

Combining Formal Concept Analysis and semantic information for building ontological structures from texts : an exploratory study

Sílvia Maria Wanderley Moraes, Vera Lúcia Strube de Lima

PUCRS- Faculdade de Informática

Avenida Ipiranga, 668. Prédio32. Sala 630. 90619-900. PortoAlegre. RS. Brasil

E-mail: silvia.moraes@pucrs.br, vera.strube@pucrs.br

Abstract

This work studies conceptual structures based on the Formal Concept Analysis method. We build these structures based on lexico-semantic information extracted from texts, among which we highlight the semantic roles. In our research, we propose ways to include semantic roles in concepts produced by this formal method. We analyze the contribution of semantic roles and verb classes in the composition of these concepts through structural measures. In these studies, we use the Penn Treebank Sample and SemLink 1.1 corpora, both in English.

Keywords: Natural Language Processing; Conceptual Structures; Non-Taxonomic Relations; Semantic Roles; Formal Concept Analysis.

1. Introduction

Conceptual structures such as terminologies, thesauri, taxonomies and ontologies are important resources for information systems. Since building and maintaining such structures is costly, semi-automatic approaches have been proposed to minimize the effort of extracting concepts and semantic relations from texts. We are interested in the learning of conceptual structures, exploring semantic relations between verbs and their arguments. Our work analyzes the contribution of semantic roles and verb classes in the building of concepts. A semantic role expresses the meaning of an argument in a situation described by the verb in a sentence. With the use of semantic roles, we can identify, for example, the agent entity of an action, even if it appears in different syntactic positions through the text. In their turn, verb classes correlate verbs according to their syntactic behavior. Since verbs of the same class have some kind of semantic relation (Levin & Hovav, 1996), this information may help disambiguating polysemic verbs and, ultimately, distinguishing the semantic context of these verbs' arguments.

The Formal Concept Analysis (FCA) method generates groups of concepts and provides an intensional

description to these groups. This description favors the traceability of the process of building ontological structures and renders the generated groups easier to interpret. Although the FCA method is not new, it has been proposed that it be used as a support resource for building and mapping ontological structures (Cimiano,2006; Pal'chunov, 2007; Priss,2006). Despite the multiple applications of this method, there are still few studies exploring the semantic aspects in conceptual structure building based on FCA, particularly regarding semantic roles and verb classes.

In this study, we extracted these semantic information from the Penn TreeBank Sample¹ and SemLink² 1.1 corpora. In order to analyze the contribution of these data to concept building, we used structural measures.

In Section 2 we describe related works. In Section 3 present a brief approach to semantic roles and verb classes. Section 4 describes the FCA method and Section 5 discusses approaches to combine FCA with semantic information. Finally, in Section 6 we present our conclusions.

¹http://nltk.googlecode.com/svn/trunk/nltk_data/index.xml

²<http://verbs.colorado.edu/semlink/>

2. Related Works

Kamphuis e Sarbo (1998) propose the representation of a sentence in natural language, associating FCA to semantic roles. In that work, these authors dealt with two types of linguistic relations: minor and major. The minor one typically related nouns to adjectives and adverbs; the major one, verbs to nouns. Though the approach seemed promising at the time it was proposed, it has not been fully explored yet, probably due to the difficulty tagging, as the appearance of automatic taggers of semantic roles is more recent. Our study, differently from the authors' work, extracts the relations from linguistically tagged texts. Furthermore, it is restricted to the relations the authors call major. It also does not aim the interpretation of texts, only approaching this interpretation, as it addresses the automatic construction of conceptual structures.

Rudolf Wille (1997) also presents examples of FCA structures combined with semantic roles. The author's objective, however, was to combine conceptual graphs with FCA structures, aiming the formalization of useful logic to knowledge representation and processing. As there are no comments, in that work, on information processing present in the conceptual graphs, we imagine that neither the construction of these graphs nor the mapping in FCA structures were performed automatically. This is, thus, another different aspect of our study. Rudolf Wille's work does not deal with the difficulties of automatically extracting information from texts to generate representation structures nor analyzes, in this sense, the limits of his approach.

