Sociopolitical Thesaurus in Concept-based Information Retrieval

Ageev M., Dobrov B., Loukachevitch N.

Research Computing Center of Moscow State University (NIVC MSU);
NCO Center for Information Research

{ageev, dobroff, louk}@mail.cir.ru

Abstract

In CLEF2005 experiments we used bilingual Russian-English Sociopolitical thesaurus that we constructed for more than 10 years specially as a tool for automatic text processing in information-retrieval tasks. The same resource and the same algorithm were used for ad-hoc and domain –specific tasks.

Categories and Subject Descriptors

H.3.1 **[Content Analysis and Indexing**]: Indexing methods, Linguistic processing, Thesauruses; J.1 **[ADMINISTRATIVE DATA PROCESSING]**: Government, Law.

Keywords

Linguistic Ontology, Information Retrieval Thesaurus, Conceptual Indexing

1. Introduction

Our group participated in two tasks: Ad-Hoc and Domain Specific Tasks. In both tasks we used the same resource bilingual Russian-English Sociopolitical thesaurus and the same algorithm. In our report we describe our resource and algorithm.

We develop the Sociopolitical thesaurus from 1994. Its domain is a broad domain of contemporary social relations – Sociopolitical domain, therefore the Thesaurus includes a lot of terminology of domains of social sphere such as politics, economy, law, defense, industry, scientific policy, education, sport, arts and others, and also thematic words and expressions of general language.

The Sociopolitical thesaurus includes more than 32 thousand concepts, 78 thousand Russian terms and 85 thousand English terms.

In construction of the thesaurus we combined three different methodologies:

- the methods of construction of **information-retrieval thesauri** (information-retrieval context, analysis of terminology, terminology-based concepts, a small set of relation types)
- the development of **wordnets** for various languages (word-based concepts, detailed sets of synonyms, description of ambiguous text expressions)
- ontology and **formal ontology** research (strictness of relations description, necessity of many-step inference).

2. Sociopolitical Domain

There are several genres of documents of considerable social significance because they concern not only specific professionals, but also life of various social groups of population. These genres of documents are: legal and normative documents, international treaties, newspaper articles, and news reports.

These different types of documents have very important similarity in their deep content. They relate (describe, discuss, regulate) to public and social relations existing in the contemporary society. The conceptual similarity leads to serious intersection of vocabulary and terminology used in these genres of texts.

The reason of this phenomenon is that all these texts can be considered as documents of the same “polythematic” domain – a domain describing life of the contemporary society. We call this domain “sociopolitical” domain (Loukachevitch, Dobrov 2004c). In large extent the sociopolitical domain comprises terminology of many specific domains as state policy, economy, law, finance, social sphere and many others (fig. 1,2).

On the other hand a lot of documents of the domain containing technical terms are addressed to non-professional people, are understandable to them. In our opinion, it means there exists an intermediate area where the general conceptual system and the upper levels of conceptual systems of specific domains intersect, and the sociopolitical domain is this intermediate area.

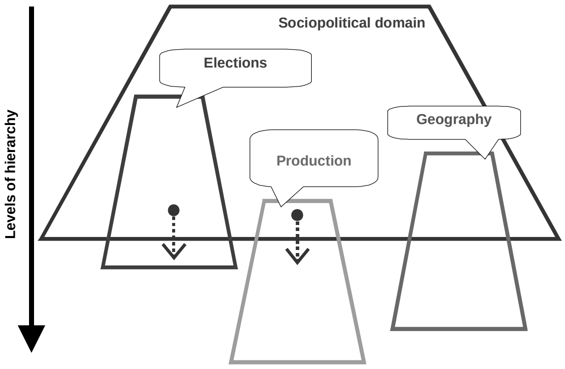


Figure 1. Special conceptual systems vs. Sociopolitical conceptual system

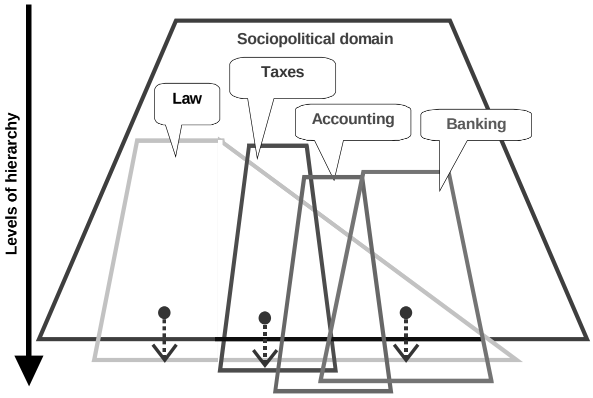


Figure 2. Interrelations of specific domains within Sociopolitical domain

So in (Rondeau, 1980) it was indicated existence of an intermediate zone between general lexicon and specific terminologies (Figure 3a), but we assume that this intermediate zone is much larger (Figure 3b).

Therefore development of linguistic resources or ontologies for the sociopolitical domain is very productive:

- they can be used for automatic text processing of important types of documents,
- they can serve as a rich source for development of resources and ontologies in specific domains.

Since 1994 we develop a concept-based resource for automatic text processing called Sociopolitical thesaurus.

