

I T H E A

International Journal
INFORMATION **MODELS**
&
ANALYSES

2014 **Volume 3** **Number 4**

**International Journal
INFORMATION MODELS & ANALYSES
Volume 3 / 2014, Number 4**

EDITORIAL BOARD

Editor in chief: **Krassimir Markov** (Bulgaria)

Adil Timofeev	(Russia)	Levon Aslanyan	(Armenia)
Albert Voronin	(Ukraine)	Luis Fernando de Mingo	(Spain)
Aleksey Voloshin	(Ukraine)	Liudmila Cheremisinova	(Belarus)
Alexander Palagin	(Ukraine)	Lyudmila Lyadova	(Russia)
Alexey Petrovskiy	(Russia)	Martin P. Mintchev	(Canada)
Alfredo Milani	(Italy)	Nataliia Kussul	(Ukraine)
Anatoliy Krissilov	(Ukraine)	Natalia Ivanova	(Russia)
Avram Eskenazi	(Bulgaria)	Natalia Pankratova	(Ukraine)
Boris Tsankov	(Bulgaria)	Nelly Maneva	(Bulgaria)
Boris Sokolov	(Russia)	Olga Nevzorova	(Russia)
Diana Bogdanova	(Russia)	Orly Yadid-Pecht	(Israel)
Ekaterina Detcheva	(Bulgaria)	Pedro Marijuan	(Spain)
Ekaterina Solovyova	(Ukraine)	Rafael Yusupov	(Russia)
Evgeniy Bodyansky	(Ukraine)	Sergey Krivii	(Ukraine)
Galyna Gayvoronska	(Ukraine)	Stoyan Poryazov	(Bulgaria)
Galina Setlac	(Poland)	Tatyana Gavrilova	(Russia)
George Totkov	(Bulgaria)	Valeria Gribova	(Russia)
Gurgen Khachatryan	(Armenia)	Vasil Sgurev	(Bulgaria)
Hasmik Sahakyan	(Armenia)	Vitalii Velychko	(Ukraine)
Iliia Mitov	(Bulgaria)	Vladimir Donchenko	(Ukraine)
Juan Castellanos	(Spain)	Vladimir Ryazanov	(Russia)
Koen Vanhoof	(Belgium)	Yordan Tabov	(Bulgaria)
Krassimira B. Ivanova	(Bulgaria)	Yuriy Zaichenko	(Ukraine)

**IJ IMA is official publisher of the scientific papers of the members of
the ITHEA® International Scientific Society**

IJ IMA rules for preparing the manuscripts are compulsory.

The rules for the papers for ITHEA International Journals are given on www.ithea.org.

The camera-ready copy of the paper should be received by ITHEA® Submission system <http://ij.ithea.org>.

Responsibility for papers published in IJ IMA belongs to authors.

International Journal "INFORMATION MODELS AND ANALYSES" Volume 3, Number 4, 2014

Edited by the **Institute of Information Theories and Applications FOI ITHEA**, Bulgaria, in collaboration with

Institute of Mathematics and Informatics, BAS, Bulgaria,

V.M.Glushkov Institute of Cybernetics of NAS, Ukraine,

Universidad Politecnica de Madrid, Spain,

Hasselt University, Belgium

Institute of Informatics Problems of the RAS, Russia,

St. Petersburg Institute of Informatics, RAS, Russia

Institute for Informatics and Automation Problems, NAS of the Republic of Armenia,

Publisher: **ITHEA®**

Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org, e-mail: info@foibg.com

Technical editor: **Ina Markova**

Printed in Bulgaria

Copyright © 2012-2014 All rights reserved for the publisher and all authors.

© 2012-2014 "Information Models and Analyses" is a trademark of ITHEA®

© ITHEA is a registered trade mark of FOI-Commerce Co.

ISSN 1314-6416 (printed)

ISSN 1314-6432 (Online)

RDFARM - A SYSTEM FOR STORING LARGE SETS OF RDF TRIPLES AND QUADRUPLES BY MEANS OF NATURAL LANGUAGE ADDRESSING

Krassimira Ivanova

Abstract: *In this paper we present results from experiments for storing middle-size and large sets of RDF triples and quadruples by means of Natural Language Addressing. For experiments we have realized program RDFArM aimed to store RDF triples and quadruples in multi-layer hash tables (information spaces with variable size). The main features of program RDFArM are outlined in the paper. Analysis of the experimental results and rank-based multiple comparison are discussed.*

Keywords: *Natural Language Addressing; Rank-based Multiple Comparison*

ACM Classification Keywords: *H.2 Database Management; H.2.8 Database Applications*

Introduction

The idea of Natural Language Addressing (NLA) [Ivanova et al, 2012a; 2012b; Ivanova et al, 2013a; 2013b; 2013c; 2013d; 2013e; Ivanova, 2013; Ivanova, 2014] consists in using the computer encoding of name's (concept's) letters as logical address of connected to it information stored in a multi-dimensional numbered information spaces [Markov, 1984; Markov, 2004; Markov, 2004a]. This way no indexes are needed and high speed direct access to the text elements is available. It is similar to the natural order addressing in a dictionary where no explicit index is used but the concept by itself locates the definition.

In this paper we present results from experiments for storing middle-size and large sets of RDF triples and quadruples [Klyne & Carroll, 2004] through Natural Language Addressing. For experiments we have realized program RDFArM based on NLA Access Method and corresponded NLA Archive Manager called NL-ArM [Ivanova, 2014]. RDFArM is aimed to store RDF triples and quadruples in multi-layer hash tables (information spaces with variable size). Each RDF element can be stored by appropriate path, which is set by a natural language word or phrase.

Below we will present shortly main features of program RDFArM and after that we will present several experiments with middle-size and large data sets. Analysis of the experimental results and rank-based multiple comparison conclude the paper.

RDFArM

The data of RDFArM are encoded in N-Triples or N-Quads format. The N-Quads is a format that extends N-Triples with context. Each triple in an N-Quads document can have an optional context value [N-Quads, 2013]:

<subject> <predicate> <object> <context>.

as opposed to N-Triples, where each triple has the form:

<subject> <predicate> <object>.

The main idea for storing RDF-graphs in RDFArM follows the one of *multi-layer representation* [Ivanova et al, 2012b]. In other words, the RDF-relations are assumed as layers and the RDF-subjects are assumed as paths valid for all layers. The objects as well as contexts are stored in the containers located by the path in the corresponded layers.

Screenshots from the RDFArM program are shown on Figures 1a and 1b. The main functions are RDF-Write and RDF-Read for which there are corresponded buttons.

By "**RDF-Write**" button the function for storing RDF triples or quadruples from a file can be activated. The recognition of the case (triples or quadruples) is made automatically. The lines of triples do not contain the fourth element, i.e. the context of the quadruples.

Each triple (subject, relation, and object) or quadruple (subject, relation, object, and context) occupy one record in the input file. There is no limit to the number of records in the file. After pressing the "RDF-Write" button, the system reads records sequentially from the file and after storing the triples or quadruples, it displays two informative lines in the panel near to the "RDF-Write" button (Figure 1a):

- Total time used for storing all instances from the file;
- Average time used for storing of one instance (in milliseconds).

The time used is highly dependent on the possibilities of the operational environment and the speed of the computer hardware.




By "**RDF-Read**" button the function for reading RDF triples or quadruples from the RDFArM archives can be activated (Figure 1b). RDF-Read uses as input a file with requests similar to SPARQL requests [SPARQL, 2013] and extracts from the archives the requested information. The requested elements may be given by <?>. In other words, if any of parameters are not given, i.e. <subject>, <predicate>, <object>, or <context>, as in SPARQL requests, the rest are used as constant addresses and omitted parameters scan all non empty co-ordinates for given position. This way all possible requests like (?S-?P-?O), (S-P-?O), (S-?P-O), (?S-P-O), etc., are covered (S stands for subject, P for property, O for object).

In the panel next to the RDF-Read button, two informative lines are shown (Figure 1b) (in milliseconds):

- Total time used for extracting of all quadruple instances;
- Average time used for extracting of one instance.

The time used is highly dependent on possibilities of operational environment and speed of computer hardware.

The RDFArN form has three **service** buttons:

- The first () serves as a transition to the form for manual input and output of data to/from the system archive;
- The second () is connected to the module for adjusting the environment of the system – archives, input and output information, etc.;
- The third () activates the help text (user guide) of the system.

In the same panel there is a button which enables deleting the work archives of the RDFArM (for test control in this version, they are stored on the hard disk but not in the computer memory). RDFArM is completed with compressing program and after storing the information prepares small archive for long time storage.

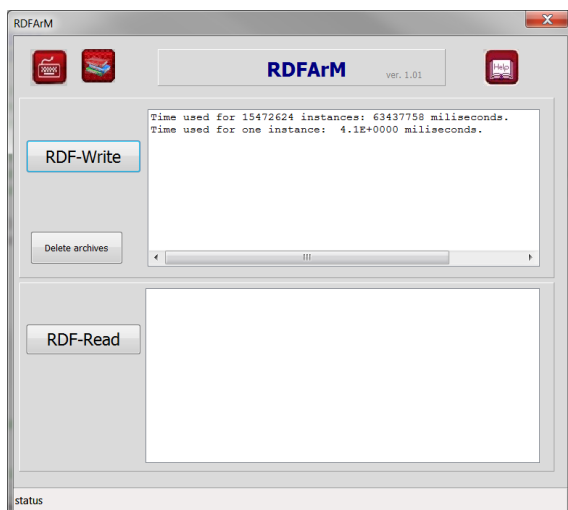


Figure 1a. Content of RDFArM RDF-Write panel with informative lines

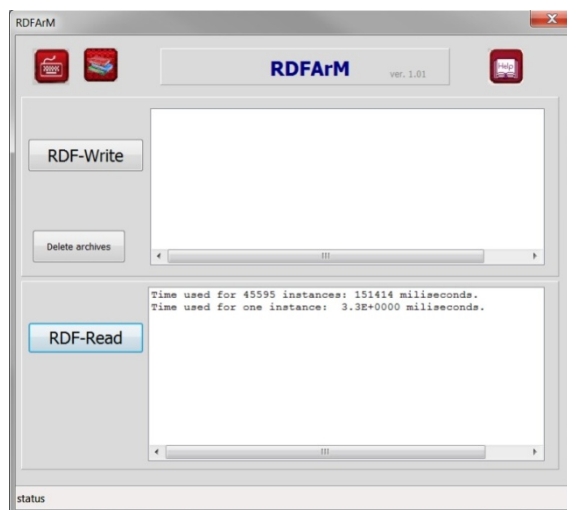


Figure 1b. Content of RDFArM RDF-Read panel with informative lines

Experiments with middle-size datasets

We have compared RDFArM with well-known RDF-stores:

- OpenLink Virtuoso Open-Source Edition 5.0.2 [Virtuoso, 2013];
- Jena SDB Beta 1 on PostgreSQL 8.2.5 and MySQL 5.0.45 [Jena, 2013];
- Sesame 2.0 [Sesame, 2012],

tested by Berlin SPARQL Bench Mark (BSBM) team and connected to it research groups [Becker, 2008; BSBMv2, 2008; BSBMv3, 2009].

We have provided experiments with *middle-size RDF-datasets*, based on selected real datasets from DBpedia [DBpedia, 2007a; DBpedia, 2007b] and artificial datasets created by BSBM Data Generator [BSBM DG, 2013; Bizer & Schultz, 2009].

The **real middle-size RDF-datasets**, we have used, consist of DBpedia's homepages and geocoordinates datasets with minor corrections [Becker, 2008]:

- *Homepages-fixed.nt* (200,036 triples; 24 MB) Based on DBpedia's homepages.nt dated 2007-08-30 [DBpedia, 2007a]. 3 URLs that included line breaks were manually corrected (fixed for DBpedia 3.0);
- *Geocoordinates-fixed.nt* (447,517 triples; 64 MB) Based on DBpedia's geocoordinates.nt dated 2007-08-30 [DBpedia, 2007b]. Decimal data type URI was corrected (DBpedia bug #1817019; resolved).

The RDF stores have different indexing behaviors: Sesame automatically indexes after each import, while SDB and Virtuoso allow for selective index activation which cause corresponded limitations or advantages. In order to make load times comparable, the data import by [Becker, 2008] had been performed as follows:

- *Homepages-fixed.nt* had been imported with indexes enabled;
- *Geocoordinates-fixed.nt* had been imported with indexes enabled.

In the case with RDFArM no parameters are needed. The data sets were loaded directly from the source files.

The **artificial middle-size RDF-datasets** are generated by BSBM Data Generator [BSBM DG, 2013] and

published in Turtle format [BSBMv1, 2008; BSBMv2, 2008; BSBMv3, 2009]. We converted it to N-triple format using “**rdf2rdf**” program developed by Enrico Minack [Minack, 2010].

We have used four BSBM datasets – 50K, 250K, 1M, and 5M. Details about these datasets are summarized in following Table 1.

Table 1. Details about used artificial middle-size RDF-datasets

Name of RDF-dataset:	50K	250K	1M	5M
Exact Total Number of Triples:	50,116	250,030	1,000,313	5,000,453
Number of Products	91	666	2,785	9,609
Number of Producers	2	14	60	199
Number of Product Features	580	2,860	4,745	3,307
Number of Product Types	13	55	151	73
Number of Vendors	2	8	34	196
Number of Offers	1,820	13,320	55,700	192,180
Number of Reviewers	116	339	1432	12,351
Number of Reviews	2,275	6,660	27,850	240,225
Total Number of Instances	4,899	23,922	92,757	458,140
File Size Turtle (unzipped)	14 MB	22 MB	86 MB	1,4 GB

In accordance with *multi-layer representation* [Ivanova et al, 2012b] the RDF-relations are assumed as layers and the RDF-subjects are assumed as paths valid for all layers. The objects are stored in the containers located by the path in the corresponded layers. Information about quantities of Subjects, Relations, and Objects in the used middle-size RDF-datasets are presented in Table 2.

Table 2. Quantities of Subjects, Relations, and Objects in used middle-size RDF-datasets

dataset	subjects (locations)	relations (layers)	objects
BSBM 50K	4900	40	50116
homepages-fixed.nt	200036	1	200036
BSBM 250K	60884	22	250030
geocoordinates-fixed.nt	152975	6	447517
BSBM 1M	92757	40	1000313
BSBM 5M	458142	55	5000453

To make experimental results comparable we have proposed special methodic and corresponded proportionality constants [Ivanova, 2014]. This way, all results from experiments are normalized to the next computer configuration:

- Processor: Intel Core2Quad Q9450@2.66GHz, CPU Launched:2008;
- Physical Memory: 8GB DDR2 667 (4 x 2GB);
- Hard Disks: 160GB (10,000 rpm) SATA2, 750GB (7,200 rpm) SATA2;
- Operating System: Ubuntu 8.04 64-bit, Kernel Linux 2.6.24-16-generic; Java Runtime: VM 1.6.0, HotSpot(TM) 64-Bit Server VM (build 10.0-b23); Separate partitions for application data (on 7,200 rpm HDD) and data bases (on 10,000 rpm HDD).

- **Loading of BSBM 50K** - The loading time' results from our experiment and from [Bizer & Schultz, 2008] are given in Table 3 and shown on Figure 2. Virtuoso has the best time. RDFArM has same loading time as Sesame and 40% better performance than Jena.

Table 3. Benchmark results for BSBM 50K

system	loading time in seconds
Sesame	3
Jena SDB	5
Virtuoso	2
RDFArM	3

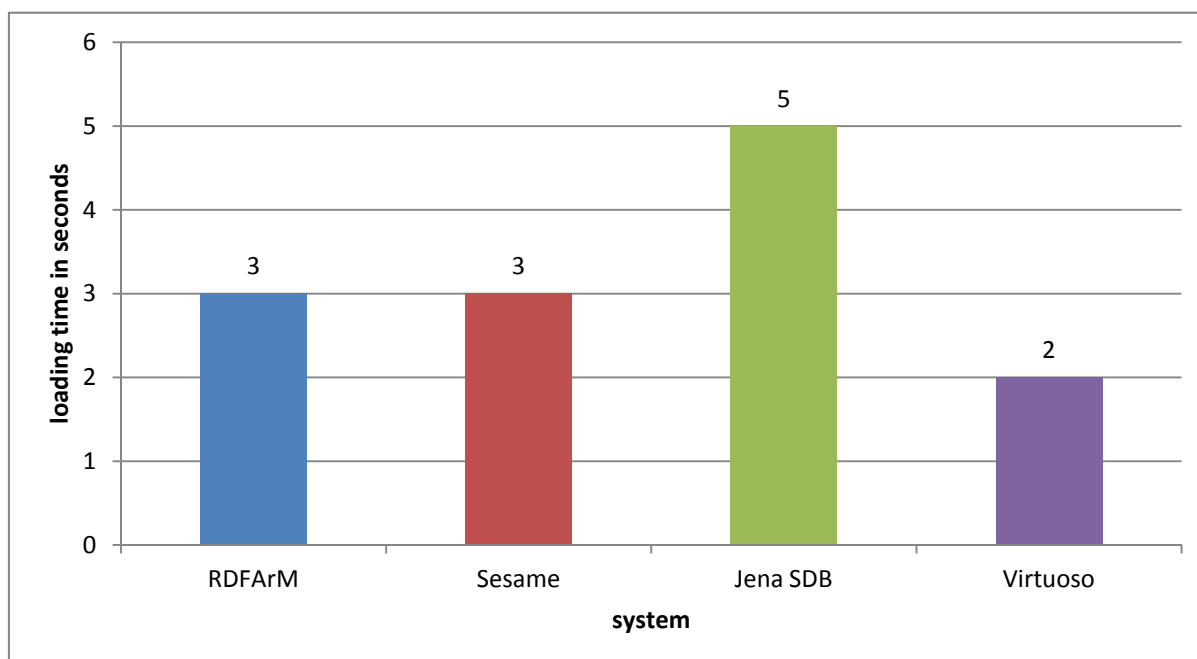


Figure 2. Benchmark results for BSBM 50K

- **Loading of homepages-fixed.nt** - The loading time' results from our experiment and from [Becker, 2008] are given in Table 4 and Figure 3. Virtuoso has the best time (about 42% better result than RDFArM). RDFArM has about 5% better time than Sesame and 36% better time than Jena (we take in account only the best result of compared system, in this case – Jena).