In more recent works, we have also found the FCA method combined with semantic roles. An example is the work of Valverde-Albacete (2008). Distinctly from our work, the author does not use the FCA as a support method to build ontological structures from texts. His effort turns to the linguistic analysis as a purpose for representing FrameNet through conceptual lattices. Accordingly, it does not use, as we do, textual information nor PropBank notations to identify the roles.

3. Verb Classes and Semantic Roles

Verbal behavior regarding syntactic inflection and

syntactic function allows us to establish morphosyntactic classifications for verbs. Beth Levin's work about the semantic classification of verbs (Levin & Hovav, 1996) is among the most quoted studies in this area. Kipper, while studying Levin's work in (Kipper,2005), pointed out that verbs of the same class are not necessarily synonyms. Some classes, such as Break (break, chip, crack, fracture, rip, ...), contain verbs with closely related meanings, while others, such as Braid (braid, brush, clip, comb, curl...) don't (Kipper,2005).

Patterns identified by Levin have fostered the development of automatic and semi-automatic tools for classifying verbs, tagging semantic roles, disambiguating the meaning of verbs and also enabling lexical resources such as VerbNet³ (Kipper,2005) and PropBank⁴ (Palmer et al., 2005). VerbNet is considered an extension of Levin's work, because it organizes verbs in hierarchical classes according to their semantic and syntactic attributes. The above mentioned Break and Braid classes are, respectively, entries 45.1 and 41.2.2 from VerbNet. PropBank is a manually written corpus that includes semantic information and provides notes on semantic roles for sentences in Treebank-2⁵.

Semantics roles, also called thematic roles or Θ -roles, are "roles within the situation described by a sentence" (Yule,1996). There is no consensual list of semantic roles, but a few are widely accepted: Agent (a human or at least animated entity, which provokes an action or event), Patient (an entity directly affected by an action, which changes its state), etc. In recent approaches, the barrier regarding the definition of roles has been circumvented by assigning numerical labels (Arg0,Arg1,Arg2 ,...) to the arguments of the verbs (Palmer et al, 2005). This is the case for PropBank. The problem with the notations provided by PropBank is that they are not uniform for verbs from different classes.

This is one of the reasons which led us to base our study on the SemLink 1.1 corpus instead of PropBank. SemLink 1.1 is an extension of PropBank, adding VerbNet information on verb classes and semantic role

³<http://verbs.colorado.edu/~mpalmer/projects/verbnnet>

⁴<http://www ldc.upenn.edu/>

⁵TreeBank-2 is composed of 2,499 articles from the 1989 editions of the Wall Street Journal.

names to the original tags. SemLink mapping is still incomplete, because not all PropBank roles were matched to VerbNet roles. Also, there are verbs yet to be associated to a class. Both the SemLink and the PropBank corpus only use the linguistic tags that come from Treebank-2. For this reason we included in our study the Penn TreeBank Sample corpus, corresponding to 10% of the Treebank-2 corpus. Although it is a small corpus (199 texts), it can be freely downloaded and contains enough verb-argument relations to be considered in our research.

4. Formal Concept Analysis

FCA is a method used for data analysis, knowledge representation and information management. The main idea behind the method is the duality known as the “Galois connection”, which establishes implicit relations between objects and attributes in such a way that objects can be described by their attributes and the attributes by the objects they characterize (Stumme *et al.*, 1995).

A key element in this method is formal context, characterized by the triple (G, M, I) , where: G is the set of domain entities, called formal objects; M consists of the features of these entities, their formal attributes; and I is the binary relation of $G \times M$, called the incidence relation, which associates a formal object to its attribute (Stumme *et al.*, 1995). We can extract formal contexts from syntactic dependencies between verbs and their complements (subject, direct and indirect object). Table 1 presents formal context based on a subset of the verb-argument pairs obtained for “share”.

The arguments of the verbs constitute the set of formal objects $G = \{share, stockholder, shareholder\}$ and the verbs constitute the set of formal attributes $M = \{receive, pay, buy, sell\}$. Set I establishes the relations between the elements in G and those in M , such as that of the formal object “*stockholder*” and its attributes, “*receive*” and “*sell*”.

