД К РАЗРЕШЕНИЮ ОМОНИМИИ ПРИ ВЫДЕЛЕНИИ ИМЕНОВАННЫХ СУЩНОСТЕЙ В РУССКОМ ЯЗЫКЕ

DICTIONARY-BASED AMBIGUITY RESOLUTION IN RUSSIAN NAMED ENTITIES RECOGNITION. A CASE STUDY¹

Brykina M. M. (m.brykina@gmail.com), ZAO Eventos, Moscow, Russia; Moscow State Lomonosov University, Moscow, Russia

Faynveyts A. V. (fainalex@lyandex.com), Freie Universität Berlin, Berlin, Germany; ZAO Eventos, Moscow, Russia

Toldova S. Ju. (toldova@yandex.ru), Moscow State Lomonosov University, Moscow, Russia; The Center for Semantic Technologies, Higher School of Economics, Moscow, Russia

The Information Extraction task and the task of Named Entities recognition (NER) in unstructured texts in particular, are essential for modern Mass Media systems. The paper presents a case study of NER system for Russian. The system was built and tested on the Russian news texts. The method of ambiguity resolution under discussion is based on dictionaries and heuristic rules. The dictionary-oriented approach is motivated by the set of strict initial requirements. First, the target set of Named Entities should be extracted with very high precision; second, the system should be easily adapted to a new domain by non-specialists; and third, these updates should result in the same high precision. We focus on the architecture of the dictionaries and on the properties that the dictionaries should have for each class of Named Entities in order to resolve ambiguous situations. The five classes under consideration are Person, Location, Organization, Product and Named Event. The properties and structure of synonyms and context words, expressions and entities necessary for disambiguation are discussed. Key words: Named Entities Recognition, Named Entities ambiguity, Named Entities disambiguation, rule-based approach.

Key words: *named entities recognition, named entities disambiguation, dictionary-based approach*

1. Introduction

The task of Named Entities recognition (NER) in unstructured texts is essential for modern Mass Media systems. Research in this area has been conducted for more than 20 years (cf. the report on the state of the art in [1], [2], [3], [8]). However the majority of works deal with English or other well-studied European languages. Some systems for Russian are discussed in [4], [5], [6].

There are two basic approaches to the NER task: handmade rule-based systems and machine learning-based systems. In this paper we discuss the architecture of a rule-based system within the task of extraction of a predefined set of Named Entities (NEs). Advantages of the suggested approach include a predictable system behavior, high precision for the user-specified list of NEs, the possibility for the user to update the system as well as to control the effects of NEs database extension. We will mainly focus on the structure of the dictionary where the NEs synonyms are stored.

Since all the entities and their synonyms within the task defined above are enumerated in the dictionary, the focus of the development shifts to the task of ambiguity resolution. There are three commonly known types of ambiguity:

- the ambiguity between a NE and a common noun (cf. *Rubin*, a football club vs. *rubin*, a jewel);
- the ambiguity between NE classes (cf. *Vladimir* as a town vs. *Vladimir* as a personal name);

¹ We would like to thank our colleagues Anastasia Bonch-Osmolovskaya, Andrei Idiatullin, Evgeny Fedko, Julia Zinova, Alexandre Arkhipov, our programmer Petr Zhalybin, and RIA Novosti project manager Philip Dudchuk for their valuable advice, support and collaboration.

- the ontology ambiguity between two entities with the same name (*Sergei Ivanov*, a politician vs. *Sergei Ivanov*, a scientist).

Some cases of NE overlapping are not discussed here, for in our case it is the user who should decide how to treat them. E.g. in (1) the user may or may not be interested in extracting Moscow as a Location while it is a part of an Organization.

(1) *Tverskoj sud Moskvy*
Tverskoj Court of Moscow

In the following sections we discuss the ways we deal with above-mentioned ambiguity cases for several classes of NEs. Afterwards we summarize the means used in the dictionaries to resolve ambiguity of a dictionary entry, and finally we present the results of our system's evaluation.

2. PLO and Events Disambiguation

2.1. System overview and task specification

Our objective was to develop a system that extracts NEs of a predefined ontology from news texts. Initial lists of NEs to be recognized were provided by the user (their size is given below); the important requirement was the following: it should be easy to add new NE and new types of NE to these lists by the user.