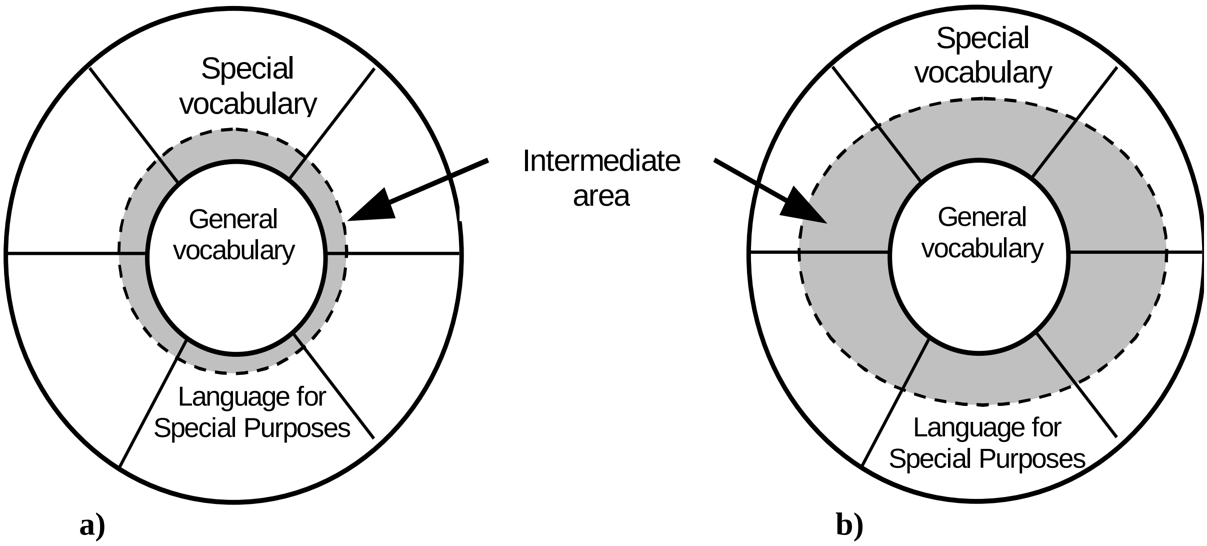


Figure 3.

3. Thesaurus

3.1. Structure of Thesaurus

The Sociopolitical thesaurus (below the Thesaurus) is a hierarchical net of concepts. We consider it as a kind of a linguistic ontology. Concepts of the Thesaurus originate from senses of language expressions, that is single words or multiword expressions.

The main unit of the Sociopolitical Thesaurus is a concept. When a new concept of the Thesaurus is introduced, it is necessary to assign its name. The name of a concept has to be clear and unambiguous for native speakers. In the Russian-English thesaurus a concept has to have a name in Russian and a name in English. These names are used in different representations of text processing results.

A concept has a set of linguistic expressions that can be used for reference to the concept in texts. A set of linguistic expressions of a concept is called 'text entries of a concept' and can be considered as a synonymic row. In Russian-English Thesaurus a concept has a set of Russian text entries and set of English text entries. These text entries are used to recognize a concept in texts.

Concepts often have more than 10 text entries including single nouns, verbs, adjectives and noun or verb groups. For example, a set of English text entries of concept *JUDICIAL COURT* look as follows: *court, court authorities, court instance, court of judiciary, court of jurisdiction, court of justice, court of law, judicature, judicial bodies, judicial court, judicial organ, judicial tribunal, law court, tribunal.* Concept *COURT SENTENCE* has 19 text entries including such as *sentence by the court*, *sentence of conviction, judgement of conviction* and others.

A concept of Thesaurus has relations with other concepts. The main types of relations are taxonomic relations and specific set of conceptual relations based on ontological dependence relations (Guarino, 2000). This set of relations was experimentally confirmed to be effective in information-retrieval applications (Loukachevitch, Dobrov 2002, 2004a).

Types of conceptual relations in the Thesaurus are:

- taxonomic relations,
- generalized part-whole relations describing internal characteristics of entities (physical parts, properties, participants for situations). In establishing of part-whole relations we use an important rule: concepts-parts have to be ontologically dependent from concepts-wholes. Therefore in the Thesaurus a tree is not a part of a forest (in fact, forest tree can describe as a part of forest). This rule provides transitivity of part-whole relations of the Thesaurus,
- external relations of ontological dependence. So in the Thesaurus concept *FOREST* is described as a dependent concept from concept *TREE*, because forests can not exist without trees, but trees can grow in many others places, not only in forests.
- related term (RT) relation is used for description of relations between very similar concepts not merged to the same concept.

Taxonomic relations and part-whole relations (with above mentioned restrictions) are considered as transitive. Taxonomic relations, part-whole relations and external relations are hierarchical relations. Therefore a concept of the Thesaurus can have a set of hierarchically lower concepts – a tree of the concept. These trees can be used for query expansion.

3.2. Development of Thesaurus

In 1994 we began development of Sociopolitical thesaurus using semi-automatic methods to find multiword terms in text collections of official documents and newspaper articles. Our procedure of terms acquisition consisted of two stages. At the first stage term-like expressions were automatically identified in the texts of the corpus. Rules defining term-like expressions included syntactical and lexical conditions. At the second stage our specialists had to look through the revealed expressions, choose terms from them and add new terms to the Thesaurus. The procedure was working during four years, processed more than 200 Mb of texts and collected more than 200 thousand term-like expressions. It was stopped because it became difficult to find new useful terms, terminology coverage became very high.