Table 4. Benchmark results for homepages-fixed.nt

system	loading time in seconds
Virtuoso (ogps, pogs, psog, sopg)	1327
Jena SDB MySQL Layout 2 Index	5245
Jena SDB Postgre SQL Layout 2 Index	3557
Jena SDB Postgre SQL Layout 2 Hash	9681
Sesame Native (spoc, posc)	2404
RDFArM	2272

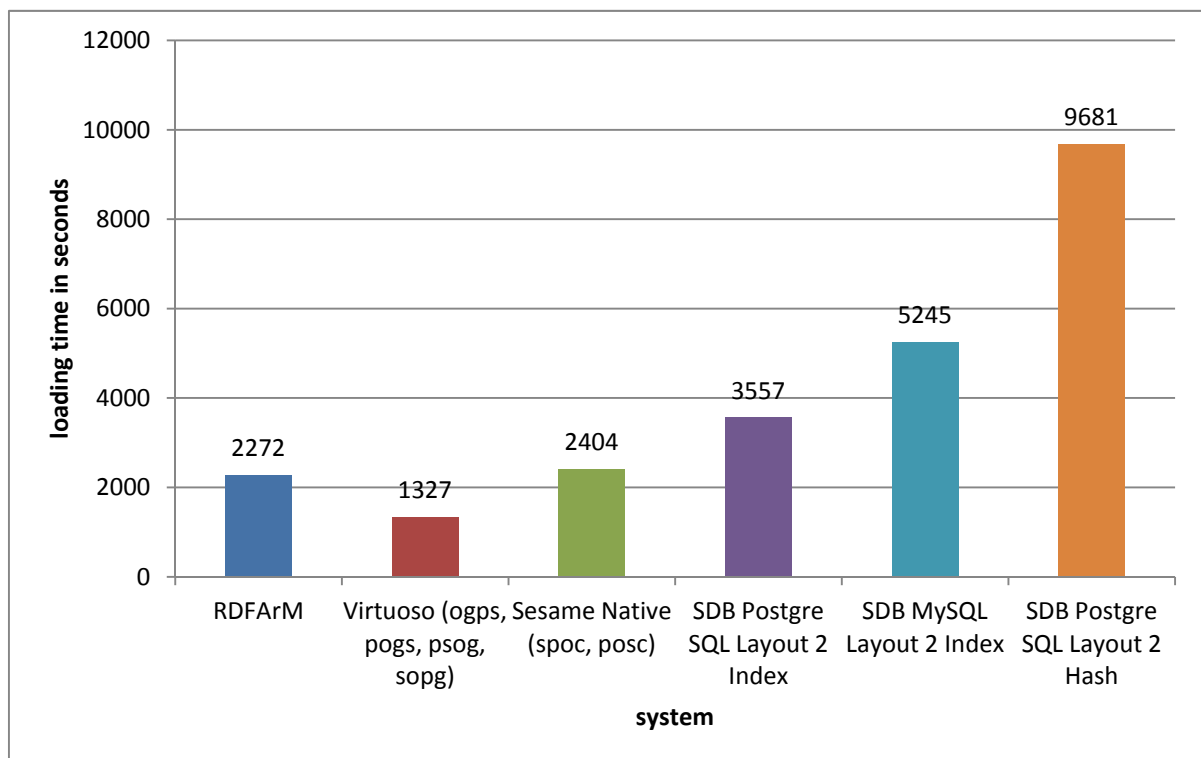


Figure 3. Benchmark results for homepages-fixed.nt

- **Loading of BSBM 250K** - The loading time' results from our experiment and from [BSBMv2, 2008] are given in Table 5 and shown on Figure 4. Virtuoso has 66% and Jena has 12% better performance than RDFArM. RDFArM has 22% better performance than Sesame.

Table 5. Benchmark results for BSBM 250K

system	loading time in seconds
Sesame	19
Jena TDB	13
Virtuoso TS	05
Virtuoso RDF views	09
Virtuoso SQL	09
RDFArM	14.79

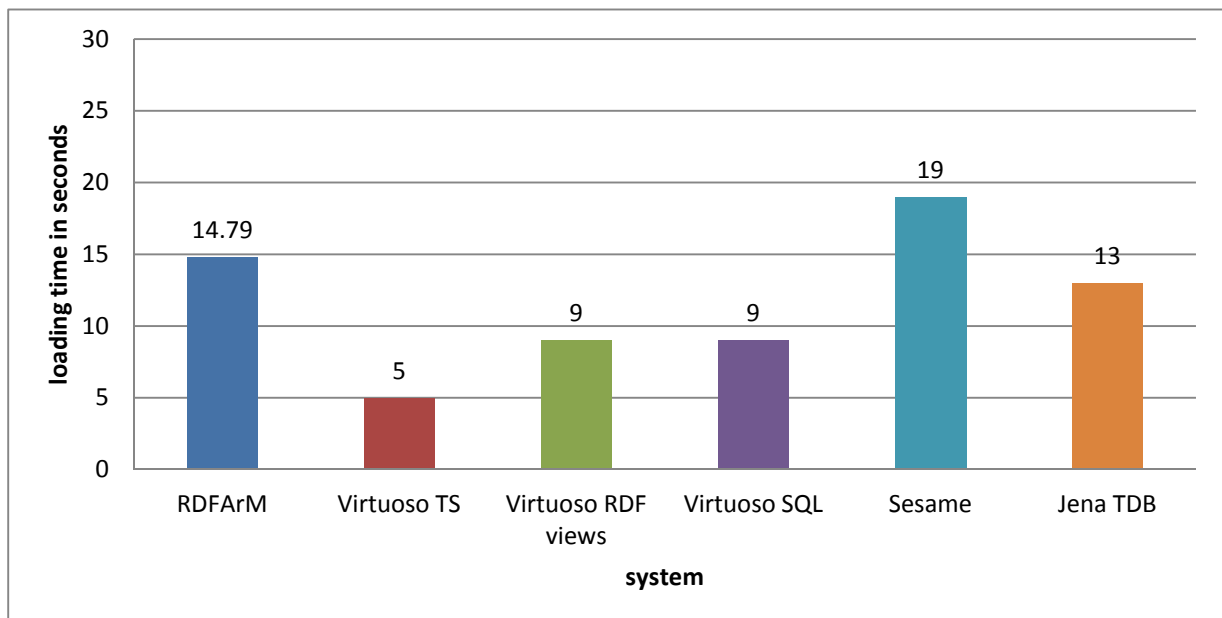


Figure 4. Benchmark results for BSBM 250K

- **Loading of geocoordinates-fixed.nt** – For this dataset we have six layers with 152975 NL-locations (containers) which contain 447517 objects, i.e. *some containers in some layers are empty*. The loading time results from our experiment and from [Becker, 2008] are given in Table 6 and Figure 5. RDFArM has the worst performance (we take the best time of Jena). Virtuoso has 64%, Sesame has 33%, and Jena has 5% better performance.

Table 6. Benchmark results for geocoordinates-fixed.nt

system	loading time in seconds
Virtuoso (ogps, pogs, psog, sopg)	1235
Jena SDB MySQL Layout 2 Index	6290
Jena SDB Postgre SQL Layout 2 Index	3305
Jena SDB Postgre SQL Layout 2 Hash	9640
Sesame Native (spoc, posc)	2341
RDFArM	3469

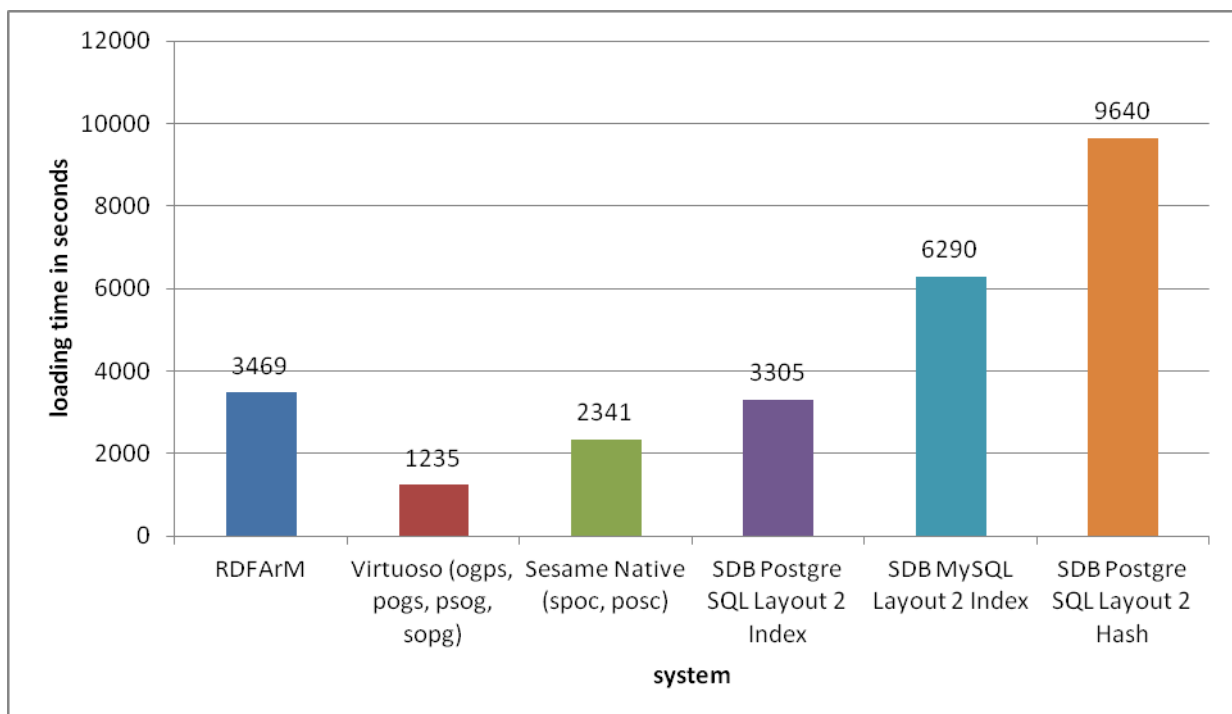


Figure 5. Benchmark results for geocoordinates-fixed.nt

- **Loading of BSBM 1M** - The loading time' results from our experiment and from [BSBMv2, 2008; BSBMv3, 2009] are given in Table 7 and shown on Figure 6. Virtuoso has 62% and Jena has 32% better performance than RDFArM. RDFArM has 67% better performance than Sesame.

Table 7. Benchmark results for BSBM 1M

system	loading time in min:sec	
	(a) [BSBMv2, 2008]	(b) [BSBMv3, 2009]
Sesame	02:59	03:33
Jena TDB	00:49	00:41
Jena SDB	02:09	-
Virtuoso TS	00:23	00:25
Virtuoso RV	00:34	00:33
Virtuoso SQL	00:34	00:33
RDFArM	01:00	01:00

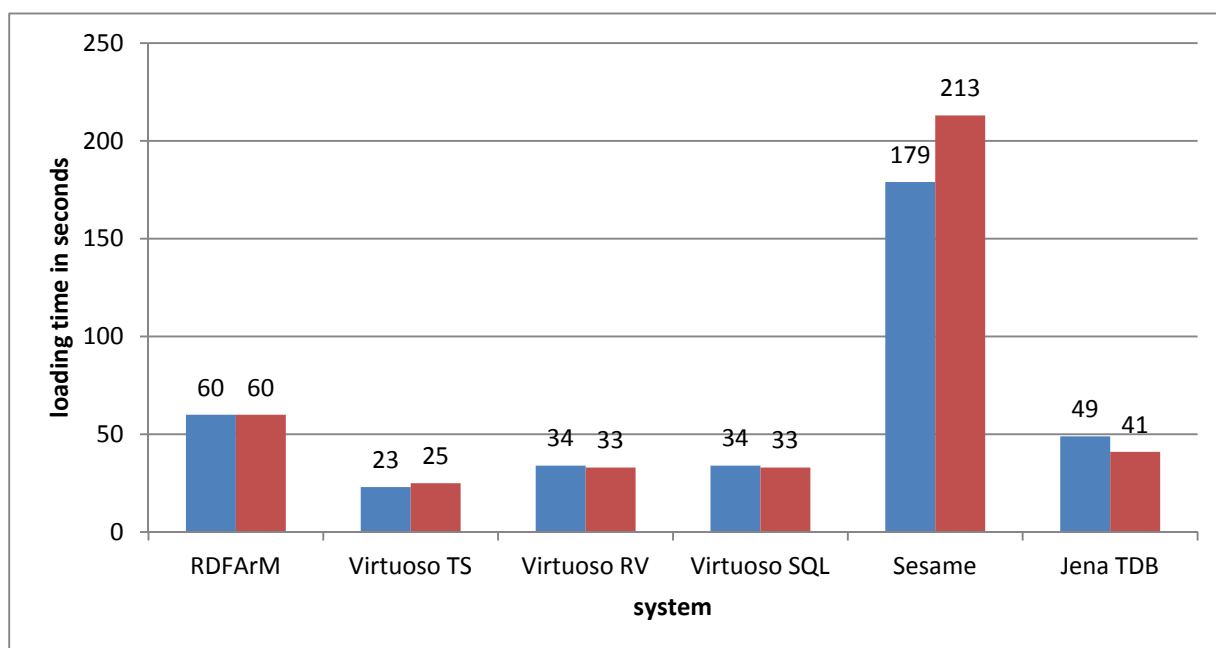


Figure 6. Benchmark results for BSBM 1M

- **Loading of BSBM 5M** - The loading time' results from our experiment and from [Bizer & Schultz, 2008] are given in Table 8 and shown on Figure 7. RDFArM has best loading time (about 85% better than Sesame, 71% than Jena, and 51% than Virtuoso).

Table 8. Benchmark results for BSBM 5M

system	loading time in seconds
Sesame	1988
Jena SDB	1053
Virtuoso	609
RDFArM	301

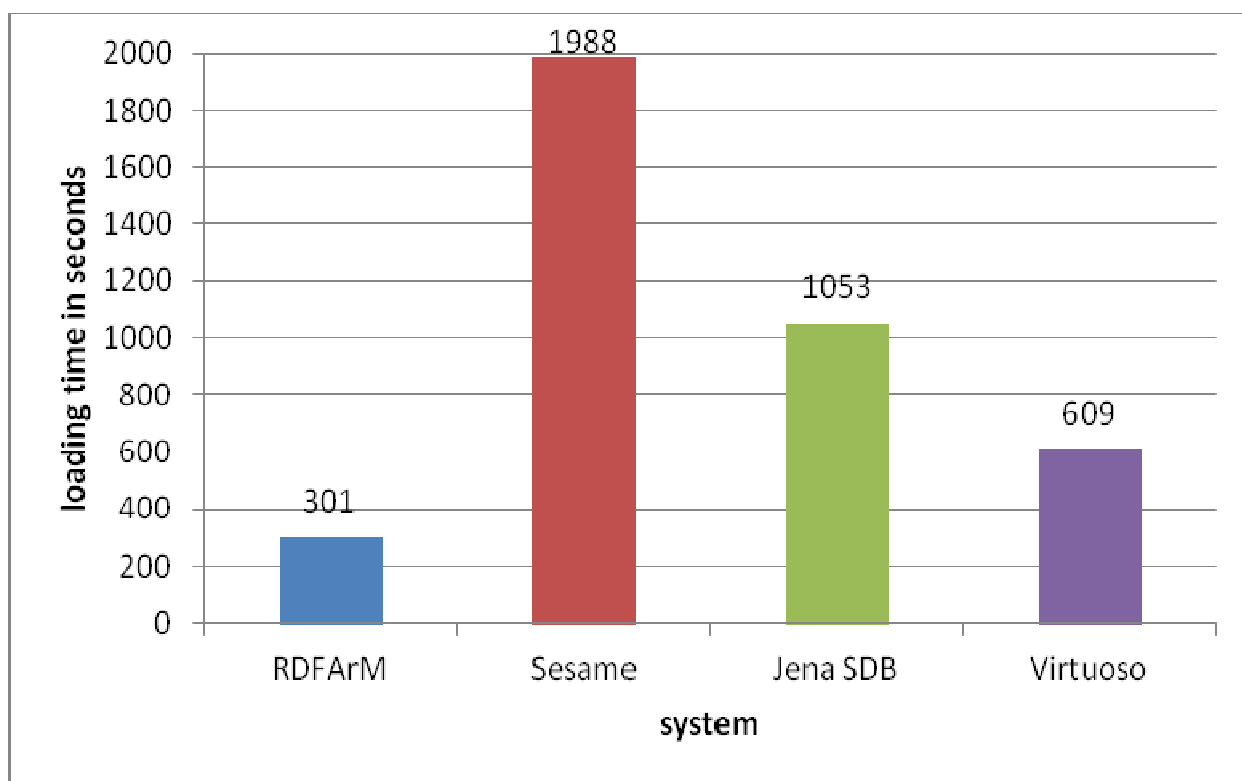


Figure 7. Benchmark results for BSBM 5M

Experiments with large datasets

We have provided experiments with **real large datasets** which were taken from DBpedia's homepages [DBpedia, 2007c] and Billion Triple Challenge (BTC) 2012 [BTC, 2012].

The real dataset is DBpedia's *infoboxes-fixed.nt* (15,472,624 triples; 2.1 GB) based on DBpedia's *infoboxes.nt* dated 2007-08-30 [DBpedia, 2007c]. 166 triples from the original set were excluded because they contained excessively large URIs (> 500 characters) that caused importing problems with Virtuoso (DBpedia bug #1871653). RDFArM has no such limitation. *infoboxes-fixed.nt* was imported with indexes initially disabled in SDB and Virtuoso. Indexes were then activated and the time required for index creation time was factored into the import time. In the case with RDFArM no parameters are needed. The datasets were loaded directly from the source file.

The RDF Stores, tested by [Becker, 2008], are:

- OpenLink Virtuoso Open-Source Edition 5.0.2 [Virtuoso, 2013];
- Jena SDB Beta 1 on PostgreSQL 8.2.5 and MySQL 5.0.45 [Jena, 2013];
- Sesame 2.0 beta 6 [Sesame, 2012].