Based on the formal context, the formal concepts are built. A formal concept in (G, M, I) is determined by the pair (O, A) if and only if $O \subset G$, $A \subset M$, so that $f(O)=A$ and $h(A)=O$, where the operator f defines

the properties shared by all the elements in O and the operator h determines the objects that contain all of the properties in A . The following pairs are examples of formal concepts of FCA illustrated in Figure 1: $(\{share, shareholder, stockholder\}, \{receive\})$, $(share, shareholder, \{receive, pay, buy\})$, $(\{share, stockholder\}, \{receive, sell\})$.

Table 1: Formal context for “share”

	receive	pay	buy	sell
share	x	x	x	x
stockholder	x			x
shareholder	x	x	x	

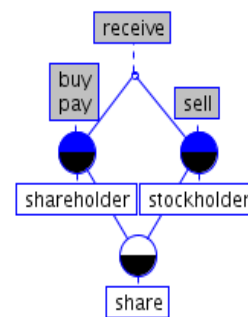


Figure 1: Conceptual lattice

5. Approaches for building ontological structures from text

We explore the contribution of verb classes and semantic roles in the building of clusters that define the formal concepts of a FCA lattice. We chose 10 seed terms from finances and searched the Penn TreeBank Sample *corpus* for all subject-verb-complement tuples that had an absolute frequency of at least 2 and that were related to one of the seeds. From each tuple we extracted the nouns of the noun phrase that belonged to the subject and its complements. Using SemLink 1.1 we associated semantic roles to the nouns and the verb classes.

We organized our research in four case⁶ studies.

5.1. Cases

In each case, the G set of formal objects is the same (i.e. it is composed by the nouns that appear as subject or complement). The M set is defined as follows:

⁶ With each case study, we specify a subscript that identifies a incidence relation from FCA lattice. We use *np* for *noun phrase*; *v* for *verb*; *sr* for *semantic role*; and *c* for *VerbNet class*.

- case 1_(np,v): FCA lattice is built as in Figure 1 (its attributes are verbs).
- case 2_(np,c): In order to analyze the influence of verb classes, we built the FCA lattice using classes verb as attributes (Figure 2).
- case 3_(np,sr): In order to analyze the influence of semantic roles, we built the FCA lattice using such roles as attributes (Figure 3).

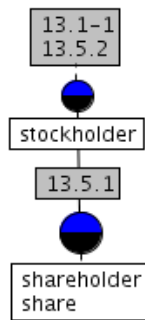


Figure 2: Conceptual lattice for case 2

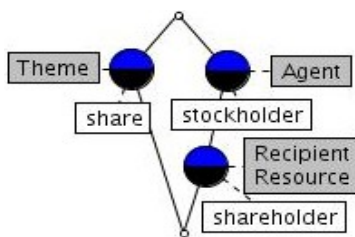


Figure 3: Conceptual lattice for case 3

- case 4_(np,sr_np): Since semantic roles are relations, we defined the set of attributes as pairs entitled “semanticRole_of_nounPhrase”. For example, the tuple *stockholder-receive-share*, where the semantic role of *stockholder* is “agent” and that of *share* is “theme”, the attributes in M are $\{agent_of_share, theme_of_stockholder\}$. Ergo, the relations $(stockholder, agent_of_share)$ and $(share, theme_of_stockholder)$ are elements of I (Figure 4).

The four lattices presented as toy examples in Figures.1, 2, 3 and 4 were generated from the same tuples verb-argument, but with different formal attributes.

In the next section, we describe the assessment criteria used to analyze the formal concepts and the effective

contribution of semantic information included in their attributes.

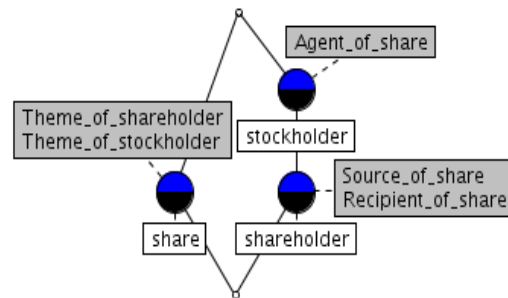


Figure 4: Conceptual lattice for case 4.