- Persons ≈ 5000 ;
- Locations ≈ 10000 ;
- Organizations ≈ 4000 ;
- Products ≈ 1000 ;
- Events ≈ 1000 .

The program we developed is based on OntosMiner technology [7] and partially on GATE software (<http://gate.ac.uk/>). The main function of the OntosMiner processor is extracting semantic information from unstructured texts and presenting in triples of a specific format Turtle (format for expressing data in the RDF data model). Thus it is possible to map the system output onto a user-defined Domain Model stored in ontology (see Fig. 1.).

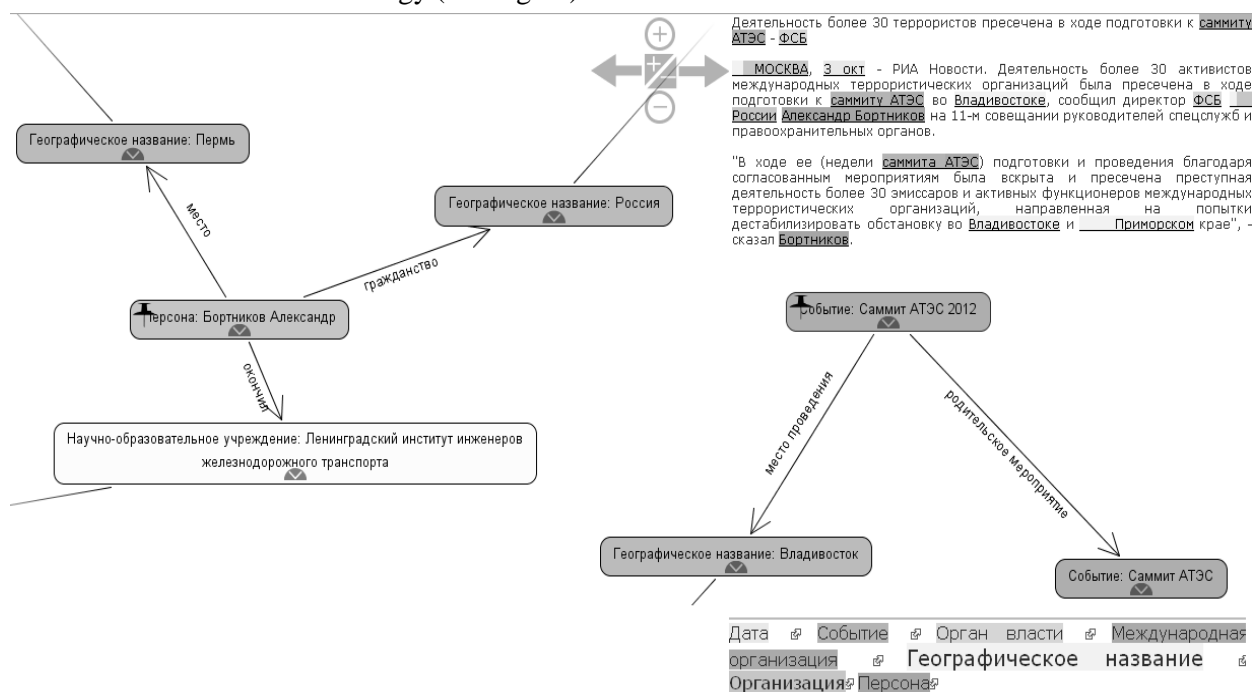


Fig. 1. Visualization of OntosMiner output.

Below we discuss the dictionary-oriented NER subsystem of OntosMiner which includes the NE ontology and linked synonyms dictionaries.

One way to present the synonyms for certain NE is to have all the synonymous full noun phrases as entries of a synonym dictionary (further on we would use the term “lookup entry,” or just a “lookup”). Our approach is quite different: we use a minimum common text material for a set of synonyms, and then expand lookup boundaries with certain rules if necessary. For instance, we need the only one dictionary entry Microsoft for three names Microsoft, Microsoft Corp., Microsoft Company etc. This approach requires less human effort to fill dictionaries and also makes it easier to treat conjunction.

In our system we distinguish between unambiguous and ambiguous lookup entries. Unambiguous lookups (=ULs) immediately identify an object, while ambiguous ones (=ALs) need to be verified through the context. Verification is performed by (manually) assigning attributes to lookups that specify the additional information needed to verify an AL (turning it into an UL). These attributes are processed by heuristic-based rules.