The RDF stores feature different indexing behaviors: Sesame automatically indexes after each import, while SDB and Virtuoso allow for selective index activation.

Artificial large datasets were taken from Berlin SPARQL Bench Mark (BSBM) [Bizer & Schultz, 2009; BSBMv3, 2009; BSBMv5, 2009; BSBMv6, 2011]. Details about the benchmark artificial datasets are summarized in the following Table 9:

Table 9. Details about artificial large RDF-datasets

Number of Triples	25M	100M
Exact Total Number of Triples	25000244	100000112
Number of Products	70812	284826
Number of Producers	1422	5618
Number of Product Features	23833	47884
Number of Product Types	731	2011
Number of Vendors	722	2854
Number of Offers	1416240	5696520
Number of Reviewers	36249	146054
Number of Reviews	708120	2848260
Total Number of Instances	2258129	9034027
File Size Turtle (unzipped)	2.1 GB	8.5 GB

Information about quantities of Subjects, Relations, and Objects in the used large RDF-datasets are presented in Table 10.

Table 10. Number of Subjects, Relations, and Objects in used large RDF-datasets

dataset	subjects (locations)	relations (layers)	objects
infoboxes-fixed.nt	1354298	56338	15472624
BSBM 25M	2258132	112	25000244
BSBM 100M	9034046	341	100000112

- **Loading of infoboxes-fixed.nt** - For this dataset we have 56338 layers with 1354298 NL-locations (containers) which contain 15472624 objects, i.e. *some containers in some layers are empty*. The loading time' results from our experiment and from [Becker, 2008] are given in Table 11 and Figure 8. RDFArM has the worst loading time. Virtuoso is 95%, Sesame is 84%, and Jena is 48% better than RDFArM (we take in account only the best results of compared systems).

Table 11. Benchmark results for infoboxes-fixed.nt

system	loading time in seconds
Virtuoso (ogps, pogs, psog, sopg)	7017
Jena SDB MySQL Layout 2 Index	70851
Jena SDB Postgre SQL Layout 2 Index	73199
Jena SDB Postgre SQL Layout 2 Hash	734285
Sesame Native (spoc, posc)	21896
RDFArM	136412

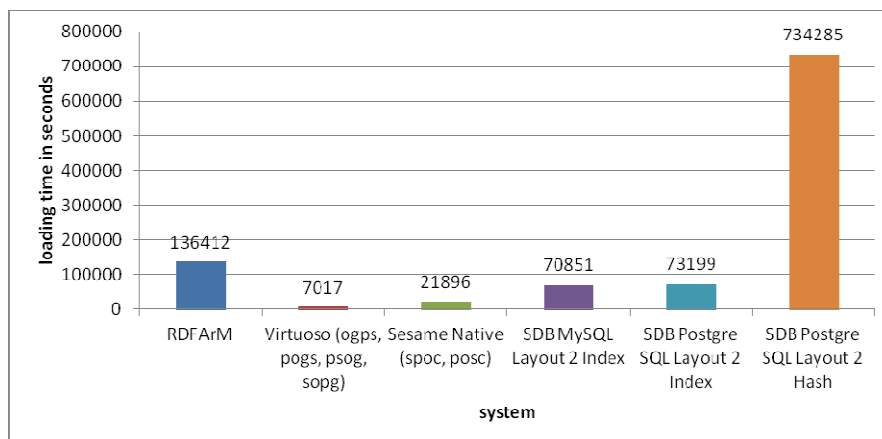


Figure 8. Benchmark results for infoboxes-fixed.nt

- **Loading of BSBM 25M** - The loading time' results from our experiment and from [Bizer & Schultz, 2009; BSBMv3, 2009] are given in Table 12 and Figure 9. Jena (with 30%) and Virtuoso (with 29%) are better than RDFArM. RDFArM has 97% better performance than Sesame.

Table 12. Benchmark results for BSBM 25M

system	loading time in seconds
Sesame	44225
Jena TDB	1013
Jena SDB	14678
Virtuoso TS	2364
Virtuoso RV	1035
Virtuoso SQL	1035
RDFArM	1453

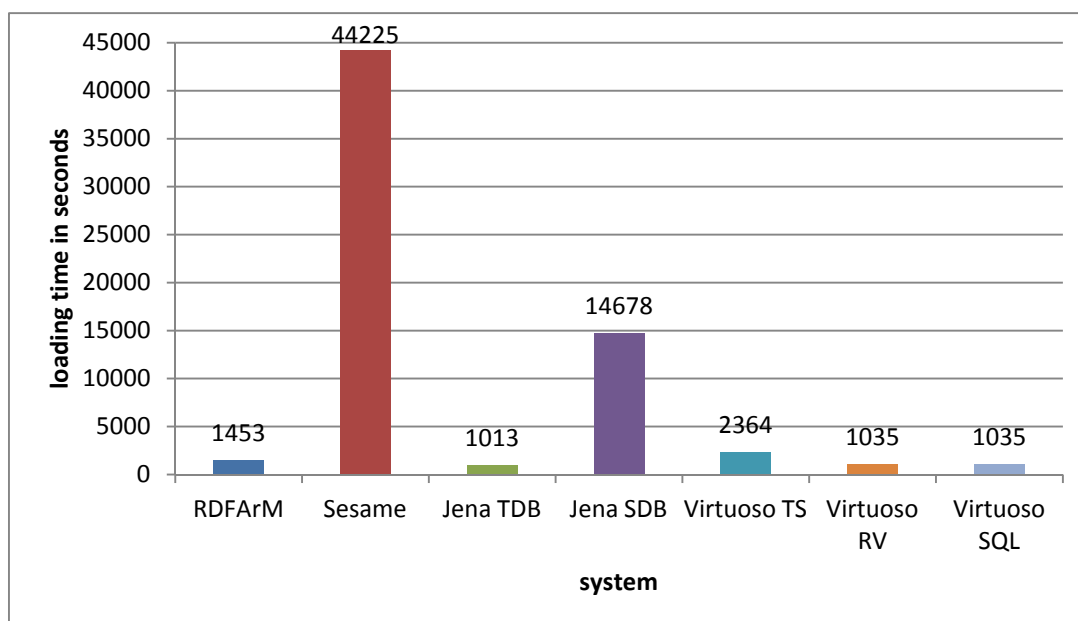


Figure 9. Benchmark results for BSBM 25M

- **Loading of BSBM 100M** - In this case we have 341 layers with 9034046 NL-locations (containers) which contain 100000112 objects, i.e. *some containers in some layers contain more than one object*. The loading time' results from our experiment and [Bizer & Schultz, 2009; BSBMv3, 2009] are given in Table 13 and Figure 10. Virtuoso is 35% better than RDFArM, and Jena is 4% better than RDFArM. RDFArM is 98% better than Sesame.

Table 13. Benchmark results for BSBM 100M

system	loading time in seconds
Sesame	282455
Jena TDB	5654
Jena SDB	139988
Virtuoso TS	28607
Virtuoso RV	3833
Virtuoso SQL	3833
RDFArM	5901

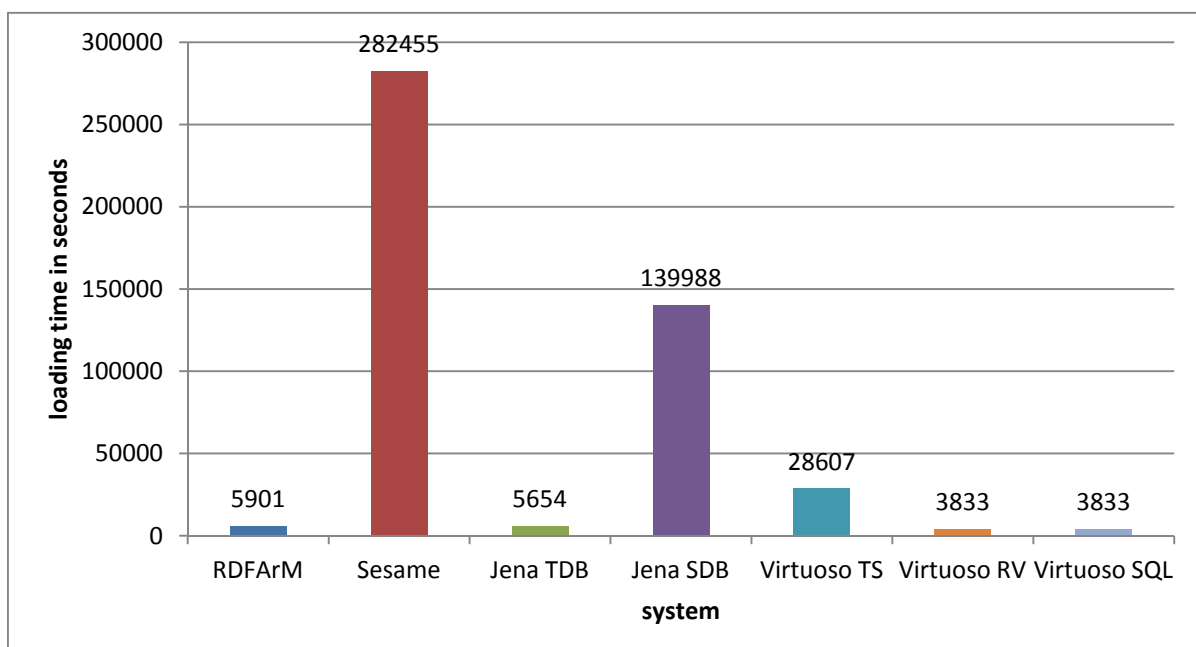


Figure 10. Benchmark results for BSBM 100M

Analysis of experiments with semi-structured datasets

In this work, the applicability of NL-addressing for middle-size and large semi-structured RDF-datasets was concerned. We have provided experiments based on selected datasets from DBpedia's homepages and Berlin SPARQL Bench Mark (BSBM) to make comparison with published benchmarks of known RDF triple stores.

We have used the Friedman test to detect statistically significant differences between the systems [Friedman, 1940]. The Friedman test is a non-parametric test, based on the ranking of the systems on each dataset. It is equivalent of the repeated-measures ANOVA [Fisher, 1973]. We have used Average Ranks ranking method, which is a simple ranking method, inspired by Friedman's statistic [Neave & Worthington, 1992]. For each dataset the systems are ordered according to the time measures and are assigned ranks accordingly. The best system receives rank 1, the second – 2, etc. If two or more systems have equal value, they receive equal rank which is mean of the virtual positions that had to receive such number of systems if they were ordered consecutively each by other.

Let n is the number of observed datasets; k is the number of systems.

Let i_j be the rank of system j on dataset i . The average rank for each system is calculated as

$$R_j = \frac{1}{n} \sum_{i=1}^k r_j^i.$$

The null-hypothesis states that if all the systems are equivalent than their ranks R_j should be equal. When null-hypothesis is rejected, we can proceed with the Nemenyi test [Nemenyi, 1963] which is used when all systems are compared to each other. The performance of two systems is significantly different if the corresponding average ranks differ by at least the critical difference

$$CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6N}}$$

where critical values q_{α} are based on the Studentized range statistic divided by $\sqrt{2}$. Some of the values of q_{α} are given in Table 14 [Demsar, 2006].

Table 14. Critical values for the two-tailed Nemenyi test

	number of systems								
	2	3	4	5	6	7	8	9	10
$q_{0.05}$	1.960	2.343	2.569	2.728	2.850	2.949	3.031	3.102	3.164
$q_{0.10}$	1.645	2.052	2.291	2.459	2.589	2.693	2.780	2.855	2.920

The results of the Nemenyi test are shown by means of critical difference diagrams.

Benchmark values from our experiments and corresponded published experimental data from BSBM team are given in Table 15. Published results do not cover all table, i.e. we have no values for some cells. To solve this problem we have taken in account only the best result for given system on concrete datasets (Table 16). Sesame had no average values for tests 10a and 10b. Because of this we did not use these tests in our comparison. They were useful to see the need of further refinement of RDFArM for big data.

The ranks of the systems for the ten tests are presented below in Table 17.

Table 15. Benchmark values for middle size datasets

system	TEST											
	1	2	3	4	5a	5b	6	7	8	9	10a	10b
RDFArM	3	2272	14.79	3469	60	60	301	136412	1453	5901	15742	31484
Sesame Native (spoc, posc)	3	2404	19	2341	179	213	1988	21896	44225	282455		
Virtuoso (ogps, pogs, psog, sopg)	2	1327		1235			609	7017			6566	14378
Virtuoso TS			05		23	25			2364	28607		
Virtuoso RDF views			09									
Virtuoso SQL			09		34	33			1035	3833		
Virtuoso RV					34	33			1035	3833		
Jena SDB	5		13		129		1053		14678	139988		
Jena TDB					49	41			1013	5654	4488	9913
Jena SDB MySQL Layout 2 Index		5245		6290				70851				
Jena SDB Postgre SQL Layout 2 Hash		3557		3305				73199				
Jena SDB Postgre SQL Layout 2 Index		9681		9640				734285				

Table 16. Chosed benchmark values for middle size datasets

system	TEST											
	1	2	3	4	5a	5b	6	7	8	9	10a	10b
RDFArM	3	2272	14.79	3469	60	60	301	136412	1453	5901	15742	31484
Sesame	3	2404	19	2341	179	213	1988	21896	44225	282455		
Virtuoso	2	1327	05	1235	23	25	609	7017	1035	3833	6566	14378
Jena	5	3557	13	3305	49	41	1053	70851	1013	5654	4488	9913

Table 17. Ranking of tested systems

system	ranks for the tests										average rank
	1	2	3	4	5a	5b	6	7	8	9	
RDFArM	2.5	2	3	4	3	3	1	4	3	3	2.85
Sesame	2.5	3	4	2	4	4	4	2	4	4	3.35
Virtuoso	1	1	1	1	1	1	2	1	2	1	1.2
Jena	4	4	2	3	2	2	3	3	1	2	2.6

All average ranks are different. The null-hypothesis is rejected and we can proceed with the Nemenyi test. Following [Demsar, 2006], we may compute the critical difference by formula:

$$CD = q_{\alpha} \sqrt{\frac{k(k+1)}{6N}}$$

where q_{α} we take as $q_{0.10} = 2.291$ (from Table 12 [Demsar, 2006; Table 5a]); k will be the number of systems compared, i.e. $k=4$; N will be the number of datasets used in benchmarks, i.e. $N=10$. This way we have:

$$CD_{0.10} = 2.291 * \sqrt{\frac{4 * 5}{6 * 10}} = 2.291 * \sqrt{\frac{20}{60}} = 2.291 * 0.577 = 1.322$$

We will use for critical difference $CD_{0.10}$ the value 1.322.

At the end, average ranks of the systems and distance to average rank of the first one are shown in Table 18.

Table 18. Average ranks of systems and distance to average rank of the first one

place	system	average rank	Distance between average rank of the system and average rank of the first one
1	Virtuoso	1.2	0
2	Jena	2.6	1.4
3	RDFArM	2.85	1.65
4	Sesame	3.35	2.15

The visualization of Nemenyi test results for tested systems is shown on Figure 11.

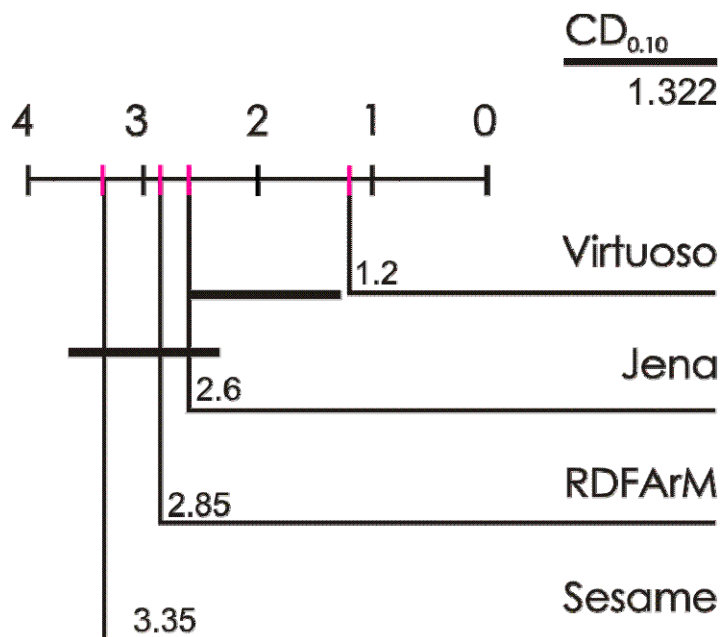


Figure 11. Visualization of Nemenyi test results

Analyzing these experiments we may conclude that RDFArM is at critical distances to Jena and Sesame. RDFArM is nearer to Jena than to Sesame. RDFArM, Jena, and Sesame are significantly different from Virtuoso.

Some recommendations to RDFArM may be given. RDF triple datasets has different characteristics depending of their origination. This causes the need to adapt NL-ArM storage engine to specifics of concrete datasets. For instance, important parameters are length of strings and quantity of repeating values of subject, relation, and object.

Conclusion

We have presented results from series of experiments which were needed to estimate the storing time of NL-addressing for middle-size and very large RDF-datasets. To make different configurations comparable, special proportionality constants for hardware and software were used.

Experiments were provided with both real and artificial datasets. Experimental results were systematized in corresponded tables. For easy reading visualization by histograms was given.

The goal of experiments for NL-storing of middle-size and large RDF-datasets were to estimate possible further development of RDFArM. What gain and loss using NL-Addressing for RDF storing?

The loss is additional memory for storing internal hash structures. But the same if no great losses we will have if we will build balanced search trees or other kind in external indexing. It is difficult to compare with other systems because such information practically is not published.

The benefit is in two main achievements:

- High speed for storing and accessing the information;
- The possibility to update and access the information immediately after storing *without recompilation* the database and rebuilding the indexes. This is very important because half or analyzed systems do not support updates.

The main conclusion is optimistic because RDFArM is at critical distances to Jena and Sesame, RDFArM is nearer to Jena than to Sesame, and, at the end, RDFArM, Jena, and Sesame are significantly different from Virtuoso.