5.2. Assessment metrics

The evaluation of conceptual structures, though extensively researched, is not yet a consolidated theme. When we evaluate FCA-based structures, there are still more difficulties because this kind of investigation is more recent. We only found two measures for this type of assessment: zeros-induced (Alqadash & Bhatnagar, 2009) and Sim (Formica, 2008) measures. Since our goal was to analyze the formal concepts adopting a semantic approach, from the two studied measures, only the Sim measure met our purpose. However, it addressed the purpose partially, since the formal context attributes of our study cases are of different types, hard to be compared. A comparison difficulty, namely the class of a verb with a semantic role, prevented the use of this measure. Thus, we focused our analysis on formal objects. We needed a measure of structural order to semantically assess the sets of objects of a formal concept. Therefore, we would be able to verify which set generated groupings whose semantic relations among objects were more representative. For this we used the Semantic Similarity Measure (SSM) from the AKTiveRank system (Alani & Brewster, 2006), that calculates how close, in a certain ontology, the concepts which exactly or partially combine with informed terms are. As the metrics were applied to formal objects of each concept of the analyzed FCA structures, it turns out SSM worked as a kind of measure for lexical cohesion. Halliday and Hasan (1976) use the term cohesion to refer to "relations of meaning that exist within the text". According to these authors, cohesion happens when

interpretation of an element is dependent on another element of discourse. It is expressed through grammar as well as vocabulary. In the last case, it is called lexical cohesion, which analyzes the semantic relation among words within the text. Lexical cohesion is based on relations such as synonym, hyponym, meronym and antonym to determine relations of meaning among the words within the text (Halliday & Hasan, 1976). Works such as Teike and Fankhauser (2005), though with a different purpose, use WordNet to measure lexical cohesion. Teike e Fankhauser aim to help the notation of texts by automatically identifying n-grams whose elements are more strongly related. Lexical cohesion is determined having as base the length of the shortest path, existent in the WordNet hierarchy, between the synsets of the terms under analysis. Looking at works of this type, we employed a measure, commonly applied to WordNet, to determine the lexical cohesion of the objects of a formal concept. The chosen measure was defined by Wu and Palmer (1994). Its use is based on two reasons. The first one, Alani and Brewster (2006) mentioned it, as one of the measures that could be used to calculate SSM metric. Another reason is that the NLTK⁷ pack implements such measure for WordNet and we could easily use it. It is also worth mentioning that the measure generates normalized values, which facilitate its interpretation.

We also opted for applying the SMM metrics to the LSDIS Finance⁸ ontology, and this decision has equally been made for two reasons. The first is that, though the extension and wealth in relations of the WordNet base, such relations do not refer to a specific domain. As this is our situation, we imagined that Wu and Palmer's measure, applied to the WordNet, might not capture the expected semantic relation and generate less expressive values. The second is that, though in the LSDIS Finance ontology the set of concepts is smaller, concepts labeled with n-grams ($n > 1$) are more usual and the relations between these concepts are known. These factors may generate results more semantically meaningful regarding the grouping quality (concepts).

⁷<http://www.nltk.org/>

⁸http://lsdis.cs.uga.edu/projects/meteor-s/wsd1-s/ontologies/LSDIS_Finance.owl

In the case of the LSDIS Finance ontology, besides implementing SSM metrics, we also had to write programs for the calculation of Wu and Palmer's measure for such conceptual structure. The measure, expressed in Equation 1, indicates the average lexical cohesion of the concepts of an FCA structure in relation to an E conceptual structure.

$$SSM_E = \frac{1}{N} \sum_{i=1}^N ssm_i \quad (1)$$

On the other hand, the measure (Equation 2) calculates the similarity between the objects of a formal concept, based on Wu and Palmer's measure. In case this set of objects G_i has cardinality 1, the measure is zero.

$$ssm_i = \frac{1}{|G_i|} \sum_{j=1}^{|G_i|-1} \sum_{k=j+1}^{|G_i|} wup_E(o_j, o_k) \quad (2)$$

for $o_j, o_k \in G_i$

Finally, the measure, shown in Equation 3, estimates the similarity between the concepts in an E structure. In this equation, a corresponds to the common and more specific ancestral of concepts c_1 e c_2 ; p , to the depth of any node, in other words, the path length (in nodes) of this node to the root node; and d , the shortest distance (em nodes) de c_1 a c_2 .