Unlike verification, rejection of lookups can be applied both to ULs and ALs. Lookups are rejected when their context changes the class of an object. E.g., location lookups can be rejected when adjoined to currency names (dollar SSHA - US dollar) or other entities class keywords (mul'tfil'm Madagaskar – “Madagascar” film). Company names may be ignored when used in product names (utjug Philips - a Philips iron).

2.2. Location lookups

Geographical names, their variants and abbreviations are the most common synonyms for Locations (Moskovskaja oblast', Podmoskov'e, MO – Moscow Region). According to our principle of minimal lookups we exclude locative keywords from the synonym: Moskovskaja would be the lookup entry for Moskovskaja oblast' (Moscow Region) entity. Instead we use an attribute which signals that a locative keyword is required. The algorithm requires an appropriate keyword to adjoin the lookup or to be the head of a conjunction phrase (v Moskovskoj i Ivanovskoj oblastjah – in Moscow and Ivanovo Regions). Locative keywords are stored in controlled vocabularies (CV) and can have their own synonyms.

2.2.1. Verification attributes

i. Location vs. Location ambiguity

To disambiguate homonymous Locations, such as the eight Soviet Regions (Sovetskij rajon) in Russia, ontological information about the hierarchy of administrative division is required. A specific algorithm searching for parent, ancestor or sibling locations has been developed to resolve this type of ambiguity.

Another case is the ambiguity between different types of Locations with the same name: gorod Tunis (Tunis) vs. strana Tunis (Tunisia). Here an appropriate keyword must be found to verify a lookup.

ii. Location vs. Person (family) names

Some city names are ambiguous with family names (Mogilev – Mogilev City) or even first names (Vladimir – the city of Vladimir). The solution is to have a heuristic module which extracts Person entities including those not present in a given ontology. Afterwards we use a minimization rule: when a Location is embedded into a Person, this Location is removed. It is indeed important to have at least a basic PLO-extracting module (which extracts all, including non-dictionary, mentions of PLO in the text), because one can never put all possible ambiguity cases in the dictionary.

iii. Location vs. common words

To distinguish Locations vs. common words ambiguity (g. Nahodka – lit. (the city of) Find) we use one of the verification attributes mentioned above (either keyword or region verification). Another

supporting context for such lookups would be their position in a prepositional phrase (v Nahodke – in Find).

Some ALs, including ambiguous abbreviations, may have no specific clues and can only be verified if a corresponding entity has already been found in the text.

2.2.2. Location attributes: overview

<i>type of attribute</i>	<i>scope</i>
requires a locative keyword	adjoined
requires region verification	Document
requires PP-verification	left-adjoined
requires entity to be already present in the text	Document

2.3. Organization lookups

Several types of organizations have been extracted: commercial organizations, government organizations, academic institutions, international organizations, and political parties. The names of organizations, abbreviations, or the shortest names without surrounding quotation marks (both in Cyrillic and Latin alphabet) were used as lookup entries.

2.3.1. Verification attributes

i. Organization vs. other entity ambiguity

First of all punctuation can be very helpful in this case, i.e. a lookup can become unambiguous if it occurs in quotation marks. However, quotation marks can be omitted, that's why punctuation cannot be the only means to rely on. Left-adjoined word or phrase can be useful, either the name of the organization type (e.g. factory, company, foundation) or other keyword(s) (director, CEO, chief accountant). These keywords are organized as in which can be expanded by the user.

To parse sentences (2) correctly one needs to mark a lookup entry for the organization whose name is ambiguous like in (2). If one of the distinguishing criteria occurs, the correspondence with a proper entity is not called in question any more.

(2) a.

<i>Lookup entry</i>	<i>attribute</i>	<i>Entity</i>
<i>Mir dereva</i> Mir dereva (lit.: the world of wood)	<i>needs quotation marks;</i> <i>needs left-adjoined</i> <i>word (sequence) from</i> <i>the list</i>	<i>Derevoobratyvjushchaja kompanija</i> <i>«Mir dereva»</i> Woodworking company “Mir dereva”

b. «Mir dereva» javljaetsja krupnym proizvoditelem izdelij iz dereva.

“Mir dereva” is a big manufacture for wooden products.

c. Kompanija Mir dereva - odin iz organizatorov ètoj vystavki.