Bibliography

- [Becker, 2008] Christian Becker, "RDF Store Benchmarks with Dbpedia", Freie Universität Berlin, 2008, <http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/> (accessed: 05.04.2013)
- [Bizer & Schultz, 2008] Christian Bizer, Andreas Schultz: Benchmarking the Performance of Storage Systems that expose SPARQL Endpoints; In: Proc. of the 4th International Workshop on Scalable Semantic Web knowledge Base Systems (SSWS2008), <http://www4.wiwiss.fu-berlin.de/bizer/pub/BizerSchulz-BerlinSPARQLBenchmark.pdf> (accessed: 31.07.2013)
- [Bizer & Schultz, 2009] Christian Bizer, Andreas Schultz, "The Berlin SPARQL Benchmark", In: International Journal on Semantic Web & Information Systems, Vol. 5, Issue 2, Pages 1-24, 2009, <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/Bizer-Schultz-Berlin-SPARQL-Benchmark-IJSWIS.pdf>; see also <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/> (accessed: 31.07.2013)
- [BSBM DG, 2013] Data Generator and Test Driver, In: Berlin SPARQL Benchmark (BSBM) - Benchmark Rules, <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/spec/BenchmarkRules/index.html#datagenerator> (accessed: 31.07.2013)
- [BSBMv1, 2008] Berlin SPARQL Benchmark Results, V1, 2008, <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/V1/results/index.html> (accessed: 31.07.2013)
- [BSBMv2, 2008] Berlin SPARQL Benchmark Results, V2 2008, <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V2/index.html> (accessed: 31.07.2013)
- [BSBMv3, 2009] Berlin SPARQL Benchmark Results, V3, 2009, <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V3/index.html> (accessed: 31.07.2013)

-
- [BSBMv5, 2009] BSBM Results (V5) for Virtuoso, Jena TDB, BigOWLIM, 2009, <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V5/index.html> (accessed: 31.07.2013)
- [BSBMv6, 2011] Berlin SPARQL Benchmark Results, V6, 2011, <http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/results/V6/index.html> (accessed: 31.07.2013)
- [BTC, 2012] Billion Triple Challenge 2012 Dataset <http://km.aifb.kit.edu/projects/btc-2012/> (accessed: 16.03.2013)
- [DBpedia, 2007a] DBpedia dataset "homepages.nt" dated 2007-08-30, <http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/homepages-fixed.nt.gz> (accessed: 31.07.2013)
- [DBpedia, 2007b] DBpedia dataset "geocoordinates.nt" dated 2007-08-30, <http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/geocoordinates-fixed.nt.gz> (accessed: 31.07.2013)
- [DBpedia, 2007c] DBpedia dataset "infoboxes.nt" dated 2007-08-30, <http://wifo5-03.informatik.uni-mannheim.de/benchmarks-200801/infoboxes-fixed.nt.gz> (accessed: 31.07.2013)
- [Demsar, 2006] Demsar J., "Statistical comparisons of classifiers over multiple data sets" J. Mach. Learn. Res., 7, 2006, pp. 1-30
- [Fisher, 1973] R. A. Fisher, "Statistical methods and scientific inference (3rd edition)", Hafner Press, New York, 1973, ISBN 978-002-844740-7
- [Friedman, 1940] Friedman, M.: "A comparison of alternative tests of significance for the problem of m rankings", Annals of Mathematical Statistics, Vol. 11, 1940, pp.86-92.
- [Ivanova et al, 2012a] Krassimira Ivanova, Vitalii Velychko, Krassimir Markov. "About NL-addressing" (К вопросу о естественно-языковой адрессации) In: V. Velychko et al (ed.), Problems of Computer in Intellectualization. ITHEA@ 2012, Kiev, Ukraine - Sofia, Bulgaria, ISBN: 978-954-16-0061 0 (printed), ISBN: 978-954-16-0062-7 (online), pp. 77-83 (in Russian).
- [Ivanova et al, 2012b] Krassimira Ivanova, Vitalii Velychko, Krassimir Markov. "Storing RDF Graphs using NL-addressing", In: G. Setlak, M. Alexandrov, K. Markov (ed.), Artificial Intelligence Methods and Techniques for Business and Engineering Applications. ITHEA@ 2012, Rzeszow, Poland; Sofia, Bulgaria, ISBN: 978-954-16-0057-3 (printed), ISBN: 978-954-16-0058-0 (online), pp. 84 – 98.
- [Ivanova et al, 2013a] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Introduction to the Natural Language Addressing", International Journal "Information Technologies & Knowledge" Vol.7, Number 2, 2013, ISSN 1313-0455 (printed), 1313-048X (online), pp. 139–146.
- [Ivanova et al, 2013b] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Introduction to Storing Graphs by NL-Addressing", International Journal "Information Theories and Applications", Vol. 20, Number 3, 2013, ISSN 1310-0513 (printed), 1313-0463 (online), pp. 263 – 284.
- [Ivanova et al, 2013c] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Storing Dictionaries and Thesauruses Using NL-Addressing", International Journal "Information Models and Analyses" Vol.2, Number 3, 2013, ISSN 1314-6416 (printed), 1314-6432(online), pp. 239 - 251.
- [Ivanova et al, 2013d] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "The Natural Language Addressing Approach", International Scientific Conference "Modern Informatics: Problems, Achievements, and Prospects of Development", devoted to the 90th anniversary of academician V. M. Glushkov. Kiev, Ukraine, 2013, ISBN 978-966-02-6928-6, pp. 214 - 215.
- [Ivanova et al, 2013e] Krassimira B. Ivanova, Koen Vanhoof, Krassimir Markov, Vitalii Velychko, "Storing Ontologies by NL-Addressing", IVth All-Russian Conference "Knowledge-Ontology-Theory" (KONT-13), Novosibirsk, Russia, 2013, ISSN 0568 661X, pp. 175 - 184.

- [Ivanova, 2013] Krassimira Ivanova, "Informational and Information models", In Proceedings of 3rd International conference "Knowledge Management and Competitive Intelligence" in the frame of 17th International Forum of Young Scientists "Radio Electronics and Youth in the XXI Century", Kharkov National University of Radio Electronics (KNURE), Kharkov, Ukraine, Vol.9, 2013, pp 6-7.
- [Ivanova, 2014] Krasimira Ivanova, "Storing Data using Natural Language Addressing", PhD Thesis, Hasselt University, Belgium, 2014
- [Jena, 2013] Apache Jena, http://jena.apache.org/about_jena/about.html (accessed: 23.03.2013)
- [Klyne & Carroll, 2004] Graham Klyne and Jeremy J. Carroll, Editors, Resource Description Framework (RDF): Concepts and Abstract Syntax, W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>. Latest version available at <http://www.w3.org/TR/rdf-concepts/> (accessed: 21.02.2013).
- [Markov, 1984] Krassimir Markov, "A Multi-domain Access Method", Proceedings of the International Conference on Computer Based Scientific Research, PLOVDIV, 1984, pp. 558 - 563.
- [Markov, 2004] Krassimir Markov, "Multi-domain information model", Int. J. Information Theories and Applications, 11/4, 2004, pp. 303 - 308
- [Markov, 2004a] Krassimir Markov, "Co-ordinate based physical organization for computer representation of information spaces", (Координатно базирана физическа организация за компютърно представяне на информационни пространства) Proceedings of the Second International Conference "Information Research, Applications and Education" i.TECH 2004, Varna, Bulgaria, Sofia, FOI-COMMERCE – 2004, стр. 163 - 172 (in Bulgarian).
- [Minack, 2010] Enrico Minack, "RDF2RDF converter", <http://www.l3s.de/~minack/rdf2rdf/> 2010, (accessed: 31.07.2013).
- [Neave & Worthington, 1992] Neave, H., Worthington, P., "Distribution Free Tests", Routledge, 1992.
- [Nemenyi, 1963] Peter Nemenyi, "Distribution-free multiple comparisons Unpublished", PhD thesis; Princeton University Princeton, NJ, 1963
- [N-Quads, 2013] N-Quads: Extending N-Triples with Context <http://sw.deri.org/2008/07/n-quads/> (accessed: 16.03.2013).
- [Sesame, 2012] Sesame, OpenRDF, <http://www.openrdf.org/index.jsp>
<http://www.openrdf.org/doc/sesame2/2.3.2/users/userguide.html#chapter-sesame2-whats-new> (accessed: 01.12.2012)
- [SPARQL, 2013] SPARQL Query Language for RDF "W3C Recommendation", 2008, <http://www.w3.org/TR/rdf-sparql-query/> (accessed: 23.03.2013).
- [Virtuoso, 2013] OpenLink Virtuoso Universal Server: Documentation <http://docs.openlinksw.com/pdf/virtdocs.pdf>, <http://virtuoso.openlinksw.com/> (accessed: 23.03.2013)

Authors' Information



Ivanova Krassimira – University of National and World Economy, Sofia, Bulgaria;
e-mail: krasy78@mail.bg

Major Fields of Scientific Research: Software Engineering, Business Informatics, Data Mining, Multidimensional multi-layer data structures in self-structured systems

MODEL OF THE MONEY CIRCULATION IN THE BULGARIAN LANDS DURING THE ANTIQUITY AND THE MIDDLE AGES

Jordan Tabov

Abstract: *This article presents a model of the money circulation in Bulgaria, based on the data about coins, excavated in Bulgaria and published in the Proceedings of the (Bulgarian) Archaeological Society (PAS) during the period 1910-1920 and in the Proceedings of the (Bulgarian) Archaeological Institute (PAI) during the period 1921-1959. We construct the Chronological Distribution of these coins and its graph using computer technology (the package MS Excel in our case). The studied material consists of over 280000 coins, which in our view is large enough to give an adequate picture in a first approximation of the chronological distribution of all coins excavated in the Bulgarian lands during the period 1910-1959. We think that it also gives a plausible picture of the monetary circulation in the eastern part of the Balkan Peninsula during the different historical periods. We discuss some considerable anomalies of the obtained chronological distribution. The results of the analysis juxtaposed with a peculiarity of the Chronological Distribution of Byzantine coins noticed by A. Kazhdan leads us to the hypothesis that there are probably certain mistakes in the accepted dating of some ancient coins.*

Keywords: *model; money circulation; coin finds; chronological distribution*

ACM Classification Keywords: *1.6 Simulation and Modeling, 1.6.3 Applications*

1. Introduction: Is it true that patterns in finds of accidentally lost coins mirror to a useful extent those of the coins in circulation at the time?

The old coins excavated in a given country are important source of information about its past. They bring information about the welfare of its population, its trade centers, trade relations with other countries, religion, the names and the titles of its rulers, etc. The well known specialist in Byzantine history A. Kazhdan wrote:

"The coin, that has been a means for trade in ancient times, in the hands of historian nowadays becomes evidence of the intensity of the trade: the abundance of coins excavated in a given old city can serve as a serious argument, which proves the existence of manufactory in this city. The chronological definiteness inherent to the numismatic material advantageously distinguishes it from many other kinds of historical sources. And at last the coin finds give a mass material, which is more or less uniformly distributed over the time and the space" ([Kazhdan, 1954]).

We ought to add to the above that the coins are a reliable "dating element" for other archaeological evidence, found together with them. Therefore the old coins are object of special interest not only for numismatic collectors, but also for archaeologists and historians.

Is it possible the quantity of the found coins, struck by a given ruler, to give us certain objective information about the monetary circulation during his reign, and thus about the intensity of the trade relations in his country and between it and other countries? In the cited article [Kazhdan, 1954] A. Kazhdan gives positive answer to this question with some reservations of little importance.

A. Gandhi [Gândilă, 2009] points to the D. M. Metcalf [Metcalf, 1958; Metcalf, 1960] as one of the pioneers in the application of statistical methods for the study of monetary circulation in the Byzantine Empire, and adds the names of V. E. Metcalf and S. Morrisson [Morrisson, 1980; Morrisson, 1981]. According to him, after 1980, there has been a real "explosion" in the use of more or less sophisticated statistical approaches.

But especially to justify the claim that the number of found old coins of a certain period of the past can serve as an indicator for the amount of monetary circulation, a special place is taken by the study of Douglas Newton [Newton, 2006]. Comparing the extant information on the quantities of coins minted in the past in a certain region with the archaeological data he established a correlation with a large coefficient of Pearson (about 0.9) between the numbers of "minted" and "found" coins [Newton, 2006]. Such a correlation gives us the opportunity to "measure" the changes in economic activity in the city, region or state by applying quantitative analysis of the found in the respective territory of old coins.

Hence to build a respective model of the money circulation one has to handle in a proper way the data about the coin finds. We do this here, applying the modern computing technology to the ideas of A. T. Fomenko about his concept of the so called "volume function" introduced in [Fomenko, 1981a] and [Fomenko, 1981b] (for more details see [Fomenko & Rachev, 1990]). We ought to stress that it is spoken about a research involving hundreds of thousands of coins; it is clear that the past attempts of the scientists to make such investigations have raised problems. To achieve our goals we used the Chronological Distribution of Coins (abbreviated CDC) described in [Tabov, 2003], which is similar to the "volume function" of A. T. Fomenko.

Variations of this concept and its construction techniques were used for building specific "historical distributions": of coins ([Tabov et al, 2003; Tabov & Panayotova, 2011]), of old manuscripts ([Tabov et al, 2004a]) and museum exhibits ([Hristova & Dobрева, 2004]) and others.

Here we use essentially the same procedure for the construction of the function of the chronological distribution and its graph.

2. Data Description

Our sources of information are the Proceedings of the (Bulgarian) Archaeological Society (PAS) during the period 1910-1920 [PAS, 1910-1920] and the Proceedings of the (Bulgarian) Archaeological Institute (PAI) during the period 1921-1959 [PAI, 1910-1920] (in a certain sense PAS has been continued by PAI). For the studied period these journals were the only scientific ones, in which regular reports of Bulgarian archaeologists on excavations and finds in Bulgaria have been published. We have extracted data about more than 280000 old coins grouped in more than 1050 finds from these journals. The high scientific verification of our data is based on this approach. There are no repeatedly reported finds, which might occur if for example additional information from newspapers or other journals is used.

We denote by CFBP-1910-1959 (Coins Found in Bulgaria and Published 1910-1959) the set of coins, excavated in Bulgaria and published in the above Proceedings during the period 1921-1959.

The rulers, who struck the respective coins, are usually given in the reports of the archaeologists. We use their time of reign for dating the coins. For about 20% of the coin finds there are given neither the rulers, nor some approximate dating – for example "there were found 70 gold Byzantine coins. I saw 4 of them". Such a piece of data is ill suited for use in this study because of uncertainty of the dating and therefore has been left out.

The data about the quantities of the coins for several coins finds in PAS and PAI is given in kilos or pots. We assumed that each kilo equals 50 coins and that a pot contains approximately 500 coins and below we will stick to this dimension.

Admittedly there were also found other ancient coins during this 50-year period (1910-1959), which have been published neither in PAS, nor in PAI. Many times some coins were found by treasure-hunters and then sold to private collectors without the civil authorities or any scientist to be informed about it. But the "split" of coins to "identified and published" and to "not published" has a random character. Furthermore the quantity of the identified and published coins is very large (more than 280000 items) and for this reason it will be plausible to suppose that our sample is representative for all the coins found in Bulgaria. Therefore we assume that our conclusions for CFBP-1910-1959 hold with a high probability for all coins found in Bulgaria.

3. Construction of the Chronological Distribution of Coins Found in Bulgaria and Published 1910-1959

We construct the Chronological Distribution (CDCFBP-1910-1959) of the set CFBP-1910-1959. The procedure of constructing this chronological distribution follows the standard procedure, described e.g. in [Tabov et al, 2004] and [Tabov et al, 2005]. The result is represented in Figure 1.

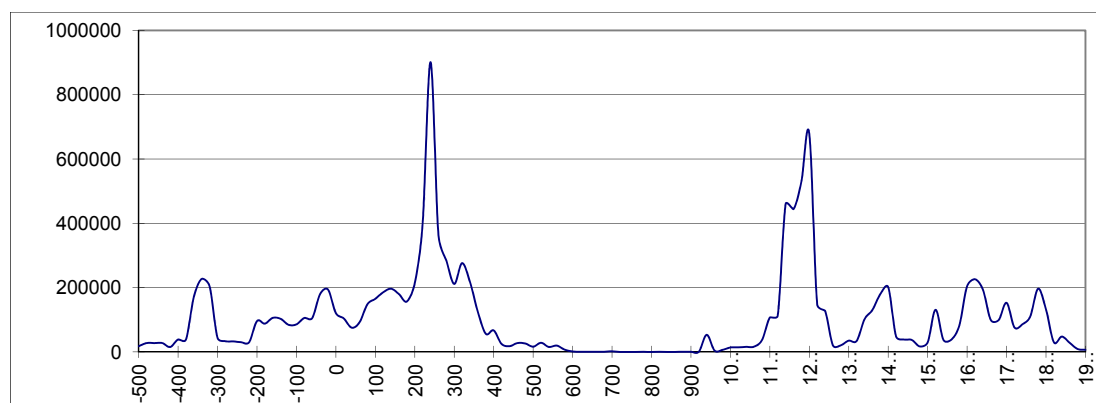


Figure 1. The graph of the Chronological Distribution of Coins Found in Bulgaria and Published 1910-1959

The above explanation gives us the opportunity to consider this graph as a model of the money circulation in the lands of modern Bulgaria.

4. Stability of the graph and representativeness of the studied data

Since the coins' excavation is a random process, it is natural to expect that after entering large enough quantity of data the shape of the graph of the CDC would become stable. This expectation proved to be true. To show this fact we consider the graphs in Figure 2 and Figure 3, which present respectively the Chronological Distribution of parts of CFBP-1910-1959, namely the Chronological Distribution of the coins found during the first 25 years of the studied period (published in [Tabov et al, 2004], and the Chronological Distribution of the coins found in 1910-1950. Comparing them to the graph in Figure 1, we conclude that the differences between the shapes of the

three graphs are very small: the high peaks and the deep downs are distributed in the same way. It might be checked that the sequence of the CDC for the first 26, 27, ..., 40, ..., 49 years "approaches" the CDCFBP-1910-1959.