Now the Thesaurus continues to grow (approximately 2000 concepts for a year). This growth is due to several factors:

- use of the Thesaurus in applications reveals additional useful concepts,
- analysis of new but already frequent words and expressions in text collections of the sociopolitical domain (normative documents, newspapers),
- adding more specific issues (usually discussed only in professional documents) of such domains as banking, taxes, customs duties, accounting and others. “Professional” concepts are usually located in lower levels of the hierarchy of the Thesaurus.

The Thesaurus was translated into English (in fact, most concepts received sets of English text entries) and now contains more than 85 thousand English text entries (Loukachevitch, Dobrov 2004). Several applications of the Thesaurus as a bilingual resource concern processing of documents in English, for example, documents of European Court for Human Rights.

3.3. Comparison to Other Resources

The Sociopolitical thesaurus differs from conventional information-retrieval thesauri and from such linguistic resources as WordNet (Miller et.al. 1990) and EuroWordNet (Climent et.al. 1998).

In developing a conventional information retrieval thesaurus the goal is to describe terms necessary for representation of documents’ main topics (LIV, 1984). More specific terms are not included. Ambiguous terms are provided with scope notes and comments convenient for human subjects. In fact a conventional information retrieval thesaurus describes an artificial language based on the real language of a certain domain. To index documents human subjects have to use their domain, common sense, and grammatical knowledge not described in a thesaurus in order to index documents. Therefore conventional information-retrieval thesauri created for manual indexing are hard to be utilized in an automatic indexing environment (Salton 1989, Soergel et.al. 2004, Tudhope et. al. 2001). To be effective in automatic text processing a thesaurus needs to include a lot of information that is usually missed in thesauri for manual indexing such as considerably more concepts and terms, ambiguous text expressions, many levels of hierarchy as we did in Sociopolitical thesaurus.

The Sociopolitical thesaurus is based on senses of linguistic expressions (words and multiword expressions), has means for description of lexical ambiguity similar to such linguistic resources as WordNet and EuroWordNet . At the same time there are important distinctions:

- concepts in the Thesaurus are the same for all parts of speech;
- inclusion of multiword expressions to the Thesaurus’s net is regulated with strict but more liberal rules (Bentovogli, Pianta 2004). It is possible to add a new multiword expression that looks as a syntactically compositional phrase if it brings new information to the thesaurus knowledge. Our policy is to find as many such useful multiword expressions as possible,
- descriptions of concepts include much thematic information: possible situations, participants, properties and so on,
- conceptual relations in the Sociopolitical thesaurus are designed for and tested in information-retrieval tasks.

Comparing the Sociopolitical thesaurus to existing ontologies we would like to stress that it is the largest linguistic ontology in the very important and broad domain of contemporary public and social relations. Specially narrowed system of relations allows us to develop resources working in real information-retrieval applications.

4. Thesaurus-based Text Processing

Processing of all received texts in Russian and English includes several stages:

- extraction of formal parameters of documents (source, date, authors and so on),
- morphological analysis,
- terminological analysis – matching with Thesaurus terms including lexical disambiguation procedures. After this stage conceptual index to a document can be built. This index does not depend on the initial language of a document.

- thematic analysis – construction of thematic representation of texts based on conceptual relations described in the Thesaurus. The thematic representation simulates the topical structure of a text dividing all terms of the text to thematic nodes of sense-related terms (Loukachevitch, Dobrov 2000). The technique is based on such properties of texts as local cohesion (Hirst, St-Onge, 1998) and global coherence. The schematic view of thematic representation of Russian and English documents is depicted in Fig. 4. During this stage weights of concepts in the conceptual index are determined. The concepts weights in a text depend not only on frequencies of concepts but of presence of semantically-related concepts in the same text.