“Mir dereva” company is one of the organizers of this exhibition.

d. Tema segodnjashnego zanjatija v detskom sadu «Mir dereva i metalla»

Theme of the today's lesson in the kindergarten is “The world of wood and metal”

ii. Organization vs. Organization ambiguity

Different companies that have nothing in common can bear the same name. Quotation marks and common words or phrases like “organization” or “CEO” in left context are helpless here. For the cases such as (3)-(3) besides the left context we should take into account the industry to which a

corresponding NE refer. We need to create two lookup entries for two different Rubin organizations and specify context words needed for each of them.

(3) a.

<i>Lookup entry</i>	<i>Attribute</i>	<i>Entity</i>
<i>Rubin</i> Rubin (lit.: ruby)	needs left-adjoined word from the list for the specific industry	<i>futbol'nyj klub «Rubin»</i> FC Rubin
<i>Rubin</i> Rubin (lit.: ruby)	needs left-adjoined word from the list for the specific industry	<i>Konstruktorskoe bjuro</i> «Rubin» design office Rubin

b. FK “Rubin” *vozglavljaet turnirnuju tablitsu.*
FC Rubin leads the table.

c. KB “Rubin” *zanimaetsja proektirovaniem.*
Design office Rubin does projecting.

Our CV gets a sort of hierarchical structure: we supplement each entry of a CV either with the name of industry or with a special attribute value which means that it is impossible to determine the industry.

The other case of organization ontological ambiguity is the organizations with the identical names situated in different regions (4a,b). To be able to link such ALs with proper entities we need to add an attribute pointing to the corresponding region. The attribute value is implemented as a link to the object dictionary entry for the location in question.

Once such a lookup is found, the nearest mention of location in this document is being looked for. If no appropriate location is mentioned in the document, the default location — Russian Federation — is set up, since we deal with Russian news agencies’ texts. Thus both in (4a) and in (4c) the Ministry of Foreign Affairs of Russian Federation should be recognized, although there is no explicit mention of Russian Federation in (4c).

(4) a. *Ministerstvo inostrannyh del RF otvetilo na pros’bu Gennadija Onishchenko*

Ministry of Foreign Affairs of Russian Federation answered to the request of Gennadij Onishchenko.

b. *Vladimir Putin posetil Parizh i provel vstrechu s glavoj MIDa.*

Vladimir Putin visited Paris and hold a meeting with the head of the Ministry of Foreign Affairs.

c. *Ministerstvo inostrannyh del otvetilo na pros’bu Gennadija Onishchenko*

Ministry of Foreign Affairs answered to the request of Gennadij Onishchenko.

2.3.2. *Types of controlled vocabulary entries*

Controlled vocabularies for keywords use two obligatory attributes for each CV entry: type and industry. Entries of different type are processed differently:

- A *prefix* (“organization”, “joint-stock company”, “factory”) designates the following AL as an organization. A *general prefix* (“company”, “society”) doesn’t imply a specific industry, while a *specific prefix* (“bank”) does.
- A *keyword* (“director”, “CEO”) means that the AL is *probably* an organization.
- A *key adjective* (“cosmic”, “aircraft”) forms a specific prefix when combined with a general prefix.
- A *postfix* (“Limited”, “& Co”) designates the previous AL as an organization.

2.3.3. *Organization attributes: overview*

<i>type of attribute</i>	<i>scope</i>
--------------------------	--------------

requires a specific location	Document
requires a general left-adjoined keyword	left-adjoined
requires a specific left-adjoined keyword	left-adjoined
requires quotation marks	surrounding

2.4. Person lookups

In Russian news texts a person is usually introduced by calling him/her with first name, family name, and optionally patronymic. Further mentions use a family name only. We normally use family names as lookup entries. An obligatory check for these lookups is name verification: somewhere in the text the family name must be used adjoined to appropriate first name(s) or initials. Exceptions are allowed for a short list of very famous people (e.g. presidents) who can be referred to by family names alone. Other lookups for Persons which do not require name-verification can be:

- nickname or stage name: *Boris Akunin* (*Boris Akunin*), *Vitas* (*Vitas*);
- first name(s) plus family name if there are several variants of name components combinations (for example, Arab person names);
- person name plus person's status-role: *Koroleva Elizaveta* (*Queen Elisabeth II*);
- person name written in English or her/his native language, which can be useful if Russian transcription is not consistent.