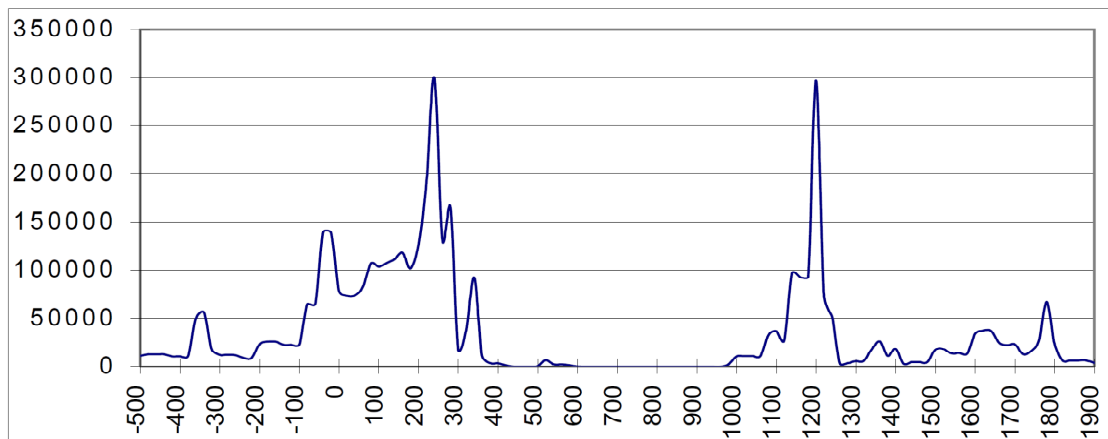


Figure 2. Chronological Distribution of Coins Found in Bulgaria and Published 1910-1934

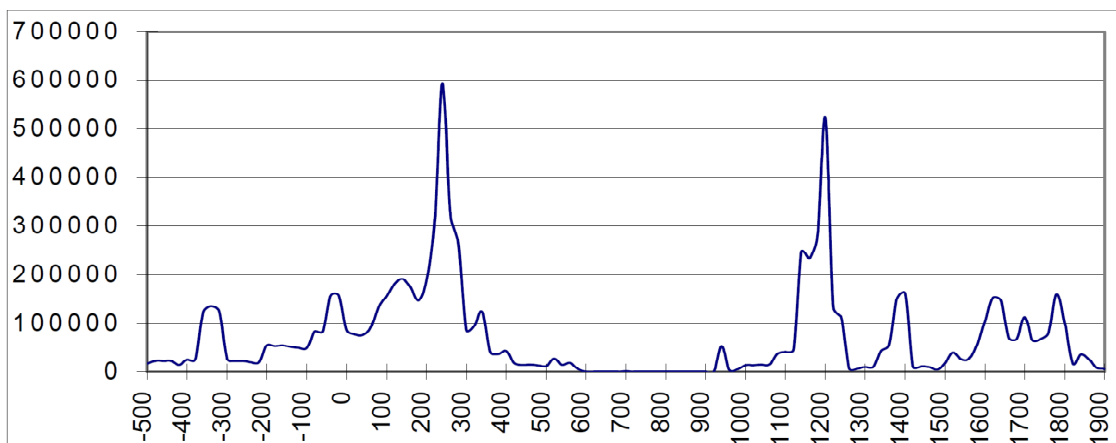


Figure 3. Chronological Distribution of Coins Found in Bulgaria and Published 1910-1950

On the basis of this observation it is natural to expect that:

1. Further data adding would not change the main properties of the graph in Figure 1, especially the distribution of the minima and maxima and the ratio between their values.
2. The CDCFBP-1910-1959 gives us an adequate picture for all excavated coins, including those, which were not published. The adding of non-published coins would not significantly change the graph in Figure 1.

5. Anomalies in the Chronological Distribution of Coins Found in Bulgaria and Published 1910-1959

We will consider four anomalies, which leap to the eye and are the most significant in our view.

1. About 60% of all found coins belong to the interval (-400, 400].
2. The amount of coins dated to the period (400, 1050] is negligible.
3. The percent of the found Turkish coins from the period of the Turkish slavery, i.e. minted after 1400, is very small, which looks very strange for the epoch of 16th -18th century.
4. The periods (200, 260] and (1180, 1240] are represented with great amounts of coins.

The first two anomalies show that the mistakes in the dating of some old coins are quite probable.

We will consider at length the second one.

6. The problem with the Byzantine coins from 8th and 9th century

In his research on Byzantine towns, published in [Kazhdan, 1954], A. Kazhdan studied every peace of data on the old coins as objective archaeological material. He used sources of information of two types:

1. Catalogues of big museums.
2. Reports on coin find (excavations and collective finds).

Kazhdan wrote:

"The catalogues of the big collections contain descriptions of hundreds of coins and provide opportunities to present more or less tentatively the minting intensity during one or another period of the Byzantine history" [Kazhdan, 1954].

In other words, without restriction in the limits of Byzantium and its history, the analyses of information of big enough and randomly chosen quantity of coins delivers opportunities to model the intensity of the monetary minting in the past.

On the basis of the data on the Byzantine coins in the collection of the British Museum, published in [Wroth, 1908], Kazhdan created the table given in Table 1:

Table 1. Data on the Byzantine coins from the collection of the British Museum, tabulated by A. Kazhdan

Period	Period in years, during which the coins were in circulation	Number of coins	Ratio of the number of coins to the years of circulation
From Anastasius I to Maurice (491-602)	111	1349	12,3
From Phocas to Constantine IV (602-685)	83	1134	13,7
From Justinian II to Michael II (685-829)	144	423	2,9

From Theophilus to Nicephorus II Phocas (829-969)	140	226	1,6
From John I Tzimisces to Nicephorus III (969-1081)	112	283	2,5
From Alexius I Comnenus to Alexius III (1081-1195)	114	349	3,0

The right column of the table defines a function, which in fact is a simple version of what we call here Chronological Distribution of Coins (CDC). We distributed the data from the above table in 20-year time units in the manner described in § 3, put it in a new MS Excel's worksheet and made the graph of what we entered in the computer. The graph is shown in Figure 4.

This graph provides us information on the Chronological Distribution of the Byzantine coins from the period 491-1195 considered by Kazhdan.

Table 1 and the graph in Figure 4 represent (almost) the same information. Comparing them we notice the advantages of the visualised presentation in Fig. 4 in comparison with the systematized numerical information in Figure 4. Such an approach and the specialized software used for visualization of the data give our study various advantages.

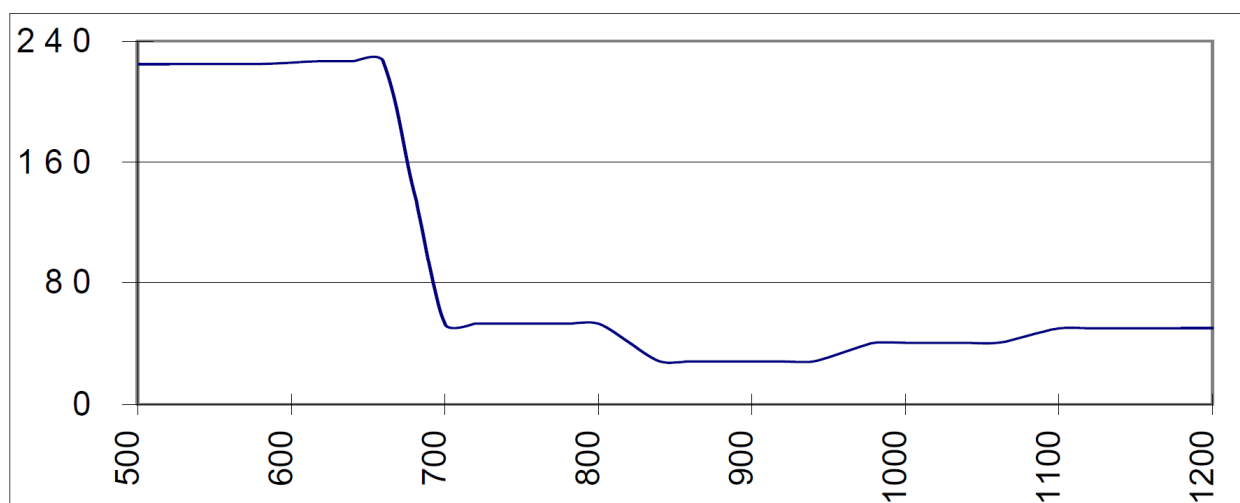
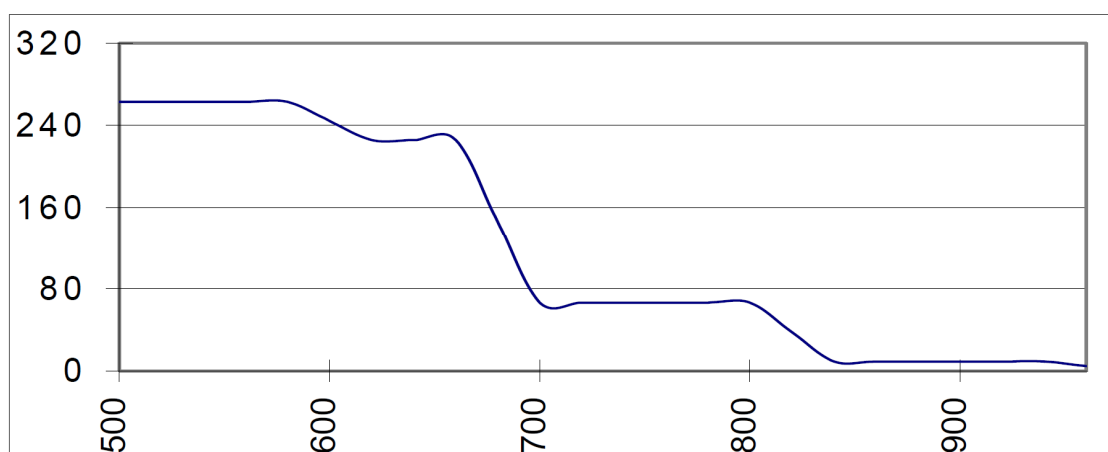


Figure 4. Chronological Distribution of Byzantine coins from the collection of the British Museum according to Kazhdan [Kazhdan, 1954]

Since the "basic intervals" in Table 1 (about 100 years and more) are longer than those used in our construction, the function in Figure 4 is rougher. Nevertheless the table gave Kazhdan the opportunity to notice a peculiarity: the period 8th-10th century is presented by lesser amounts of coins. The analogous table and graph in Table 2 and Figure 5 constructed for the Byzantine coins from the collection of count I. I. Tolstoy [Tolstoy, 1912-1914] show the same peculiarity.

Table 2. Data on the Byzantine coins from the collection of count I. I. Tolstoy, tabulated by A. P. Kazhdan

Period	Period in years, during which the coins were in circulation	Number of coins	Ratio of the number of coins to the years of circulation
From Anastasius I to Maurice (491-602)	111	1578	14,3
From Phocas to Constantine IV (602-685)	83	1129	13,6
From Justinian II to Michael II (685-829)	144	532	3,7
From Theophilus to Michael III (829-867)	38	72	1,9

**Figure 5.** Chronological Distribution of the Byzantine coins included in the collection of count I. I. Tolstoy

Considering further the generalized data of finds of Byzantine coins Kazhdan makes the following conclusion:

“Considering the data in these catalogues one may suppose that at the border between 7th and 8th century Byzantium had entered in a period of significant economical alterations, which found expression in diminishing the intensity of minting. This conclusion, however, remains hypothetical and should be seriously checked out. The matter is that the collections always have subjective and random character: in particular the more common is one or another kind of coins the milder is the collector’s interest in it. Therefore there are mostly rare coins in the collections. Moreover the coins struck by the Comnens and Palaeologs during the last centuries of the empire are of milder interest for the collectors than the earlier ones. All these factors of course deform the picture and therefore the catalogue data should be verified with the data in the publications on archaeological excavations and collective coin finds... Having in mind all these facts, all the following reasoning and calculations might have preliminary and tentative character, even though it seems to me that the mass character of the coin finds is a

certain guarantee for the relevancy of the conclusions deduced on the base of analyses of the numismatic material" [Kazhdan, 1954].

Following this logic Kazhdan directs his attention to publications on archaeological excavations and collective coin finds and uses their data to confirm his expectations that (this phrase deserves to be repeated) "the mass character of the coin finds is a certain guarantee for the relevancy of the conclusions deduced on the base of analyses of the numismatic material." Kazhdan finds out that the numismatic material found by excavations and treasure findings has the same property – negligible quantity of coins from 8th and 9th century. Details can be found by the readers in his article [Kazhdan, 1954]. It is important to notice that in his research he involves data for more than 50000 Byzantine coins; we put all them together and construct their Chronological Distribution in the interval (500, 1000]. Its graph is shown in Figure 6.

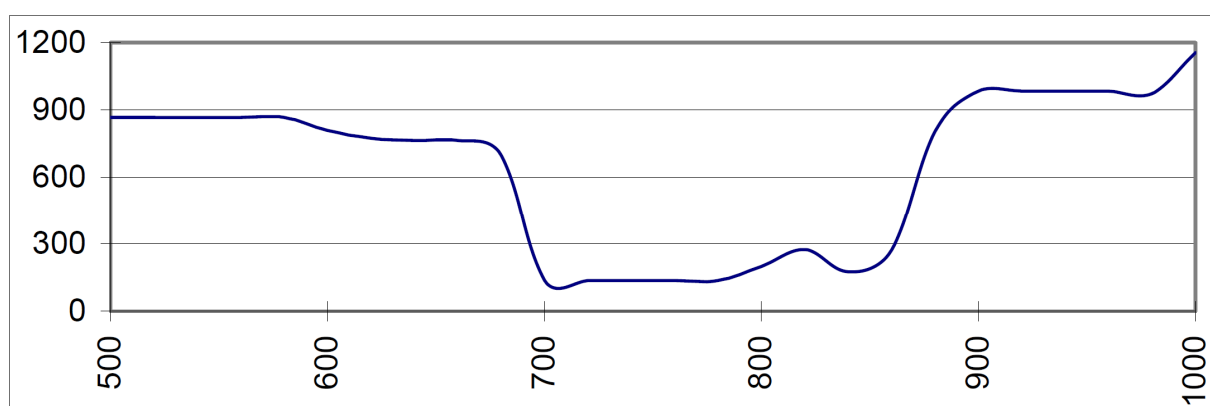


Figure 6. Chronological Distribution of all Byzantine coins considered in [Kazhdan, 1954]

Thus Kazhdan convincingly shows the existence of a “monetary crisis” in the Byzantine history. It comprises the period 8th-9th century. Similar conclusion is proved also by the data of R. Lazer [Lazer, 1980] (cited by A. Romanchuk in [Romanchuk, 2003] on Roman and early Byzantine coins (1641 coins in total) excavated in East Germany, and the results of V. Kropotkin [Kropotkin, 1961] and [Kropotkin, 1962] on finds of the same type coins on the territory of the ex-USSR and East Europe (11729 coins in total). The number (given by V. Kropotkin) of Byzantine coins found in East Europe and dated to 8th-9th century is clear enough: 2 coins from 8th century and 11 from 9th century [Kropotkin, 1962, Table 1].

This phenomenon appears to be a serious scientific problem, because the life in a city, even a small one, is impossible without coins. The majority of the citizens do not produce agricultural products like bread, milk, meat, etc and they should purchase their food, needing money for this. It's known that in some countries instead of money were used leathers; but as far as it is known from the sources in Byzantium for “money” were used only coins. Hence in Constantinople and other big cities of the empire, especially in Serdica (Sofia) and Philippopolis (Plovdiv), which are in Bulgaria today, there should have been monetary circulation. Hence coins were struck in Byzantium for sure.

But where they are? Why the archaeologists don't find them in Bulgaria?

7. Why no coins came to us from 8th – 9th century?

Kazhdan suggests an answer, which likely follows from his research: during 8th-9th century there were decays and degradation of the towns. According to him all Byzantine towns had stopped to exist or became villages and some towns like Constantinople, Thessalonica and Nicaea survived but had lost their leading economical status. Kazhdan wrote in his article [Kazhdan, 1954]:

"The material studied by us does not allow us to agree with the statement, wide-spread in the bourgeois historical literature, that Byzantine Empire has always been an urban country. Indeed with the fall of the slave-holder way of manufacture in the East Roman Empire gradually disappeared almost all slave-holder Antique pollises on Balkan Peninsula and in Asia Minor: the slave-holder polis had collapsed together with the slave-holder way of producing...Only a few towns besides Constantinople survived the fall of the slave-holder society, but they also had lost their economical functions" [Kazhdan, 1954].

The wording here outwardly reminds the attacks on the "bourgeois historical literature" but the core idea is that the towns disappeared. It is in accordance with the concept of the "dark ages" – namely in the 8th, 9th and 10th centuries.

It is worth to pay a special attention to the "exceptions from the rule", i.e. to those towns, for which Kazhdan wrote that they had preserved their urban appearance. It makes at once the impression that for none of them Kazhdan gave any data about the number of the coins found in the respective town or in the surrounding area. Considering the catalogue data of the British Museum and the collection of count Tolstoy we see that the number of coins in the "survived" towns, which came to us from 8th and 9th centuries, is too little. Hence the quantity of the found coins with a high probability is small not only in the survived towns. Therefore no objective data is visible except probably of some written sources, which might allow making such "exceptions".

All this creates the feeling that Kazhdan had tried to soften the clear results from his research and thus partially to hide the contradiction between them and the generally accepted historical picture of the epoch 8th – 10th century.

Several years after the publication of the article of Kazhdan another soviet historian – I. V. Sokolova – suggested other explanation for the lack of coins during the period mentioned in her article [Sokolova, 1959] and tried to solve the problem stated by Kazhdan.

First of all we ought to point out that she had understood very well the meaning of the coin finds for the work of the historians and especially for the decay of the towns during 8th-9th century. Her position on it is expressed in the introduction of her article:

"Coins are very rich, many-sided and interesting historical source. Designed for serving as a tool for trading they are in position to reflect to a considerable degree the state and the level of the economical development of a country during given period. Therefore the investigations of the coins themselves, of the monetary system, of the distribution and the content of the collective coin finds, random coin finds and the coin finds in cases of archaeological excavations as well create opportunities to make clear the picture of the monetary circulation, characterizing the economical centers and regions, the inner trade paths, international trade relations, etc" [Sokolova, 1959].