$$wup_E(c_1, c_2) = \frac{2p(a(c_1, c_2))}{d(c_1, a(c_1, c_2)) + d(c_2, a(c_1, c_2)) + 2p(a(c_1, c_2))} \quad (3)$$

We also analyzed the relation between cardinality of the attribute set of each structure with the amount of formal concepts produced and, still, the height and width estimation of these structures. Another element assessed is the amount of edges in these structures: the higher is this value, the greater the complexity in the construction of the correspondent lattice.

In the following section, the analyses performed are commented in the study cases proposed according to the structural measures described in this section.

5.3. Case analysis

Our goal, as mentioned, was to investigate the influence of semantic roles in the building of formal concepts.

With this objective, we analyzed two settings. One of them included only the 4 most frequent semantic roles

(Agent, Theme, Patient and Topic) and another considered all semantic roles available in the selected tuples (between 18 and 20 roles).

Table 2 describes the results regarding the SSM structural measure for the 4 studied cases. In this table we present data about the number of objects and attributes of each formal context analyzed. In the table, we also included the amount of formal concepts generated and the SSM measures calculated. The SSM measure corresponds to the lexical cohesion concerning the WordNet base; and the SSM, to the lexical cohesion concerning the LSDIS Finance ontology. The last column of the table presents the arithmetic mean of these 2 measures.

By observing the data in Table 2, we can notice that only case 4, which contains the relation (np, sr_np), obtained, in the mean, lexical cohesion below case 1's. That happened due to the attribute specificity of the semanticRole_of_nounPhrase (sr_np) form. Few objects shared such attributes, what turned out to produce many concepts whose objects set cardinality was 1 (around 58.3% of the formal concepts, according to Table 3).

Analyzing case 2 (Table 3), it is perceived that, besides the improvement in cohesion, there was a reduction: in the total number of concepts (largest grouping), in the number of concepts with a unitary set of objects (best grouping) and the amount of edges (less processing).

Of all cases analyzed, case 3 was the one that obtained the highest index of lexical cohesion. However, it concentrated the objects in few concepts. The generality of its attributes, which are semantic roles, might be the reason for high cohesion. The presence of more objects in the same formal concept increases the amount of combinations of pairs of objects that are submitted to Wu and Palmer's similarity measure. Since these objects have some semantic relation (at least the one defined by the semantic role itself), the resulting similarity ends up being greater.

However, when we included all the semantic roles in the formal contexts of the study cases, we noticed a decrease in case 3's lexical cohesion (Table 4). This cohesion fall might be related to the addition of attributes (from 4 to up to 20 semantic roles) and to low frequency in most of

them. In general, more attributes cause a broader object distribution in concepts. In this specific case, it increased not only the amount of concepts but also of concepts with unitary set of objects, which raised to 25%.

Table 2: Results of the SSM structural measure regarding the 4 studied cases (4 semantic roles)

case	#obj.	#atr.	#conc.	SSM _w	SSM _f	mean
1 (np,v)	178	71	119	0,21	0,11	0,16
2 (np,c)	178	45	99	0,23	0,13	0,18
3 (np,sr)	178	4	12	0,48	0,47	0,48
4 (np, sr_np)	178	214	151	0,09	0,05	0,07

Table 3: Complementary structural measures applied to the 4 studied cases (4 semantic roles)

case	#edges	#unitary(%)	height	width
1 (np,v)	237	35 (29,4)	5	56
2 (np,c)	211	34 (34,3)	5	40
3 (np,sr)	20	0	4	4
4 (np, sr_np)	271	88 (58,3)	5	100

We also analyzed the non-taxonomic relations existent in the LSDIS Finance ontology. Conversely, we did not find any "perfect" matching between formal concepts and these relations.

Still with the goal of improving results, particularly case 4's, we included some heuristics in the pre-processing of formal contexts in the studied cases. These heuristics are next section's topic.