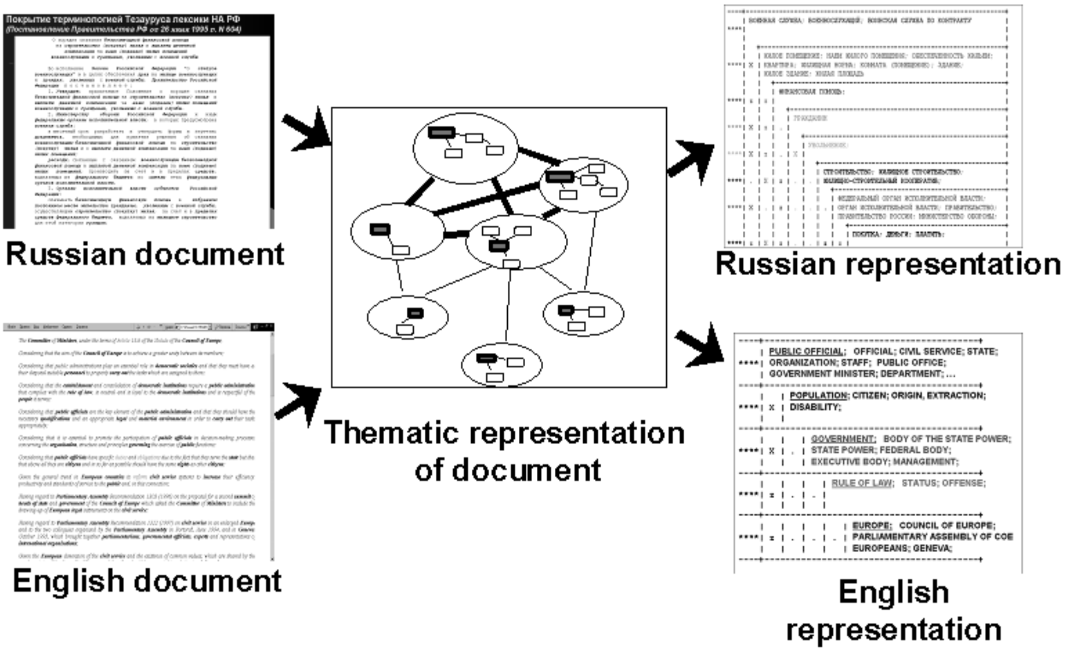
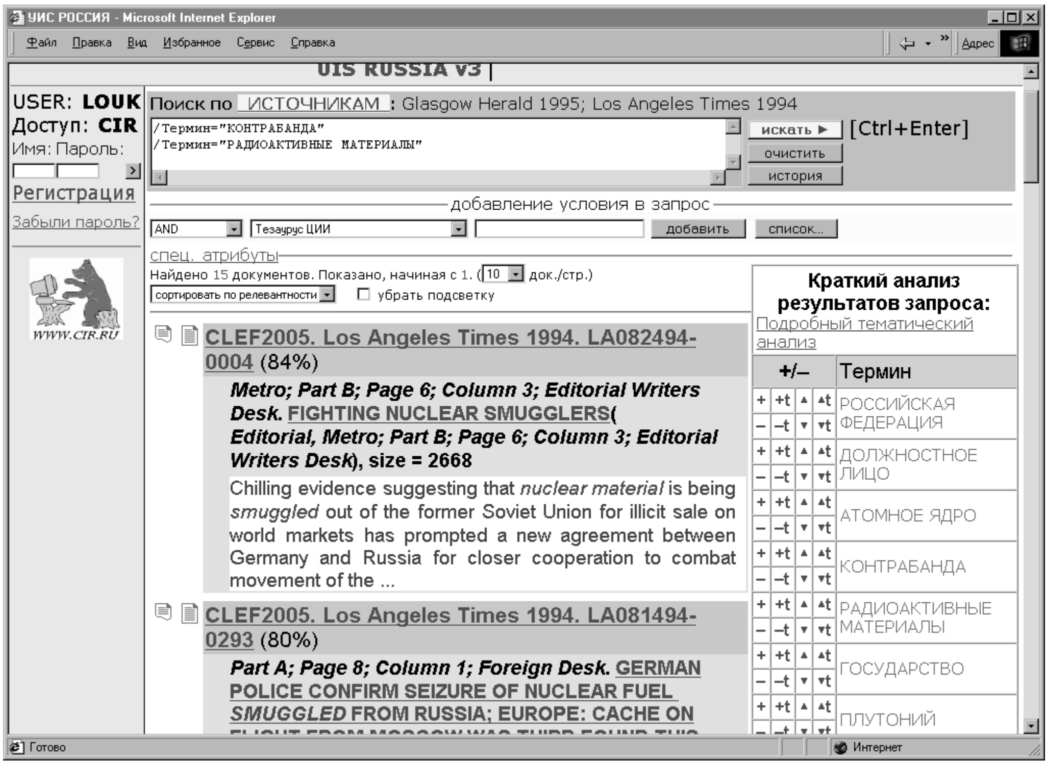

Russian document
Thematic representation of document
Russian representation
English document
English representation
PUBLIC OFFICIAL; OFFICIAL; CIVIL SERVICE; STATE; ORGANIZATION; STAFF; PUBLIC OFFICE; GOVERNMENT MINISTER; DEPARTMENT; ...
POPULATION; CITIZEN; ORIGIN, EXTRACTION; DISABILITY;
GOVERNMENT; BODY OF THE STATE POWER; STATE POWER; FEDERAL BODY; EXECUTIVE BODY; MANAGEMENT;
RULE OF LAW; STATUS; OFFENSE;
EUROPE; COUNCIL OF EUROPE; PARLIAMENTARY ASSENBLY OF COE EUROPEANS; GENEVA;

Figure 4.


УИС РОССИЯ - Microsoft Internet Explorer
Файл Правка Вид Избранное Сервис Справка
Адрес
искать ▶
очистить
история
добавить
список...
WWW.CIR.RU
Готово
Интернет

UIS RUSSIA v3

USER: **LOUK**
Доступ: **CIR**
Имя: Пароль:
Регистрация
Забыли пароль?

Поиск по ИСТОЧНИКАМ : Glasgow Herald 1995; Los Angeles Times 1994

```
/Термин="КОНТРАБАНДА"
/Термин="РАДИОАКТИВНЫЕ МАТЕРИАЛЫ"
```

[Ctrl+Enter]

добавление условия в запрос

AND | Тезаурус ЦИИ

спец. атрибуты

Найдено 15 документов. Показано, начиная с 1. (10 док./стр.)