But this acknowledgement of the role of the coin finds does not mean at all that she was ready to accept of Kazhdan's methodology and conclusions. This is clear from her words further down:

"One has to take into consideration that the numismatic material can be valuable only when it is studied from all the sides, with correct methods and comparing it with the data from the written sources. The one-sided selective usage of it may lead to wrong conclusions" [Sokolova, 1959].

This gives us the opportunity to see that Sokolova has seen well the contradiction between the conclusions of Kazhdan and the generally accepted picture of the past and in final account with many pieces of information from the narrative sources.

Sokolova puts the specific understanding of the concept *collective coin finds* in the root of her disagreement. Since the coins are frequently found "many on the same place" in sets of some or several hundreds or even thousands (namely for such coin sets she uses the term "collective coin finds"), she starts her criticism to the work of Kazhdan with the note:

"Therefore Kazhdan is not right considering a collective coin find not as an overall memorial from the epoch of its burial but as a randomly accumulated material. Decomposing the finds in parts of coins from same centuries and then summarizing the respective parts of all the collective coin finds, A. P. Kazhdan compares the results thus obtained considering them as a measure of the intensity of coins' minting or their distribution in foreign countries during one or another historical period. At that he doesn't take into account the size of the minting area and the metal of the coins, which form the collective coin find" [Sokolova, 1959].

Kazhdan really has ignored the material of the coins and also a number of insignificant other things, which influence on the studied problem is relatively small in first approximation. For example for the minting area size Sokolova is not right, because Kazhdan considers a group of towns of the Byzantine Empire, which is representative for the latter and as such it gives completely adequate view on the economic life of the whole country.

But her main objection to the conclusions of Kazhdan is based on the following:

"Influxes and refluxes in burying collective coin finds in one or another region depend directly on how much the population of these regions was threatened by war." [Sokolova, 1959]

With other words she states that the people in the past had buried their money in the ground because of fear of enemy invasions. But then we must ask: what the people had done to hide money from their relatives – wife, children, and grand children? And from friends and enemies, from thieves and burglars? Put their money in Swiss banks? This concept of Sokolova is unacceptable. In the most cases people hide money not because of war danger but for other reasons.

But even if we suppose that Sokolova is right, then taking into account the frequent Bulgarian raids in Byzantium during the period 8th – 9th century, it is natural to expect that because of warlike danger in today's South Bulgaria lots of coin treasures have been buried there. However, as we have already seen, the archaeological data shows just the opposite.

So, the objections of Sokolova against the conclusions of Kazhdan are unacceptable. First because they are based on a wrong concept of the "peace" on the Balkans during the period 8th – 9th century, and second because they are not consistent with the logic of the people's behaviour when they try to hide their money from other's eyes and hands.

Recently in his article [Romanchuk, 2003] A. Romanchuk took a medium position between the described opinions of Kazhdan on one hand and Sokolova on the other about how far the quantities of the found old coins reflect the state of the coins' minting during different periods of time. He took a note of the fact that some coins were

probably in circulation during periods of time, considerably surpassing the period of reign of the ruler, who struck the respective coins. He advances arguments noticed by V. Kropotkin [Kropotkin, 1961] that in the archaeological finds of Roman coins from 3rd century there are always coins from 1st and 2nd century as well. If this is true, it points in our opinion to the fact that there are systematic errors in coins' dating.

8. The third and the fourth anomalies in CDCFBP-1910-1959

We will consider another two anomalies, pointed in § 5.

Let compare the quantities of the coins dated to the intervals (220, 320] and (1480, 1580]. From the modern historical point of view of the picture of the past there should be expected that the coins from the second interval are much more than these from the first one. We point two arguments pro this thesis:

1. The Roman reign on the Balkans was in crisis during the first half of the first period. The Barbarian raids lead to stagnation in many aspects of life over the territories of today's Bulgaria, especially in the field of economics. During (1480, 1580] the territories of today's Bulgaria were parts of the Ottoman Empire and undergo economical growth. The stories of travellers, who had crossed the Balkan Peninsula on their way to Constantinople, mention caravanserais and dozens and sometimes even hundreds of shops in each town (see for example the recent monograph of L. Klusakova ([Klusakova, 2002]), i.e. obvious signs for brisk trade.
2. The second interval is later and the percent of the extant coins should be greater (at least not lesser) than the same percent for the first interval. But from the graph in Fig. 2 it follows that the most intensive monetary circulation through; Bulgarian lands was namely in the period (220, 320]. It means ten times more coins in circulation than those in the period (1480, 1580]. Such levels, so high, are unrealistic, having in mind that a part of Balkan population was slaves during the 3rd century and they didn't partake in the monetary circulation.

9. Conclusions

The analysis of the graph of the CDCFBP-1910-1959 (chronological distribution of the coins found in Bulgaria and published during the period 1910-1959) from the point of view of today's historical science on the picture of the past points a number of contradictions. Their explanations have to be searched in different directions and the most natural of them is to propose that much of the discrepancies are due to certain incorrect dating of coins. ***The scale of the anomalies gives us merits to suspect that there are mistakes in the dating of some old coins.*** For the justification of this hypothesis further investigations are necessary.

Bibliography

- [Fomenko & Rachev, 1990] Fomenko, A. and S. Rachev. Volume Functions of Historical Texts and The Amplitude Correlation Principle. Computers and the Humanities, 11 (1990), 187-206.
- [Fomenko, 1981a] Fomenko A. T. New experimental-statistic methods of dating events and applications in the Global chronology of the Ancient world. Preprint Gos. Com. Telev. Rad. 3672, № B07201 (9/1981), Moscow. (In Russian)
- [Fomenko, 1981b] Fomenko A.T. Informational functions and related statistical laws. Theses of papers presented at the 3-d International Conference in Probability Theory and Math. Statistics, Vilnius, Vol. 2, 211-212. (In Russian)

-
- [Gândilă, 2009] Gândilă, A. Early Byzantine Coin Circulation in the Eastern Provinces: A Comparative Statistical Approach. *AJN Second Series* 21 (2009) 151–226.
- [Hristova & Dobрева, 2004] S. Hristova, M. Dobрева. Some observations on the chronological distribution of mediaeval manuscripts and church items preserved in Bulgaria. In: *Mathematics and Education in Mathematics. Proc. Of the Thirty Third Spring Conference of the Union of Bulgarian Mathematicians, Borovets, April 1-4, 2004*, 214-217.
- [Kazhdan, 1954] Kazhdan A. P. Byzantine towns during 7th-11th century. *Soviet Archaeology*, XXI (1954), 164-188. (In Russian)
- [Klusakova, 2002] Klusakova L. *The Road to Constantinople. Sixteenth-Century Ottoman Towns through Christian Eyes*, ISV Publishers, Prague, 2002.
- [Kropotkin, 1961] Kropotkin V. V. Finds of Roman Coins in the USSR. *Soviet Archaeological Institute*, Issue G4-4. (In Russian)
- [Kropotkin, 1962] Kropotkin V. V. Finds of Roman Coins in the USSR. *Soviet Archaeological Institute*, Issue E4-4. (In Russian)
- [Lazer, 1980] Lazer R. *Die Roemischer und Fruehbyzantinischen Fundmuenzen aus dem Gebiet der DDR*, Berlin, 1980.
- [Metcalf, 1958] Metcalf, D. M. Statistische Analyse bei der Auswertung von Munzfundmaterialien. *Jahrbuch fur Numismatik und Geldgeschichte* 9 (1958), 187–196.
- [Metcalf, 1960] Metcalf, D.M. The currency of Byzantine Sirmia and Slavonia. *Hamburger Beitrage zur Numismatik* 14 (1960), 429–444.
- [Morrisson, 1980] Morrisson, C. Les monnaies. In: J.-P. Sodini, G. Tate, B. Bavant, S. Bavant, J.-L. Biscop, D. Orssaud, and C. Morrisson, *Dehes (Aurie du nord) campagnes I–III (1976–1978), recherches sur l'habitat rural*. Paris: Librairie Orientaliste Paul Geuthner, 1980, 267–287.
- [Morrisson, 1981] Morrisson, C. Estimation du volume des emissions de Solidi de Tibere et Maurice a Carthage (578–602). In: C. Carcassone and T. Hackens, eds. *Statistique et numismatique: table ronde organisee par le Centre de mathematique sociale de l'Ecole des hautes etudes en sciences sociaux de Paris et le Seminaire de Numismatique Marcel Hoc de l'Universite. Catholique de Louvain, Paris, 17–19 sept. 1979*, Strasbourg, Conseil de l'Europe, Assemblée parlementaire, 1981, 267–284.
- [Newton, 2006] Newton, D. P. Found Coins as Indicators of Coins in Circulation: Testing Some Assumptions. *European Journal of Archaeology* 2006; 9; 211-227. <http://eja.sagepub.com>
- [PAI, 1910-1920] *Proceedings of the (Bulgarian) Archaeological Society*. Vol. 1(1910), Vol. 2(1911), Vol. 3(1912-1913), Vol. 4(1914), Vol. 5 and 6 (1915-1918), Vol. 7(1919-1920).
- [PAS, 1910-1920] *Proceedings of the (Bulgarian) Archaeological Institute*. Vol. 1(1921-1922), Vol. 2(1923-1924), Vol. 3(1925), Vol. 4(1926-1927), Vol. 5(1928-1929), Vol. 6(1930-1931), Vol. 7(1932-1933), Vol. 8(1934), Vol. 9(1935), Vol. 10(1936), Vol. 11(1937), Vol. 12(1938) Vol. 13(1939), Vol. 14(1940-1943), Vol. 15(1944-1946), Vol. 16(1947-1949), Vol. 17(1950), Vol. 18(1951-1952), Vol. 19(1953-1955), Vol. 20(1956), Vol. 21(1957), Vol. 22(1958-1959).
- [Romanchuk, 2003] Romanchuk. A. I. On the problem of the usage of numismatic data in studying the Byzantine town during the period of the "dark ages". In: *The world of Alexander Kazhdan, Alateya, Sanct-Petersburg, 2003*, 132-137. (In Russian)
- [Sokolova, 1959] Sokolova I. V. Finds of Byzantine Coins as a source of the Byzantine History in 7th -11th century. *Viz. Vremennik*, XV (1959), 50-63. (In Russian)

-
- [Tabov & Panayotova, 2011] Tabov, J., G. Panayotova. Model of the dynamics of the coin circulation in medieval Pliska. In: Krassimir Markov, Vitalii Velychko (editors). Applicable Information Models, No. 22. ITHEA, Sofia, 2011, 149-154. (In Russian) http://www.foibg.com/ibs_isc/ibs-22/ibs-22-p21.pdf
- [Tabov et al, 2003] Tabov, J., Vasilev K. and Velchev A. A mathematical model of monetary circulation in Medieval Bulgaria. Storiadelmondo. 2003: <http://www.storiadelmondo.com/14/tabov.monetary.pdf>
- [Tabov et al, 2004] Tabov J., Vassilev K. and Velchev A. Mathematical modelling of monetary minting in mediaeval Bulgaria. Review of the National Center for Digitization, Belgrade, 4 (2004), 99-104.
- [Tabov et al, 2004a] Tabov, J., A. Velchev, M. Dobрева, K. Sotirova. Chronological distribution of the Bulgarian mediaeval manuscripts preserved in Bulgaria. In: Mathematics and Education in Mathematics. Proc. Of the Thirty Third Spring Conference of the Union of Bulgarian Mathematicians, Borovets, April 1-4, 2004, 257-261.
- [Tabov et al, 2005] Tabov J., Vassilev K. and Velchev A. (2005) Chronological Distribution Of The Coin Finds In Bulgaria Reported in the Scientific Literature For a Quarter Century (1910-1934), Review of the National Center for Digitization, Belgrade, 6, 103-107.
- [Tabov, 2003] Tabov, J. Chronological Distribution of Information in Historical Texts. Computers and the Humanities, 24 (2003), 235-240.
- [Tolstoy, 1912-1914] Tolstoy I. I. Byzantine coins. Sanct-Petersburg, (1912-1914). (In Russian)
- [Wroth, 1908] Wroth W. Catalogue of the imperial Byzantine coins in the British Muzeum, London, 1908.

Information for the Author



Jordan Tabov – Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences, Acad. G.Bonchev Str. Block 8, 1113 Sofia, Bulgaria; e-mail: tabov@math.bas.bg

Major Fields of Scientific Research: Applications of mathematics and informatics in the humanities, Didactics of mathematics and informatics

КОМПЛЕКСНЫЕ ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ИНЖЕНЕРИИ ОНТОЛОГИЙ

Виталий Величко, Кирилл Малахов, Виталий Семенов, Александр Стрижак

Аннотация: В статье представлен обзор актуальных специализированных инструментальных средств инженерии онтологий, а также средств аннотирования текстов на основе онтологий. Рассмотрены основные функции и возможности данных инструментальных средств, их достоинства и недостатки. Описываются основные компоненты системы формирования онтологий на основе семантического анализа текстовых массивов. Дается системный сравнительный анализ средств инженерии онтологий.

Ключевые слова: онтология, инженерия знаний, проектирование онтологии предметной области, инструментальный комплекс онтологического назначения.

ACM Classification Keywords: I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods, H. Information Systems – H.2 DATABASE MANAGEMENT – H.2.4 Systems

Введение

Методология проектирования онтологии предметной области (ПрО) [Палагин, 2012] основывается на конструктивном использовании следующих категорий – множества концептов, отношений, функций интерпретации и аксиом. Построение указанных множеств является трудоёмким процессом, как по времени, так и по количеству вовлечённых в процесс проектирования высококвалифицированных специалистов. Особенно трудоёмко ручное проектирование онтологий, которое мало чем отличается от тематического проектирования экспертных систем.

Понимание важности решения проблемы создания эффективных инструментальных средств поддержки процессов проектирования онтологии заданной ПрО, пришло практически одновременно с осознанием парадигмы компьютерных онтологий. В настоящее время известно более ста инструментальных программных систем [Гаврилова, 2000; Палагин, 2009], но только ограниченное количество комплексных программных систем включают редактор онтологических структур, автоматизированное построение онтологий ПрО, средства поверхностного семантического анализа текстовых документов, используемых для построения онтологии.

Основные характеристики комплексных программных систем проектирования онтологий

К основным характеристикам комплексных программных систем онтологического инжиниринга можно отнести: поддерживаемые формализмы и форматы представления онтологий, архитектура программного обеспечения, интерфейс пользователя, функционал редактора онтологии, средства хранения онтологий, доступность, дополнительные возможности. Ниже рассмотрим более подробно некоторые из приведенных характеристик.

Поддерживаемые формализмы и форматы представления онтологий

Под формализмом понимается некоторая формальная теория [Клини, 1957], лежащая в основе способа представления онтологических знаний (логика предикатов, фреймовые модели, дескриптивная логика, концептуальные графы и др.). Выбранный формализм существенно влияет на организацию внутренних (компьютерных) структур данных и может определять их формат представления.

Формат представления онтологий задаёт вид их хранения в библиотеке, способ передачи онтологических описаний другим потребителям и метод обработки ее концептов. В качестве форматов онтологических описаний разработаны определенные языки представления онтологий, наиболее известными из которых являются OWL, RDFS, KIF.

Некоторые из известных редакторов онтологий поддерживают работу с несколькими формализмами представления, однако следует учитывать тот факт, что обычно конкретный формализм является предпочтительным для конкретного редактора [Палагин, 2009].

Функциональность редактора онтологии

Функциональность редактора онтологии является одной из самых важных характеристик, под которой понимается множество предоставляемых пользователю сервисов работы с онтологическими структурами.

Базовый набор функций редактора онтологии обычно обеспечивает:

- работу с одним или несколькими онтологическими описаниями (проектами) одновременно;
- графический интерфейс пользователя;
- редактирование онтологии (создание, редактирование, удаление концептов, отношений, аксиом и прочих структурных элементов онтологии);
- инкапсулирование онтологий в среду информационных систем.

Дополнительные возможности

К дополнительным возможностям относят поддержку языка запросов, анализ целостности, использование механизма логического вывода, поддержку удалённого доступа через Интернет, документирование.

Известны три группы инструментальных средств (ИнС) онтологического инжиниринга [Овдей, 2006]. К первой группе относят инструменты создания онтологий, которые предполагают поддержку совместной разработки и просмотра, создание онтологии в соответствии с заданной (произвольной) методологией, поддержку рассуждений.

Ко второй группе относят инструменты объединения, отображения и выравнивания онтологий. Объединение предполагает нахождение сходств и различий между исходными онтологиями и создание результирующей онтологии, которая содержит элементы исходных онтологий. Для этого ИнС автоматически определяют соответствия между концептами или обеспечивают графическую среду, в которой пользователь сам находит эти соответствия. Процедура отображения заключается в нахождении семантических связей между концептами различных онтологий. Процедура выравнивания онтологий устанавливает различные виды соответствия между двумя онтологиями, информация о которых сохраняется для дальнейшего использования в приложениях пользователя [Noy, 1999].

К третьей группе относят инструменты для аннотирования Web-ресурсов на основе онтологий.

Содержательный обзор известных инструментов инженерии онтологий, в котором рассмотрены основные функции и возможности ИнС, их достоинства, недостатки, сравнительный анализ и описание известных доступных онторедакторов, также приведен в [Овдей, 2006; Noy, 1999; Calvanese, 2007; Филатов, 2007].

Общими недостатками большинства известных инструментальных средств являются:

- отсутствие процедур автоматического (автоматизированного) формирования компонент онтологии;
- англоязычный интерфейс с пользователем, в котором (для большинства ИнС) не предусмотрено присвоение имён компонентам онтологии на русском или украинском языке;
- структуризация концептов выполняется только по одному типу отношений;
- для большинства общедоступных ИнС не предусмотрена работа с большими по объёму онтологиями (например, для OntoEditFree – до 50 концептов);
- большинство инструментов хранит свои онтологии в текстовых файлах, что ограничивает скорость доступа к онтологиям;
- задекларированные функциональные возможности для общедоступных инструментов зачастую так и остаются нереализованными;
- недостаток информации для пользователей в инструкциях.