Table 4: Results of SSM structural measure regarding the 4 studied cases (all semantic roles)

case	#obj.	#atr.	#conc.	SSM _w	SSM _f	mean
1 (np,v)	377	153	289	0,21	0,11	0,16
2 (np,c)	377	86	261	0,23	0,13	0,18
3 (np,sr)	377	20	356	0,32	0,17	0,25
4 (np, sr_np)	377	579	151	0,09	0,04	0,07

5.4. Heuristics application

Taking the example of Otero *et al.* (2008), which inspired the case 4, we applied a heuristic to group similar attributes, based on the Dice coefficient. For each attribute analyzed, new relations are generated from their neighbor k (more similar attributes, according to Dice's measure). Otero *et al.* used $k=5$. We tested values 4, 5 and 6 for k . We established as similar those attributes whose measures generated values in the (0.5;1) interval.

We also deleted less frequent attributes. We tested 3, 4 and 5 as cut-off values for m , which corresponds to the minimum frequency required for the attributes.

Though the heuristics improved the SSM measure of all 4 cases, the grouping heuristics based on the Dice coefficient was more effective for case 4. For the other cases, k values were not as decisive for the SSM results. The cut-off point was what prevailed for these cases. Table 5 shows data of the best SSM means obtained for the 4 cases after the application of heuristics. The set that generated more significant results was $k=4$ and $m=5$ for all 4 cases.

It is interesting to highlight that, even though with the lowest SSM mean, case 4 generated concepts as dense as case 2 in number of objects (around 5 by concept). Another aspect to take into consideration is that lexical cohesion of concepts in this case, particularly in relation to domain ontologies, improved significantly. Among all cases, it obtained the highest SSM value for the LSDIS Finance ontology.

Table 5: Results of SSM structural measure regarding the 4 studied cases, after application of heuristics considering $k=4$ and $m=5$ (all semantic roles)

case	#obj.	#conc	%unit.	SSM _W	SSM _L	mean
1 (np,v)	215	109	17,4	0,39	0,18	0,29
2 (np,c)	234	114	19,3	0,39	0,26	0,33
3 (np,sr)	263	53	20,8	0,33	0,25	0,29
4 (np, sr np)	140	28	21,4	0,15	0,28	0,22

Case 2 generated the most cohesive concepts. We observed that in relation to WordNet as well as in relation to the LSDIS Finance ontology, the SSM measure, for this case, had significant results. Case 1, in average cohesion, was as good as case 3. The applied heuristics significantly increased the SSM measure in relation to WordNet. However, for the Finance ontology, such measure was the lowest among the cases. This might be an indication that semantic information included in the formal concepts makes a valuable contribution to capture domain relations.

To qualitatively analyze the concepts that include information related to semantic roles, we generated small examples for cases 3 and 4 (Figures 5 and 6). Analyzing

the structures, we noticed a certain similarity between concepts. Nevertheless, attributes in case 4 (Figure 6) are more informative, as they are based on relations. Apparently, such attributes can better delineate the domain semantics because they express the context in which the semantic roles are applied. On the other hand, attributes in case 3 (Figure 5) form sets of objects apparently more comprehensive. This might be one explanation for the fact that the SSM measure, when applied to WordNet, generates higher values for this case.

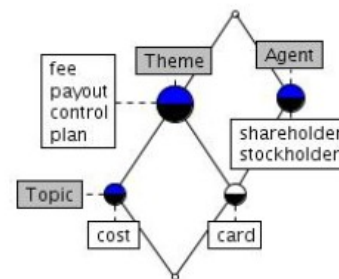


Figure 5: Conceptual lattice for case 3.

We observed that the unigrams *share*, *price*, *trade* and *company* that appear in case 4's attributes, are not part of its set of formal objects. They might have been discarded due to their association to less frequent attributes (lexico-semantic contexts). We also noticed that the semantic roles Cause and Predicate do not appear in case 4's FCA structure, only in case 3's. Therefore, cases 2 and 3 are not necessarily based on the same semantic roles.

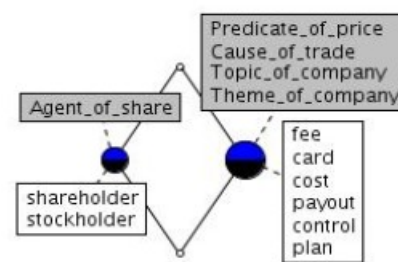


Figure 6: Conceptual lattice for case 4.