2.4.1. Verification attributes

i. Person vs. Location ambiguity

Proper names designating first names, and more often family names can also stand for Locations: *Anton Chekhov* vs. *Chekhov* (city); *Lion Izmajlov* vs. *Lion* (city). As mentioned, this ambiguity is resolved by means of minimization: when a possible Location is embedded into Person, this Location is removed.

ii. Person vs. Person ambiguity

Ambiguous Person names are not often found in top-news texts in Russian, though the more widespread name a Person has, the more often we expect to find his/her namesake. Wikipedia lists 25 persons for *Sergej Ivanov*, most of which can be expected in news texts (though only one of them is being mentioned constantly through the past years, and one more was several years ago). To disambiguate homonymous person names the following ontology information is used:

- (a) profession, academic rank, title (some constant status of a Person);
- (b) place of employment (this can be an Organization or a Location);
- (c) position at the place of employment.

We use CVs for (a) and (c), so that synonyms of one concept form a group, and each of the words could be used as verification marker. The search scope is the paragraph containing the AL.

iii. Person vs. other entity ambiguity

We don't use specific attributes to distinguish between a Person and another entity, although family names can be ambiguous with common words. It seems sufficient to have one verified mention of a person in the text to consider all the other occurrences verified (except for when there is more than one person with the same family name).

2.4.2. Person attributes: overview

<i>type of attribute</i>	<i>scope</i>
requires name verification	adjoined

requires title/employment/position verification	Paragraph
---	-----------

2.5. Event lookups

Some cultural, governmental and other actions are described as Events (exhibitions, film festivals, messages of the President to the Federal Assembly, United Nations summit). Shortest versions of names, types of events without quotes are used as lookup entries.

2.5.1. Verification attributes

i. Event vs. another entity ambiguity

Event names are not always unique and could mean something totally different in other contexts. Quotation marks are not regularly used with Event names (even more rarely than with organizations), that's why the only reliable clue is a left-adjoined word or phrase of a proper class.

(5) a.

Lookup entry	Attributes	Entity
Zolotoj lev	needs left-adjoined word (sequence) from the list for the theater festival	Theater festival "Zolotoj Lev" (lit.: Golden lion)

b. *Ivan Ivanov – glavnyj redaktor zhurnala «Zolotoj Lev».*

Ivan Ivanov is the chief editor of the "Zolotoj Lev" (lit.: Golden lion) magazine.

c. *Teatral'nyj festival' «Zolotoj lev» proshel v Permi.*

Theater festival "Zolotoj Lev" took place in Perm.

d. *Priz kinofestivalja v Venetsii – statuètka zolotogo l'va.*

The prize of the film festival in Venice is a golden lion statuette.

Among the entities named Zolotoj Lev in (5), only in (5). the Event should be recognized. To ensure this, the lookup entry Zolotoj Lev needs an attribute saying that an appropriate keyword is required (5). Possible keywords are also organized in a CV.

ii. Parent Event vs. Child Event

Many Events like festivals are periodical. For instance, in (6) we would like to see three different events. In (6), a class of periodical song contests called Eurovision (parent Event); in (6) and (6), specific child Events should be recognized.

(6) a. *Evrovidenie – vazhnoe sobytie v mire pop-muzyki.*

Eurovision is an important event in pop music world.

b. *Poslednee Evrovidenie proshlo v Moskve.*

The last Eurovision took place in Moscow.

c. *Na Evrovidenii v 2009 godu pobedil Dima Bilan.*

Dima Bilan won Eurovision contest in 2009.

A periodical Event like in (7a) must be associated with a unique place, number or year to be recognized as a specific child Event:

(7) a. *Evrovidenie v Baku OR Evrovidenie-2012*

Eurovision in Baku OR Eurovision-2012

b.

Lookup entry	Attributes	Entity
Evrovidenie Eurovision	needs a time-marker: year (2012) OR needs a place-marker (Baku)	Evrovidenie-2012 Eurovision-2012

For the cases like in (7), the system should be aware that: (i) both events took place once in 2012; (ii) they are the same; (iii) they did not take place in 2011 or in 2010. Thus, lookup entries must have an attribute requiring a mention of a year, a number or a place. In (7) it is essential to point out that we need a time-marker OR a place-marker: both at once are also helpful but not necessary.