сортировать по релевантности ☐ убрать подсветку

CLEF2005. Los Angeles Times 1994. LA082494-0004 (84%)

***Metro; Part B; Page 6; Column 3; Editorial Writers Desk.* FIGHTING NUCLEAR SMUGGLERS(*Editorial, Metro; Part B; Page 6; Column 3; Editorial Writers Desk*), size = 2668**

Chilling evidence suggesting that *nuclear material* is being *smuggled* out of the former Soviet Union for illicit sale on world markets has prompted a new agreement between Germany and Russia for closer cooperation to combat movement of the ...

CLEF2005. Los Angeles Times 1994. LA081494-0293 (80%)

***Part A; Page 8; Column 1; Foreign Desk.* GERMAN POLICE CONFIRM SEIZURE OF NUCLEAR FUEL *SMUGGLED* FROM RUSSIA; EUROPE: CACHE ON**

Краткий анализ результатов запроса:
Подробный тематический анализ

+/–				Термин
+	+t	▲	▲t	РОССИЙСКАЯ ФЕДЕРАЦИЯ
–	–t	▼	▼t	
+	+t	▲	▲t	ДОЛЖНОСТНОЕ ЛИЦО
–	–t	▼	▼t	
+	+t	▲	▲t	АТОМНОЕ ЯДРО
–	–t	▼	▼t	
+	+t	▲	▲t	КОНТРАБАНДА
–	–t	▼	▼t	
+	+t	▲	▲t	РАДИОАКТИВНЫЕ МАТЕРИАЛЫ
–	–t	▼	▼t	
+	+t	▲	▲t	ГОСУДАРСТВО
–	–t	▼	▼t	
+	+t	▲	▲t	ПЛУТОНИЙ

Figure 5.

After text processing documents and all types of extracted information (formal parameters, word and conceptual indexes) are loaded to a version of University information system RUSSIA (www.cir.ru) (fig. 5).

Thesaurus-based retrieval in our system is independent of a language used in a query and in a text, and a retrieval set can contain texts in both languages.

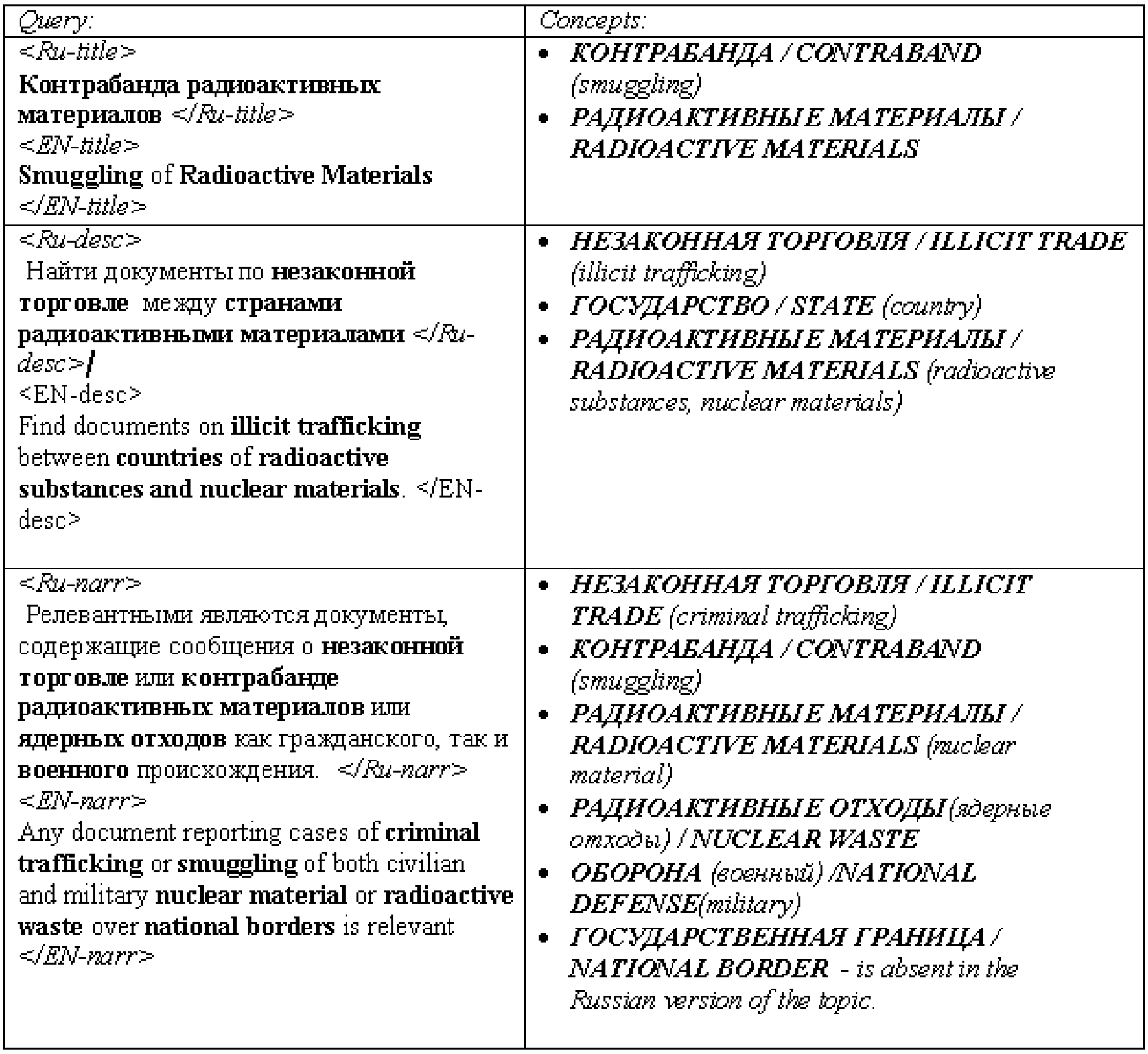
The right column of the screen shows concepts specific for the retrieval set. Top-rank terms are computed using a technique similar to blind relevance feedback.