6. Conclusions

From the structural and lexical point of view, we observed that the inclusion of semantic information in attributes of formal contexts, in general, had more cohesive formal concepts as a result. In this sense, the verb classes, for the formal context set proposed in case 2, proved to be more effective than verbs. The classes,

besides increasing lexical cohesion, helped reducing the complexity of the building of the FCA lattice, as they generated less concepts and edges. On the other hand, the semantic roles proved to be more effective, still in the cohesion aspect, especially when the formal context set proposed in case 3 was used.

In spite of these results, the interpretation of the structures generated here is not as objective as those structures in which verbs are used as attributes. Under the intensional aspect, the use of numerical labels for the verb classes, as well as the use of semantic roles as classes and not as relations, makes such elements, while attributes, less informative than verbs. However, the formal context set described by case 4, in which the semantic roles are used as relations, presents attributes that seemed more intensionally descriptive to us, yet initially (before the application of heuristics) such configuration had produced less cohesive concepts.

Considering the conducted study, we believe our approach to be promising. Our next step is to analyze it using metrics of functional order. We will test our approach applicability in other *corpora* and domains starting from the task of text categorization.

7. References

- Alani,H. and Brewster,C. (2006) Metrics for ranking ontologies, In Proceedings of the 4th International Workshop on Evaluation of Ontologies for the Web (EON2006) at the 15th International World Wide Web Conference (WWW 2006), Edinburgh, Scotland, 24--30.
- Alqadash, F. and Bhatnagar, R. (2009). Similarity measures in formal concept analysis, *Knowledge-Based System* 21, 80--87.
- Cimiano, P. (2006). *Ontology learning and population from text: Algorithms, evaluation and applications*, Springer.
- Formica, A. (2008). Concept similarity in formal concept analysis: An information content approach, *Knowledge-Based System* 21, 80--87.
- Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English (English Language)*, Longman, Pub Group.
- Kamphuis, V. and Sarbo, J. (1998) Natural language concept analysis, In Proceedings of International Conference on New Methods in Language Processing and Computational Natural Language Learning, ACL, Yellow Book, 205--214.
- Kipper,K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*, These, University of Pennsylvania.
- Levin, B. and Hovav,M. (1996). *Lexical semantic and syntactic structure*. The Handbook of Contemporary Semantic Theory, Blackwell, Oxford.
- Otero, P. G, Lopes,G. P. and Agustini (2008). A. Automatic acquisition of formal concepts from text.*LDV Forum* 23,1, 59--74.
- Palmer, M., Gildea, D. and Kingsbury, P. (2005). The Proposition Bank: An annotated corpus of semantic roles", *Computational Linguistic*,31 (1):71-106.
- Pal'chunov, D.E. (2007). Lattices of relatively axiomatizable classes", *ICFCA*, 221-239.
- Priss, U. (2006). Formal Concept Analysis in Information Science *Annual Review of Information Science and Technology*, 40.
- Stumme, G., Darmstadt, T. H. and Mathematik, F. (1995). Exploration tools in formal concept analysis, *In Proceedings of Studies in Classification, Data Analysis, and Knowledge Organization* 8, Springer, 31-44.
- Teike, E. and Fankhauser (2005) Exploring lexical patterns in text:lexical cohesion analysis withWordNet. *Interdisciplinary Studies on Information Structure* 02. Heterogeneity in Focus: Creating and Using Linguistic Databases, 129—145.
- Valverde-Albacete,F.J.(2008). Extracting frame-semantics knowledge using lattice theory, *J Logic Computation* 18 (3): 361—384.
- Wille,R. (1997). Conceptual graphs and formal concept analysis, In Proceedings of the Fifth International Conference on Conceptual Structures: Fulling Peirce's Dream (London, UK), ICCS'97, Springer-Verlag, 290--303.
- Wu, Z. and Palmer, M. (1994). Verb semantics and lexical selection. *In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, New Mexico, 133--138.
- Yule, G. (1996). *The study of language*. Cambridge University Press, Cambridge, 2nd edition.