For some other types of Events, attributes should rather be combined with AND. Consider (8):

- (8)
- a. *avtosalon v Toronto*
automobile show in Toronto
 - b. *avtosalon, otkryvshijsja vchera v Toronto*
automobile show (that) opened yesterday in Toronto
 - c.

Lookup entry	Attribute	(Hypothetical) proper entity
<i>Avtosalon</i> automobile show	needs a place-marker (Toronto)	<i>Avtosalon v Toronto</i> Automobile show in Toronto

The events in (8) must both correspond to the same Event. It may be impossible to estimate all the language constructions that describe the connection between this type of events – automobile show – and its location – Toronto. Instead we use an attribute signaling that a place-marker is needed (8c).

If such an event is periodical, a time-marker (number or year) is also to be found. Thus, to determine a correspondence with a proper entity both types of markers (time AND place) should be present in the document (9).

(9)

Lookup entry	Attributes	(Hypothetical) proper entity
<i>avtosalon</i> automobile show	needs a time-marker: year (2011) AND needs a place-marker (Toronto)	<i>Avtosalon v Toronto 2011</i> Automobile show in Toronto – 2011
<i>avtosalon</i> automobile show	needs a time-marker: year (2012) AND needs a place-marker (Toronto)	<i>Avtosalon v Toronto 2012</i> Automobile show in Toronto – 2012

2.5.2. Event attributes: overview

<i>type of attribute</i>	<i>Scope</i>
needs a place-marker	Sentence
needs a time-marker	Sentence
needs a left-adjoined keyword	left-adjoined
needs both place and time markers	Sentence
needs either place or time marker	Sentence

3. Unified Template for an Arbitrary Object Type (NE)

All the attribute systems described in previous sections were designed for specific object types. But the list of object types was not fixed and could be modified by the user. We have then created a unified template of preset attributes through which ALs can be marked and then processed by disambiguating rules.

3.1. Components of a unified attributes template:

- i. User can operate both substrings and ontology entities. E.g. if a mention of Russian Federation in the context is needed to disambiguate an AL, the user can just give a link to the corresponding entity. For example, if we want to extract Russian Football National Team, we can create an AL *national team*, requiring *Russian Federation* or its synonyms to verify this entry. An entity from any dictionary (also CV) can be chosen. An ontological feature that contains a link to an object and requires verification can also be used.
 - ii. Key-words/entities or stop-words/entities can be set. This means that user can define not only obligatory context but also a context that blocks the association of an AL to an entity.
 - iii. Users can choose the scope of each key- or stop-parameter: immediate left or right context, sentence, paragraph, or document.
 - iv. Key- and stop-parameters can be combined with each other with OR or AND operators.
 - v. Users can set whether quotation marks can be used or are required to identify an object.
 - vi. Users can add ALs that are verified only if a corresponding entity has already been found in the text
- This unified template was used to make a lookup dictionary for Product entities. The results can be seen in Table 1.

4. Evaluation and Discussion

The general assessment of the system is based on a procedure similar to one described in the MUC2 conferences. A random set of 300 texts from RIA Novosti news agency was chosen as test corpus. It was manually annotated in GATE3 by two persons⁴, the cases of annotators' disagreement were further revised. In the evaluation procedure only entities from the Ontology were taken into consideration. The results are shown in Table 1.

Table 1. Evaluation results

NE	Number of cases	Recall	Precision	F-measure
Location	2270	0,98	0,99	0,98
Organization	1654	0,93	0,95	0,94
Person	453	0,94	0,99	0,96
Event	55	0,98	0,85	0,91
Product	138	0,96	0,98	0,97

Three notes are in order concerning the results above:

- we only provide counts for NPs that correspond to the NE from the Domain ontology (we've also used heuristic mechanisms to extract NEs, but these are not counted);
- the quality of the system depends crucially on the quality of manually filled dictionaries, and thus these results cannot be reproduced unless based on the same dictionaries;
- the system gives its user control over any particular object; the metric based on the quality evaluation for each ontology entity is out of discussion in present paper.

The evaluation however demonstrates that the lightweight system based on user dictionaries and rather simple rules could be quite helpful if one's goal is extracting a limited number of NEs, and that the cases of ontology ambiguity are not too frequent in general news texts to influence significantly the performance of the system based on predefined high-quality NEs dictionaries.

² http://www-nlpir.nist.gov/related_projects/muc/

³ <http://gate.ac.uk/>

⁴ We are grateful to our colleagues E. Tarasov and I. Novikov who have helped us to annotate the corpus.