Рассмотрим более подробно некоторые инструментальные программные системы, предназначенные для построения онтологий и их использования для решения задач. Универсальные и специализированные оболочки программных систем являются средством, упрощающим процесс создания интеллектуальной системы [Артемяева, 2008]. Универсальные оболочки основаны на использовании некоторого универсального языка представления знаний. В специализированных оболочках при представлении знаний используется специфичная для предметной области схема, определяемая онтологией той области, для которой создается оболочка, что позволяет создавать базу знаний эксперту предметной области без участия посредника, которым является инженер знаний. В сложно-структурированных предметных областях, связанных с наукой, могут изменяться не только знания, но и онтологии, и, как следствие, множество классов решаемых задач. Описание особенностей специализированных оболочек интеллектуальных систем для сложно-структурированных предметных областей приведено на основе статьи «Интеллектуальная система, основанная на многоуровневой онтологии химии» [Артемяева, 2008].

Специализированная оболочка интеллектуальной системы для сложно-структурированных предметных областей

Информационными компонентами специализированной оболочки для сложно-структурированной ПрО являются многоуровневая модульная онтология и модульная база знаний. Создание и редактирование информационных компонент осуществляется многоуровневым редактором онтологий и редактором знаний, разработка которых основывается на онтологии уровня *п*.

Редакторы многоуровневых онтологий и знаний должны позволять создание и редактирование модульных онтологий и знаний, а также обеспечивать возможность повторного использования модулей

при создании онтологий и знаний новых разделов и подразделов области, причем процесс создания и редактирования модуля онтологии уровня $i-1$ должен управляться онтологией уровня i , а процесс создания и редактирования модуля знаний – онтологией уровня 2.

Редактор онтологии должен обеспечивать возможность выбора того из существующих модулей онтологии уровня i , который управляет процессом редактирования создаваемого модуля. Аналогично при редактировании модуля знаний должна обеспечиваться возможность выбора «управляющего» модуля онтологии уровня 2.

Редакторы онтологии и знаний должны обеспечивать возможность задания структурированной и неструктурированной части онтологии, а также структурированной и неструктурированной части знаний, т. е. программным компонентом этих редакторов должен быть специализированный редактор утверждений, позволяющий вводить онтологические соглашения и законы предметной области.

Редактор знаний должен обеспечивать возможность ввода/вывода значений нестандартных величин при редактировании знаний. Для значений нестандартных величин в предметной области может существовать способ их графического представления. Например, для химии [Артемьева, 2008] графически может быть задана краткая структурная формула или структурная формула химического соединения. Поэтому редактор знаний должен обеспечивать возможность использования принятого в предметной области графического способа представления значений нестандартных величин при создании и редактировании знаний. Величина, которой принадлежит значение некоторого свойства, задается онтологией уровня 2. Поэтому редактор знаний должен обеспечивать автоматический выбор (управляемый онтологией уровня 2) средств для графического представления значений нестандартных величин при редактировании знаний.

Редактор онтологии интерпретирует онтологию уровня i при создании модуля онтологии уровня $i-1$. Редактор знаний интерпретирует онтологию уровня 2 при создании модуля знаний. Одна и та же онтология может интерпретироваться разными способами в разных редакторах знаний. Редакторы знаний могут отличаться не только способом интерпретации знаний, но и интерфейсом. Очевидно, что более удобный интерфейс и более понятный эксперту способ интерпретации можно обеспечить для редактора, предназначенного для интерпретации одной онтологии, а не класса онтологий. Поэтому специализированная оболочка должна позволять использование редакторов, поддерживающих разные способы интерпретации модуля онтологии уровня 2 и предоставлять возможность эксперту выбора требуемого ему редактора знаний.

Значения нестандартных величин используются не только при редактировании знаний, но также при вводе исходных данных задач. Графический способ задания исходных данных задач более удобен для специалиста предметной области, поскольку в этом случае отсутствует необходимость громоздкого вербального описания этих данных. Графическое представление результатов решения является более наглядным способом представления. Поэтому оболочка должна обеспечивать возможность ввода/вывода значений нестандартных величин при задании исходных данных задач, а также позволять использование

принятого в предметной области графического способа представления значений нестандартных величин при вводе исходных данных задач и выводе результатов их решения.

Как уже отмечалось, величина, которой принадлежит значение некоторого свойства, задается онтологией уровня 2. Оболочка должна обеспечивать автоматический выбор (управляемый онтологией) средств для графического представления значений нестандартных величин при задании исходных данных задач.

Каждый раздел сложно-структурированной ПрО характеризуется своим множеством классов прикладных задач, причем разные множества могут содержать как общие классы задач, так и специфичные для раздела. Решатель задач может быть предназначен для решения классов задач одного раздела (в этом случае он использует онтологию и знания этого раздела), либо разных разделов (в этом случае он может использовать разные онтологии и знания). В первом случае используемая решателем онтология определяется классом задач. Во втором случае требуется дополнительное указание, какие онтология и знания должны использоваться в процессе решения. Специализированная оболочка интеллектуальных систем для сложно-структурированной предметной области должна обеспечивать возможность решения задач разных классов, причем пользователь должен иметь возможность указания модуля онтологии и модуля знаний, которые надо использовать при решении задач.

Таким образом, специализированная оболочка должна содержать расширяемые библиотеки систем для решения задач разных классов, системы автоматического построения методов решения задач по их спецификации (рис. 1).

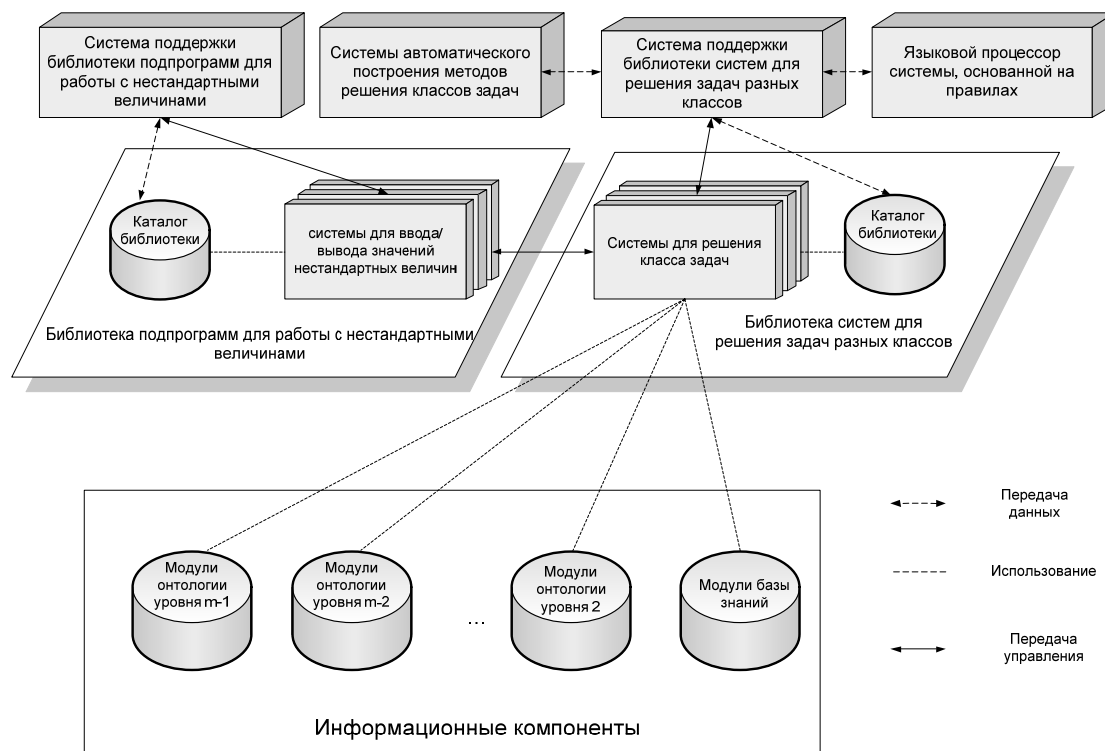


Рис. 1. Системы для решения задач и средства их разработки

Метод решения задач может быть представлен либо в виде алгоритма, либо в виде множества правил системы продукций. В первом случае для создания решателя задач используется процессор алгоритмического языка, во втором случае – процессор языка, основанного на правилах, который является одним из программных компонент специализированной оболочки.

Специализированная система обработки текстовых документов «Лота»

Специализированная система обработки текстовых документов «Лота» [Невзорова, 2001] является системой класса Text Mining. Система предназначена для анализа специализированных текстов «Логика работы», описывающих логику работы сложной технической системы в различных режимах функционирования. Основной задачей анализа является извлечение из данных текстов информационной модели алгоритмов, решающих определенную задачу в определенной проблемной ситуации, и контроль структурной и информационной целостности выделенной схемы алгоритмов.

Информационная модель алгоритма включает:

- описание входного информационного потока (типы информационных сигналов или семантическое описание информационного потока с указанием источника информации – конкретный алгоритм, конкретное измерительное устройство);
- описание процессов преобразования входных данных в выходные (допустимый способ разрешения проблемы);
- описание выходного информационного потока (типы информационных сигналов или семантическое описание информационного потока с указанием точки приема информации).

Решение основной задачи обеспечивается комплексом технологий обработки текстов, включающих:

- технологии морфосинтаксического анализа;
- технологии семантико-синтаксического анализа;
- технологии взаимодействия с прикладной онтологией.

Указанная сумма технологий формируется на основе центрального ядра – прикладной онтологии (в дальнейшем, авиаонтология [Лукашевич, 2004]), обеспечивающей согласованное взаимодействие различных программных модулей. Авиаонтология концептуально описывает предметную область информационного обеспечения различных полетных режимов антропоцентрических систем. Авиаонтология представляет собой сеть понятий предметной области. Авиаонтология относится к классу лингвистических (лексических) онтологий и предназначена для встраивания в различные лингвистические приложения.

Программный комплекс состоит из трех взаимодействующих подсистем: подсистемы лингвистического анализа технических текстов «Анализатор», подсистемы ведения онтологии «OntoEditor+» и подсистемы «Интегратор». Взаимодействие подсистем реализовано на базе технологии «клиент-сервер», причем в различных подзадачах подсистемы выступают в различных режимах (режим сервера или режим клиента).

Инструментальная система визуального проектирования «OntoEditor+» является специализированной СУБД. Система предназначена для ручного редактирования онтологий, хранящихся в реляционной базе

данных в формате TPS, а также обслуживания запросов пользователей и внешних программ. Новые возможности системы обеспечиваются функциональным набором «Лингвистический инструментарий», посредством которого реализуется встраивание прикладной онтологии в лингвистические приложения. Наиболее типичными задачами, решаемыми с помощью инструментария системы «OntoEditor+», являются: изучение структурных свойств прикладной онтологии с помощью исследовательского инструментария системы «OntoEditor+»; построение лингвистической оболочки прикладной онтологии; задача покрытия текста онтологическими входами; построение выводов по прикладной онтологии и др.

Подсистема «Анализатор» реализует основные этапы лингвистической обработки текста (графематический, морфосинтаксический и частичный синтаксический анализ).

Подсистема «Интегратор» исполняет внешний запрос на извлечение знаний из текста. Структура внешнего запроса содержит компоненты информационной модели алгоритма. Внешний запрос интерпретируется при взаимодействии с подсистемой «OntoEditor+» как структура, привязанная к прикладной онтологии. Выделение компонент информационной модели происходит на основе механизмов отождествления элементов дерева сегментов входного текста (взаимодействие с подсистемой «Анализатор») и элементов структуры запроса (взаимодействие с подсистемой «OntoEditor+»).

Инструментальная система визуального проектирования онтологий «OntoEditor+» включает:

- лингвистический инструментарий (задачи корпусного исследования (загрузка корпуса; сегментация на предложения; автоматическое ведение статистики по различным объектам корпуса), построение лингвистической оболочки онтологии, задача покрытия текста онтологическими входами, построение выводов по онтологии, поддержка протоколов информационного обмена системы «OntoEditor+» с внешними программными модулями, в том числе с внешними информационными ресурсами);
- исследовательский инструментарий.

Основные функции подсистемы «Анализатор»:

- графематический анализ;
- морфосинтаксический анализ;
- покрытие текста онтологическими входами (взаимодействие с системой «OntoEditor+»).

Основные функции подсистемы «Интегратор»:

- анализ и исполнение внешнего запроса (информационная модель алгоритма);
- интерпретация внешнего запроса в терминах прикладной онтологии (взаимодействие с системой «OntoEditor+»);
- интерпретация внешнего запроса в структурных компонентах дерева сегментов (взаимодействие с системой «Анализатор»);
- контроль информационной целостности (анализ компонент внешнего запроса).

Метод контекстного разрешения омонимии является базовым методом в интегральной технологии разрешения омонимии в системе «ЛoTA». Однако практические задачи системы выявили ряд важных

аспектов лингвистического анализа, которые стимулировали развитие новых методов разрешения многозначности.

Интегральная технология разрешения многозначности, разрабатываемая в системе «ЛОНА», включает следующие методы:

- метод контекстного разрешения функциональной омонимии;
- метод разрешения функциональной, грамматической и лексической омонимии на основе индексируемой базы устойчивых коллокаций;
- метод разрешения функциональной, грамматической и лексической омонимии на основе лингвистической оболочки онтологии.

Для эффективного встраивания в лингвистические приложения система «OntoEditor+» поддерживает группу протоколов информационного обмена с внешними программными модулями системы и внешними словарными базами данных, обеспечивая работу в режиме клиент-сервер. Разрешение многозначности (функциональной, морфологической и лексической) во входных текстах происходит на основе механизма распознавания контекстов омонимов, зафиксированных в индексируемой базе контекстов.

Разработаны три основных механизма пополнения индексируемой базы контекстов функциональных омонимов:

- ручной ввод и редактирование данных по типовым контекстам омонимов;
- импорт типовых контекстов омонимов из текстового файла, подготовленного в специальном формате представления данных;
- импорт типовых контекстов омонимов, обнаруженных специальными механизмами поиска подсистемы «Анализатор».

Данный механизм организован как запрос к подсистеме «Анализатор» с передачей ему от подсистемы «OntoEditor+» текстового корпуса, по которому проводится поиск. В процессе обработки подсистема «Анализатор» передает подсистеме «OntoEditor+» информацию об обнаруженных контекстах омонимов, которая записывается либо в индекс омонимов, либо в автоматическом режиме, либо в режиме диалога с оператором. Отличительной особенностью режима диалога является режим самообучения, который реализуется с использованием механизма журнала событий. В данном журнале в зависимости от его настройки фиксируются те или иные важные события в системе, например, изменение информации в индексе омонимов или операции взаимодействия с подсистемой «Анализатор». В режиме самообучения сохраняется и контролируется последовательность ранее сгенерированных диалогов, что обеспечивает генерацию только уникальных диалогов на разрешение омонимии без повторений.

Лингвистический инструментарий подсистемы «OntoEditor+» обеспечивает встраивание онтологии в различные приложения, связанные с обработкой текстов. Лингвистический инструментарий реализует функции загрузки корпуса текстов; автоматическое ведение статистики по различным объектам корпуса; функции предсинтаксической обработки текста (сегментация предложений, распознавание аббревиатур, разрешение омонимии на основе специальных протоколов взаимодействия с внешними словарными ресурсами); построение лингвистической оболочки онтологии; распознавание терминов прикладной

онтологии во входном тексте (задача покрытия). Сопряжение онтологического и лингвистического (грамматического) ресурсов реализуется через механизмы лингвистической оболочки онтологии. Лингвистическая оболочка онтологии создается с помощью разработанного программного инструментария, посредством которого фиксируется грамматическая информация об онтологических концептах и их текстовых формах. Каждый онтологический вход (как правило, многословный термин) снабжается соответствующей грамматической информацией, при этом для омонима разрешается соответствующая (функциональная, лексическая, морфологическая) омонимия. Грамматическая информация передается в подсистему «OntoEditor+» от подсистемы «Анализатор» на основе специальных протоколов взаимодействия. Разрешение лексической, функциональной и морфологической омонимии выполняется на основе специальных диалогов с экспертом-лингвистом. Отдельные процедуры реализуют проверки словоформ в составе терминологического входа на согласованность их грамматических характеристик, также осуществляется контроль достоверности словарной информации. Контроль достоверности обеспечивает отслеживание изменений, как в составе грамматического словаря, так и в составе онтологии. Учитывая сложность и многоступенчатость вышеперечисленных процедур, в подсистеме «OntoEditor+» разработан мастер построения лингвистической оболочки, который вызывается командой основного меню.

Подсистема «Анализатор» обеспечивает реализацию метода разрешения омонимии на основе контекстных правил, т. е. фактически используются лингвистические знания системы. Этот метод является универсальным, не зависит от специфики предметной области и обеспечивает в текущей версии точность распознавания не ниже 95 %. Однако, для данного метода существуют крайне сложные типы функциональной омонимии, например, тип «частица/союз». Разрешение данной омонимии возможно во многих случаях лишь после завершения полного синтаксического анализа.

Взаимодействие подсистемы «OntoEditor+» и подсистемы «Анализатор» осуществляется на основе специальных протоколов взаимодействия. При применении интегральной технологии разрешение многозначности происходит в два этапа. На первом этапе подсистема «Анализатор» (клиент) передает запрос на разрешение омонимии входного текста подсистеме «OntoEditor+» (сервер). Подсистема «OntoEditor+» возвращает подсистеме «Анализатор» информацию о разрешенных омонимах на основе своих методов. На втором этапе подсистема «Анализатор» разрешает омонимию оставшихся неразрешенных омонимов на основе метода контекстных правил.

Интегральная технология разрешения многозначности эффективно применяется на этапе предсинтаксического анализа в системе «ЛoTA». По существу, интегральная технология представляет собой сочетание инженерного и лингвистического подхода к решению поставленной задачи. В основе проектирования интегральной технологии лежат процессы скоординированного взаимодействия различных языковых уровней, прежде всего онтологического уровня (обеспечивающего системные модели знаний о мире) и различных языковых уровней (морфологического и синтаксического). В системе реализован эффективный механизм взаимодействия различных подсистем, обеспечивающих реализацию различных методов в составе интегральной технологии. При этом, следует признать, что сам процесс

согласования языковых взаимодействий достаточно сложен и требует определения большого числа дефиниций-правил, описывающих условия обработки концептов онтологии.