A user can modify the query, add or delete the concepts of the right column from the query using only one key. Names of these concepts can be also formulated in both languages. Therefore a user can refine a query using his/her native language, and only after this refinement stage a user has to begin reading or translation of texts in another language.

5. Processing of CLEF Topics and Results of Experiments.

The main idea of thesaurus-based processing of CLEF topics was as follows.

We supposed that matching of topics with Thesaurus concepts has to highlight important entities and miss abstract words that can be easily substituted by other words in documents of the collection. Ambiguity of terms in the Thesaurus is much lower than for general vocabulary (Loukachevitch, Dobrov 2004c). So we decided to construct Boolean queries only from Thesaurus Concepts found in a topic.



Query:	*Concepts:*
<Ru-title> **Контрабанда радиоактивных материалов** *</Ru-title>* *<EN-title>* **Smuggling** of **Radioactive Materials** *</EN-title>*	• ***КОНТРАБАНДА / CONTRABAND*** *(smuggling)* • ***РАДИОАКТИВНЫЕ МАТЕРИАЛЫ / RADIOACTIVE MATERIALS***
<Ru-desc> Найти документы по **незаконной торговле** между **странами радиоактивными материалами** *</Ru-desc>*| <EN-desc> Find documents on **illicit trafficking** between **countries** of **radioactive substances and nuclear materials**. </EN-desc>	• ***НЕЗАКОННАЯ ТОРГОВЛЯ / ILLICIT TRADE*** *(illicit trafficking)* • ***ГОСУДАРСТВО / STATE*** *(country)* • ***РАДИОАКТИВНЫЕ МАТЕРИАЛЫ / RADIOACTIVE MATERIALS*** *(radioactive substances, nuclear materials)*
<Ru-narr> Релевантными являются документы, содержащие сообщения о **незаконной торговле** или **контрабанде радиоактивных материалов** или **ядерных отходов** как гражданского, так и **военного** происхождения. *</Ru-narr>* *<EN-narr>* Any document reporting cases of **criminal trafficking** or **smuggling** of both civilian and military **nuclear material** or **radioactive waste** over **national borders** is relevant *</EN-narr>*	• ***НЕЗАКОННАЯ ТОРГОВЛЯ / ILLICIT TRADE*** *(criminal trafficking)* • ***КОНТРАБАНДА / CONTRABAND*** *(smuggling)* • ***РАДИОАКТИВНЫЕ МАТЕРИАЛЫ / RADIOACTIVE MATERIALS*** *(nuclear material)* • ***РАДИОАКТИВНЫЕ ОТХОДЫ*** *(ядерные отходы)* ***/ NUCLEAR WASTE*** • ***ОБОРОНА*** *(военный)* ***/NATIONAL DEFENSE****(military)* • ***ГОСУДАРСТВЕННАЯ ГРАНИЦА / NATIONAL BORDER*** *- is absent in the Russian version of the topic.*

Figure 6.

All parts of topics were compared to the Thesaurus concepts. Figure 6 shows Topic C264 and Thesaurus concepts found in its zones. In parentheses text entries of a concept, different from concept names, are indicated.

Let us denote concepts found in the title of a topic as C_{t1}… C_{tn}, concepts found in the description of a topic - C_{d1}… C_{dm}, concepts found in the narrative of a topic – concepts C_{n1}… C_{nk} .

Search of documents included several steps. New documents received at every next step are added to the end of the document list received from previous steps.

Step 1. In the first step we suppose that main entities of a topic are named in the title. Concepts found in the description and the narrative inform us about additional properties of concepts from the title.

Then the main type of topic representation was as follows:

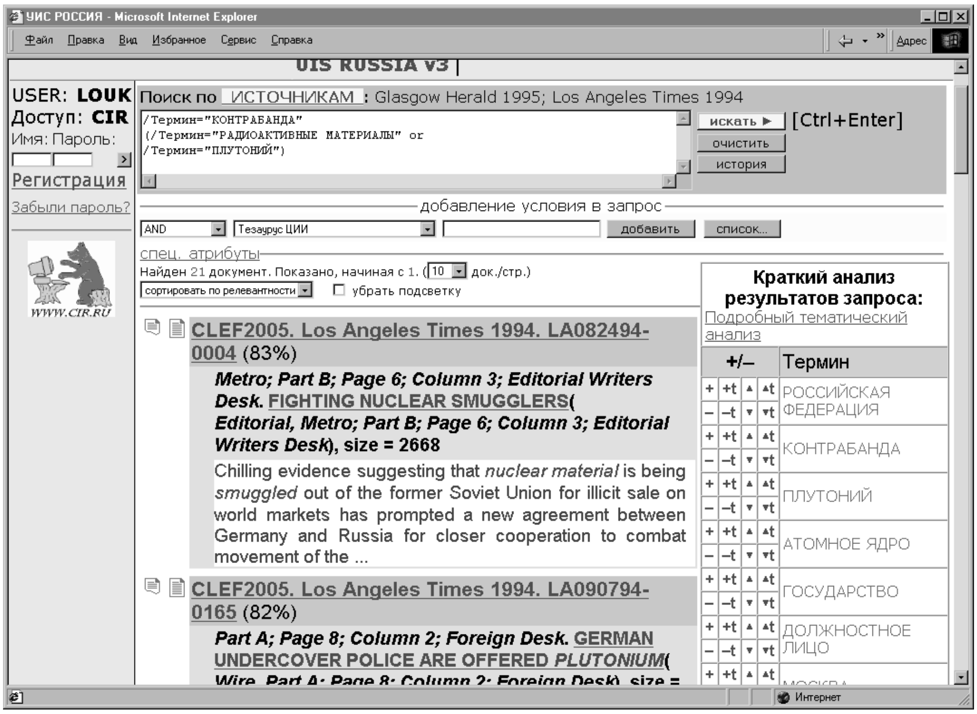
(C_{t1} AND ..AND… C_{tn}) AND (C_{d1} OR ...OR C_{dm} OR C_{n1} OR .. OR C_{nk})

Fig.5 shows results of retrieval of query (C_{t1} AND ..AND… C_{tn}) = CONTRABAND and RADIOACTIVE MATERIALS for topic 264.

Step 2. We try to expand a query using Thesaurus concepts subordinate to the concepts of a query. But it is well-known that the context of a concept in a query can restrict expansion of this concept. Therefore in this stage for expansion we try to justify expansion with a technique similar to blind relevance feedback. For expansion we use only that subordinate concepts of query concepts that are top-ranked 20 concepts from the top-ranked 100 documents. List of such top-ranked concepts is shown in the right column of the screen.

Subordinate concepts are added using OR to their superordinate concepts, forming disjunction. For example at this stage for Query 264 Smuggling of Radioactive Materials concepts URAN and PLUTONIUM are added to concept RADIOACTIVE MATERIALS. So we receive disjunction (RADIOACTIVE MATERIALS or URAN or PLUTONIUM). Fig. 7 shows results of such expansion.

We fulfill expanded queries and add subordinate concepts while new such concepts appear.



USER: **LOUK**
Доступ: **CIR**
Имя: Пароль:

Поиск по ИСТОЧНИКАМ : Glasgow Herald 1995; Los Angeles Times 1994

```
/Термин="КОНТРАБАНДА"
(/Термин="РАДИОАКТИВНЫЕ МАТЕРИАЛЫ" or
/Термин="ПЛУТОНИЙ")
```

добавление условия в запрос

AND | Тезаурус ЦИИ

спец. атрибуты

Найден 21 документ. Показано, начиная с 1. (10 док./стр.)

сортировать по релевантности ☐ убрать подсветку

CLEF2005. Los Angeles Times 1994. LA082494-0004 (83%)

***Metro; Part B; Page 6; Column 3; Editorial Writers Desk.* FIGHTING NUCLEAR SMUGGLERS(*Editorial, Metro; Part B; Page 6; Column 3; Editorial Writers Desk*), size = 2668**

Chilling evidence suggesting that *nuclear material* is being *smuggled* out of the former Soviet Union for illicit sale on world markets has prompted a new agreement between Germany and Russia for closer cooperation to combat movement of the ...

CLEF2005. Los Angeles Times 1994. LA090794-0165 (82%)

***Part A; Page 8; Column 2; Foreign Desk.* GERMAN UNDERCOVER POLICE ARE OFFERED *PLUTONIUM*(*Wire, Part A; Page 8; Column 2; Foreign Desk*), size =**

Краткий анализ результатов запроса:

Подробный тематический анализ

+/–				Термин
+	+t	▲	▲t	РОССИЙСКАЯ ФЕДЕРАЦИЯ
–	–t	▼	▼t	
+	+t	▲	▲t	КОНТРАБАНДА
–	–t	▼	▼t	
+	+t	▲	▲t	ПЛУТОНИЙ
–	–t	▼	▼t	
+	+t	▲	▲t	АТОМНОЕ ЯДРО
–	–t	▼	▼t	
+	+t	▲	▲t	ГОСУДАРСТВО
–	–t	▼	▼t	
+	+t	▲	▲t	ДОЛЖНОСТНОЕ ЛИЦО
–	–t	▼	▼t	
+	+t	▲	▲t	МОСКВА

Figure 7.

Step 3. At this stage we continue to expand the initial query. Now we use full trees of lower concepts for title concepts C_{t1} & ..& C_{tn}. So at this stage we work with the following query ($C_{t1+tree}$ AND ..AND $C_{tn+tree)}$ & (C_{d1} OR ..OR C_{dm} OR C_{n1} OR .. OR C_{nk})

Step 4. At this step we reduce initial query to concepts only from the title, so we have the query (C_{t1} AND ..AND C_{tn}).