5. Conclusion

We have presented a system for NEs extraction and disambiguation based upon manually created dictionaries⁵. Ambiguous lookups can be assigned multiple attributes depending on the contextual information needed for disambiguation. Verification is performed by heuristics-based rules. Thus, our system doesn't require train corpus, it has predictable behavior, and is extensible to other types of objects. The user has access to every object's properties and a possibility to use a synonymous set of keywords.

We have discussed clues that help disambiguate NEs in Russian news texts, such as:

- text properties (quotation marks);
- ontological properties:
 - object class hierarchy;
 - location associated with an object;
 - class-specific features (status-role of a Person, organization industry etc.)⁶.

Clues may have different scope, e.g. a sentence, a paragraph, or even the whole text (with special search algorithm). In some cases adposition of a clue is required.

We argue that for such systems:

- it is possible to suggest a unified template allowing to add attributes that serve to disambiguate different types of objects (other than PLO);
- along with dictionary-based extraction, it can be useful to have a guessing-module to help resolve some types of ambiguity and thus ascribe less attributes to lookup entries;
- there can be specific rules based on metatextual information, which recover implicit attributes (for example, Russia can be set as default location for Russian news agencies texts).

6. References

1. Antonov E. (2012) Models of automatic disambiguation of ontologic homonymy. [Modeli avtomaticheskogo razresheniya ontologicheskoy omonimii], PhD thesis, manuscript.
2. Bunescu R. and Paska M. (2006), Using Encyclopedic Knowledge for Named Entity Disambiguation, Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006), Trento, Italy: Association for Computational Linguistics, pp. 9–16.
3. Hoffart, J., Yosef, M. A., Bordino, I., Fürstenau, H., Pinkal, M., Spaniol M., Taneva B., Thater S., Weikum, G. (2011), Robust Disambiguation of Named Entities in Text, Proceedings of Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 782–792.
4. Kiselev S., Ermakov A., Pleshko V. (2004), A search of facts in texts on natural language on the base of net descriptions, [Poisk faktov v tekste estestvennogo jazyka na osnove setevykh opisaniy], Computational linguistics and intellectual technologies ("Dialog-2004"). Proceedings of the International Conference [Kompjuternaja lingvistika i intellektual'nye tehnologii ("Dialog-2004"). Trudy Mezhdunarodnoj Konferentsii]. Zvenigorod, pp. 282–285.
5. Krejdlin L. (2006), TagLite: Programma vydelenija russkih individualizirovannyh imennyh grupp TagLite [TagLite: The Program of Identification of Russian Individualized NPs]// Kompjuternaja lingvistika i intellektual'nye tehnologii ("Dialog-2006"). Trudy Mezhdunarodnoj Konferentsii

⁵ Automatic synonyms detection and attributes filling would be more appropriate, but has not been implemented yet.

⁶ The frequency of different ambiguity situations and the impact of attributes on the final result need further quantitative analysis and evaluation methods.

[Computational linguistics and intellectual technologies (“Dialog-2006”). Proceedings of the International Conference]. Zvenigorod, pp. 292–297.

6. Kuznetsov I. (2012), The methods of discovery of objects and their links presented implicitly in text. [Metodiki vyjavlienija objektov i zvjazej, zadannyh v nejavnom vide], available at: <http://www.dialog-21.ru/digests/dialog2012/materials/pdf/%D0%9A%D1%83%D0%B7%D0%BD%D0%B5%D1%86%D0%BE%D0%B2%D0%98%D0%9F.pdf>
7. Minor S. Starostin. A. (2007), Ontos: Technology of knowledge extraction from unstructured texts and semantic indexing, [Ontos: Tehnologija izvlechenija znaniy iz nestruktirovannyh tekstov i semanticheskoe indeksirovanie], Computational linguistics and intellectual technologies (“Dialog-2007”). Proceedings of the International Conference [Kompjuternaja lingvistika i intellektual’nye tehnologii (“Dialog-2007”). Trudy Mezhdunarodnoj Konferentsii]., Bekasovo.
8. Nadeau D. and Sekine S. A survey of named entity recognition and classification, *Linguisticae Investigationes*, Amsterdam, Netherlands: John Benjamins Publishing Company, 1: Vol. 30. pp. 3-26.