Step 5. Concepts from the title are expanded with lower concepts ($C_{t1+tree}$ AND ..AND $C_{tn+tree}$).

Step 6. At this stage we change AND of title concepts to OR and return concepts from the description and narrative to the query ($C_{t1+tree}$ OR ..OR. $C_{tn+tree}$) and (C_{d1} OR ..V C_{dm} OR C_{n1} OR .. OR C_{nk})

Step 7. At last all concepts of a topic are used in OR-query ($C_{t1+tree}$ OR ..OR. $C_{tn+tree}$) OR (C_{d1} OR ..OR C_{dm} OR C_{n1} OR .. OR C_{nk}).

To compare results of the bilingual concept-based run we fulfilled several monolingual English tf.idf runs. Our best tf.idf run was based only on titles (Tabl. 1). In the system tf.idf technique similar to (Callan et.al, 1992) is implemented.

Run Ad-hoc	Type of information	Average precision %	Number of topics better than average
Concept-based Bilingual Russian to English	TDN	22.82	34
tf.idf (Okapi BM25) Monolingual English	T	20.90	11

Table 1.

Conclusion

During more than 10 years we developed bilingual Russian-English Sociopolitical Thesaurus as a resource for automatic text processing in a broad domain of social relations of the contemporary society.

We considered the Thesaurus as a resource useful for application in two tasks of CLEF: in the ad-hoc task based on newspapers and the domain-specific task based on social sciences documents. For automatic processing of documents and queries we used only the Sociopolitical Thesaurus and therefore we can state that the concepts of the Thesaurus indeed provide broad coverage of newspaper texts, scientific abstracts and corresponding CLEF queries.

For the first participation in CLEF we received promising results: average precision in the bilingual ad-hoc task X2En is more than medium.

In future we plan to experiment with weighting schemes for results of Boolean queries, for example, weights for OR queries over thesaurus concepts were evidently unsuccessful and led to serious decrease of results.

Bibliography

Bentivogli L., Pianta E., 2004. Extending WordNetwith Syntagmatic Information. – International Wordnet Conference (GWC – 2004). – 2004. – pp. 47-53.

Callan, J.P., Croft, W.B. and Harding, S.M., 1992. The INQUERY Retrieval System. In A.M. Tjoa and I. Ramos (eds.), *Database and Expert System Applications*. Springer Verlag, New York.

Climent S., Rodriguez H., Gonzalo J., (1996): Definitions of the links and subsets for nouns of the EuroWordNet project. - Deliverable D005, WP3.1, EuroWordNet, LE2-4003.

Guarino, N., 2000. Some Ontological Principles for Designing Upper Level Lexical Resources. In Proceedings of First International Conference on Language Resources and Evaluation.

Hirst G., St-Onge D., 1997. Lexical Chains as representation of context for the detection and correction malapropisms, In: C. Fellbaum, editor, *WordNet: An electronic lexical database and some of its applications*, Cambridge, MA: The MIT Press.

LIV (Legislative Indexing Vocabulary), (1994): Congressional Research Service. The Library of Congress. Twenty-first Edition.

Loukachevitch N., Dobrov B., 2000. Thesaurus-Based Structural Thematic Summary in Multilingual Information Systems. *Machne Translation Review*, 11:10-20.

Loukachevitch N., Dobrov B., 2002. *Evaluation of Thesaurus on Sociopolitical Life as Information Retrieval Tool.* In: “Proceedings of Third International Conference on Language Resources and Evaluation (LREC2002)”, M.Gonzalez Rodriguez, C. Paz Suarez Araujo, eds. – Vol.1 – Gran Canaria, Spain, pp.115—121.

Loukachevitch N., Dobrov B., 2004a. Development of Ontologies with Minimal Set of Conceptiul Relations. In: “Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC2004)”, – Vol.6 – pp.1885—1889.

Loukachevitch N., Dobrov B., 2004b. Development of Bilingual Domain-Specific Ontology for Automatic Conceptual Indexing. In: “Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC2004)”, – Vol.6 – pp.1993—1996.

Loukachevitch N., Dobrov B., 2004c. Sociopolitical Domain as a Bridge from General Words to Terms of Specific Domains // Proceedings of Second International WordNet Conference GWC 2004. – pp.163-168.

Miller G., Beckwith R., Fellbaum C., Gross D., Miller K., 1990. Five papers on WordNet, *CSL Report, 43*, Cognitive Science Laboratory, Princeton University.

Rondeau G., 1980. Introduction a terminologie. Quebec, 1980.

Salton G., 1989. Automatic Text Processing - The Analysis, Transformation and Retrieval of Information by Computer. Addison-Wesley, Reading, MA, 1989.

Soergel D., Lauser B., Liang A., Fisseha F., Keizer J., Katz S., (2004). Reengineering Thesauri for New Applications: the AGROVOC Example. – Journal of Digital Information. Volume 4, Issue 4. - Article No. 257, 2004-03-17.

Tudhope D., Alani H., Jones Cr., (2001). Augmenting Thesaurus Relationships: Possibilities for Retrieval. – Journal of Digital Libraries. Volume 1, Issue 8. – 2001.