Интеллектуальная система извлечения данных и их анализа (на основе текстов) ИСИДА-Т

Целью ИСИДА-Т [Кормалев, 2006]; [Киселев, 2004], является извлечение значимой информации определенного типа из (больших массивов) текста для дальнейшей аналитической обработки. Результатом работы систем является получение структурированных данных и отношений на них.

Основные компоненты ИСИДА-Т:

- Инфраструктурные службы (конфигурирование, параллельная обработка, взаимодействие модулей);
- Лингвистический процессор;
- Модули работы со знаниями ПрО;
- Интерпретатор правил извлечения информации.

Разработанные в рамках проекта ИСИДА-Т технологии, инструменты и продукты позволяют:

- обнаруживать в электронных документах, извлекать и структурировать информацию о представляющих интерес фактах, событиях, объектах и отношениях;
- выполнять мониторинг сайтов в сети Интернет на предмет появления там значимой для пользователя информации.

Основные рабочие характеристики технологии и продуктов:

- поддержка русского языка;
- быстрая настройка на предметную область при помощи эффективных инструментальных средств;
- высокая точность и полнота анализа за счет использования предметных знаний;
- наличие встроенных средств визуализации результатов анализа в виде диаграмм и схем;
- легкая интегрируемость в другие информационные системы на любом уровне (программный или сетевой интерфейс, БД);
- функционирование под управлением ОС Windows и большинства Linux-систем;
- близкая к линейной масштабируемость при параллельной архитектуре анализа. Возможность работы на вычислительных машинах кластерного типа.

Некоторые области применения технологий семантического анализа и структурирования текстовой информации:

- информационная поддержка бизнеса (business intelligence) и управление знаниями (knowledge management);
- маркетинговые исследования;
- финансовая аналитика;
- военная и коммерческая разведка и мониторинг;
- информационная поддержка органов государственной власти (в рамках направления «Электронное правительство»);

- работа библиотек, издательств и СМИ.

Рассмотрим общую организацию инфраструктуры системы ИСИДА-Т. Краеугольным камнем системы ИСИДА-Т является точная настройка на предметную область и конкретную задачу извлечения. С одной стороны, это достигается за счет редактирования лингвистических ресурсов, ресурсов знаний, правил извлечения и правил трансформации. С другой стороны, настройка может потребовать включения в процесс обработки дополнительных специализированных методов обработки текста. Кроме того, для каждой задачи необходимо подобрать наиболее подходящие алгоритмические средства анализа из набора имеющихся. Эти аспекты требуют создания такой архитектуры, при которой легко могут добавляться и замещаться алгоритмические компоненты процесса извлечения.

Проблема конфигурирования на алгоритмическом уровне потребовала создания модульной архитектуры и декларативного подхода к определению процесса извлечения. Модули получили название обрабатывающих ресурсов в противовес лингвистическим ресурсам и ресурсам знаний. В конфигурации декларируется порядок обработки документа аналитическими модулями, потоки данных между ними, а также параметры их работы.

Обрабатывающие ресурсы можно разделить на следующие группы.

- Ресурсы предобработки. Сюда относятся средства определения кодировки документа, извлечения текста и стилевой разметки из документа, предварительной фильтрации.
- Ресурсы лингвистического анализа. Осуществляют разбор текста на отдельные слова, морфологический анализ (в том числе специализированные варианты для различных категорий имен собственных), поверхностный синтаксический анализ и определение границ предложений.
- Ресурсы извлечения. Осуществляют поиск в документе целевой лексики и синтаксических конструкций, а также первичное структурирование информации.
- Ресурсы унификации знаний и вывода. Осуществляют унификацию и отождествление элементов знаний, вывод производных знаний.
- Ресурсы подготовки результата. Осуществляют приведение извлеченной информации к определенному формату и передачу за пределы последовательности обработки (в БД, глобальный ресурс знаний, файл, приложение).

В системе ИСИДА-Т все модули, в том числе средства общего лингвистического анализа, используют структуру данных – аннотация. Аннотация – объект, который приписывается фрагменту текста (например, слову, словосочетанию, предложению, ссылке на сущность предметной области и т. д.) и описывает свойства этого фрагмента. Аннотации разбиты на конечное множество классов. Каждый класс аннотаций описывает текст в определенном аспекте. Информация о фрагменте представлена значениями именованных атрибутов аннотации. Наборы классов и атрибутов аннотаций намеренно не специфицированы, чтобы можно было использовать произвольный набор обрабатывающих модулей и представлять необходимую лингвистическую и предметную информацию. Обмен данными между модулями тоже идет в терминах аннотаций: новые аннотации могут строиться на основании полученных на предыдущих этапах анализа [Овдей, 2006]. В реализации системы ИСИДА-Т модель аннотаций была

дополнена некоторыми полезными средствами. В частности, было снято ограничение на атомарность атрибутов и добавлена возможность устанавливать ссылки между аннотациями.

Для распознавания текстовых ситуаций используется набор правил, описывающих характерные для конкретной задачи способы выражения ситуации в тексте. Эти правила задают образец для сопоставления и действия, которые должны быть произведены после успешного сопоставления. Ряд современных систем извлечения информации (в том числе, система ИСИДА-Т) берут за основу различные диалекты языка CPSL [Noy, 1999]. Использование этого языка подразумевает разметку текста при помощи аннотаций.

Язык правил, используемый в системе ИСИДА-Т, является расширением CPSL. Предлагаемые расширения преследуют две цели: 1) обеспечить возможность описывать более сложные контексты, в которых встречается целевая информация, и 2) снизить объем рутинной работы при создании системы правил за счет более компактного описания контекста [Гаврилова, 2000].

Отличия от других реализаций, например, JAPE [Calvanese, 2007] или диалекта CPSL состоят в следующем.

- Реализована встроенная поддержка расширенного спектра типов данных, в том числе, ссылок на аннотации и множественных значений. Данные этих типов могут использоваться в качестве значений переменных и значений атрибутов аннотаций.
- Логика работы интерпретатора правил приведена в максимальное соответствие поведению интерпретатора обычных регулярных выражений. Отличия от современной реализации JAPE и Montreal transducer [Calvanese, 2007] заключаются в поддержке «жадных» и «нежадных» квантификаторов и опережающей проверки.
- Поддерживаются кванторы существования (по умолчанию) и всеобщности, связывающие элементарные тесты. К кванторам может добавляться отрицание.
- Существуют языковые средства, позволяющие гибко проверять взаимное расположение аннотаций, рассматриваемых в контексте сопоставления, и прочих аннотаций во входной коллекции.
- В тестах могут использоваться функции для обращения к ресурсу знаний, например, проверки таксономической принадлежности элементов. Для более сложных запросов к ресурсу знаний используется предметно-ориентированный язык, совпадающий с языком описания левой части правил трансформации.
- Для передачи информации между элементарными тестами, а также в правую часть правил могут использоваться именованные переменные, значения которых присваиваются явно в ходе сопоставления. Множество значений переменных входит в контекст сопоставления.

Инструментальное средство проектирования онтологий Protégé

Protégé [Noy, 1999] – локальная, свободно распространяемая Java-программа, разработанная группой медицинской информатики Стэнфордского университета. Программа предназначена для построения

(создания, редактирования и просмотра) онтологий прикладной области. Её первоначальная цель – помочь разработчикам программного обеспечения в создании и поддержке явных моделей предметной области и включение этих моделей непосредственно в программный код. Protégé включает редактор онтологий, позволяющий проектировать онтологии, разворачивая иерархическую структуру абстрактных или конкретных классов и слотов. Структура онтологии сделана аналогично иерархической структуре каталога. На основе сформированной онтологии, Protégé может генерировать формы получения знаний для введения экземпляров классов и подклассов. Инструмент имеет графический интерфейс, удобный для использования неопытными пользователями, снабжен справками и примерами.

Protégé основан на фреймовой модели представления знания ОКВС (Open Knowledge Base Connectivity) [Chaudhri, 1998] и снабжен рядом плагинов, что позволяет адаптировать его для редактирования моделей, хранимых в разных форматах (стандартный текстовый, в базе данных JDBC, UML, языков XML, XOL, SHOE, RDF и RDFS, DAML+OIL, OWL).

Используемые формализмы и форматы

Изначально единственной моделью знаний, поддерживаемой Protégé, была фреймовая модель. Этот формализм сейчас является «родным» для редактора, но не единственным.

Protégé имеет открытую, легко расширяемую архитектуру и, помимо фреймов, поддерживает все наиболее распространенные языки представления знаний (SHOE, XOL, DAML+OIL, RDF/RDFS, OWL). Protégé поддерживает модули расширения функциональности (plug-in). Расширять Protégé для использования нового языка проще, чем создавать редактор этого языка «с нуля».

Protégé основан на модели представления знаний ОКВС (Open Knowledge Base Connectivity). Основными элементами являются классы, экземпляры, слоты (представляющие свойства классов и экземпляров) и фасеты (задающие дополнительную информацию о слотах).

Пользовательский интерфейс

Пользовательский интерфейс состоит из главного меню и нескольких вкладок для редактирования различных частей базы знаний и ее структуры. Набор и названия вкладок зависят от типа проекта (языка представления) и могут быть настроены вручную. Обычно имеются следующие основные вкладки: Классы, Слоты (или Свойства для OWL), Экземпляры, Метаданные.

Инструментальный комплекс автоматизированного построения онтологий Про

Инструментальный комплекс онтологического назначения (ИКОН) для автоматизированного построения онтологии в произвольной предметной области [Палагин, 2012] является системой, реализующей одно из направлений комплексных технологий Data & Text Mining, а именно – анализ и обработку больших объёмов неструктурированных данных, в частности лингвистических корпусов текстов на украинском и/или русском языке, извлечение из них предметных знаний с последующим их представлением в виде системно-онтологической структуры или онтологии предметной области.

Извлечение информации (Information Extraction) [Палагин, 2012] — это подход, позволяющий сузить круг задач, требующих специфического предметно-ориентированного решения при анализе текста. В рамках

этого подхода задача обработки текста ограничена распознаванием множества классов ключевых понятий конкретной предметной области и игнорированием всякой другой информации.

Несмотря на то, что системы извлечения информации могут строиться для выполнения различных задач, подчас сильно отличающихся друг от друга, существуют компоненты, которые можно выделить практически в каждой системе.

В состав почти каждой системы извлечения информации входят четыре основных компонента, а именно: компонент разбиения на лексемы, некоторый тип лексического или морфологического анализа, синтаксический анализ (микро- и макроуровень), модуль извлечения информации и модуль для анализа на уровне конкретной предметной области. В зависимости от требований к конкретному программному продукту в приведённую выше схему добавляют дополнительные модули анализа (специальная обработка составных слов; устранение омонимии; выделение составных типов, которое может также быть реализовано на языке правил извлечения информации; объединение частичных результатов).

Разбиение на слова при анализе европейских языков не является проблемой, поскольку слова отделяются друг от друга пробелом (или знаками препинания). Тем не менее, для обработки составных слов, аббревиатур, буквенно-цифровых комплексов и ряда других особых случаев требуются специфические алгоритмы. С границами предложений, как правило, тоже больших проблем не возникает. Однако при анализе таких языков как японский или китайский, определение границ слова на основе орфографии невозможно. По этой причине системы извлечения информации, работающие с такими языками, должны быть дополнены модулем сегментирования текста на слова.

В некоторые системы наряду с обычными средствами лексического и морфологического анализа могут быть включены модули для определения и категоризации атрибутов частей речи, смысловых нагрузок слов, имен или других нетривиальных лексических единиц.

ИКОН предназначен для реализации множества компонентов единой информационной технологии:

- поиск в сети Интернет и/или в других электронных коллекциях (ЭлК) текстовых документов (ТД), релевантных заданной ПрО, их индексацию и сохранение в базе данных;
- автоматическая обработка естественно-языковых текстов;
- извлечение из множества ТД знаний, релевантных заданной ПрО, их системно-онтологическая структуризация и формально-логическое представление на одном (или нескольких) из общепринятых языков описания онтологий. Кроме того, внутри этой технологии реализуется процедура построения, визуализации и проверки семантических структур синтаксических единиц ТД и понятийных структур заданной ПрО в виде онтографа, названного начальной онтологией ПрО;
- создание, накопление и использование больших структур онтологических знаний в соответствующих библиотеках;
- системная интеграция онтологических знаний как одна из основных компонент методологии междисциплинарных научных исследований;
- другие процедуры, связанные с автоматизацией приобретения знаний из множества естественно-языковых объектов.

ИКОН состоит из трёх подсистем и представляет собой интеграцию разного рода информационных ресурсов (ИР), программно-аппаратных средств обработки и процедур естественного интеллекта (ЕИ), которые, взаимодействуя между собой, реализуют совокупность алгоритмов автоматизированного, итерационного построения понятийных структур предметных знаний, их накопления и/или системной интеграции. Обобщённая блок-схема ИКОН представлена на рис. 2.

Подсистема «Информационный ресурс» включает блоки формирования лингвистического корпуса текстов, баз данных языковых структур и библиотек понятийных структур. Первый компонент представляет собой различные источники текстовой информации, поступающей на обработку в систему. Второй компонент представляет собой различные базы данных обработки языковых структур, часть из которых формируется (наполняется данными) в процессе обработки ТД, а другая часть формируется до процесса построения онтологии ПрО и, по сути, является ЭлК различных словарей. Третий компонент представляет собой совокупность библиотек понятийных структур разного уровня представления (от наборов терминов и понятий до высокоинтегрированной онтологической структуры междисциплинарных знаний) и является результатом реализации некоторого проекта (проектирования онтологии ПрО и/или системной интеграции онтологий).

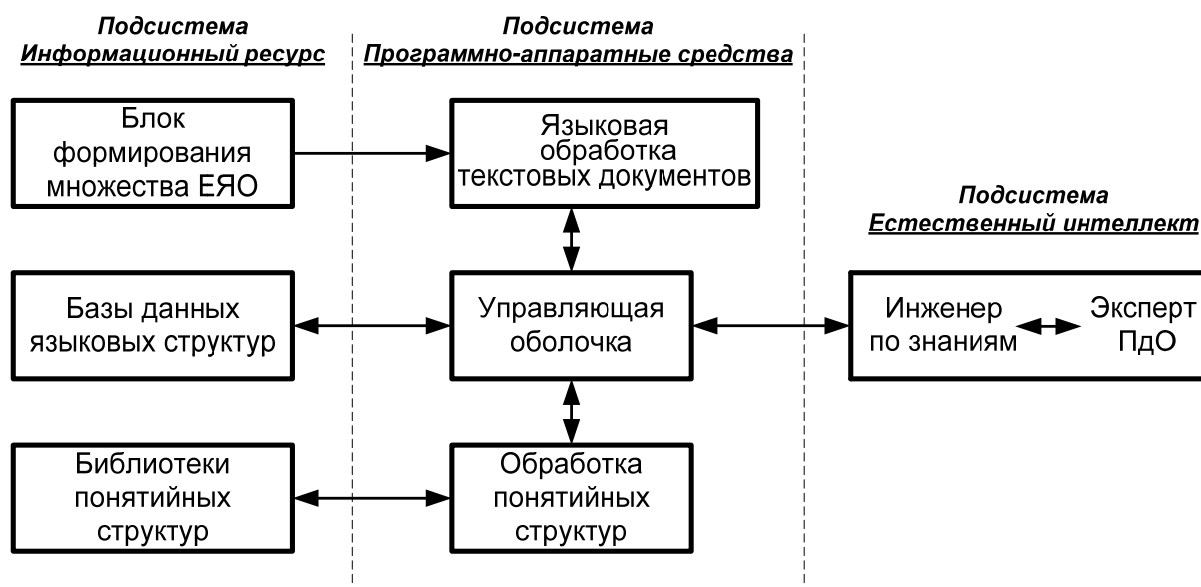


Рис. 2. Обобщённая блок-схема ИКОН

Подсистема «Программно-аппаратные средства» включает блоки обработки языковых и понятийных структур и управляющую графическую оболочку, с помощью которой инженер по знаниям осуществляет общее управление процессом использования связанных информационных технологий.

Подсистема «Естественный интеллект» осуществляет подготовку и реализацию процедур предварительного этапа проектирования, а на протяжении всего процесса осуществляет контроль и проверку результатов выполнения этапов проектирования, принимает решение о степени их завершенности (и в случае необходимости – повторении некоторых из них).

ТОДОС – IT-платформа онтологических информационно-аналитических экспертных систем

ТОДОС – инновационный комплекс программно-информационных и методических средств управления знаниями с использованием подходов онтологического управления корпоративными информационными ресурсами, где человек рассматривается как источник определения новых знаний для передачи их в форме собственного знания через инструментарий [Стрижак, 2013]. ТОДОС обеспечивает для пользователя единую интегрированную точку доступа – «единое окно» – к информации и приложениям системы для обеспечения интерактивного взаимодействия при решении прикладных задач.

Принцип ТОДОС – «Ситуационная осведомленность» за счет предоставления пользователям необходимой информации, касающейся направлений их деятельности и достаточной для принятия эффективного решения:

- Удобное, интуитивно-понятное, многоаспектное представление аналитической информации;
- Обеспечение работы с неструктурированной и слабо структурируемой информацией;
- Работа с информацией и результатами анализа из любой точки сетевого доступа;
- Удовлетворение поиска и запросов пользователей – извлечение знаний;
- Обработка и анализ контента, агрегирование и упорядочивание по заданному критерию;
- Поддержка принятия решений на основе анализа больших объемов информации;
- От данных - к пространственно-распределённым системам управления (ГИС) - от ГИС - к информации;
- Обеспечение взаимодействия и обратной связи.

Концепция ТОДОС:

- консолидация и интеграция всей имеющейся корпоративной информации и предоставление ее через систему «единого окна», за счет чего повышается уровень осведомленности всех категорий пользователей в их деятельности;
- обеспечение «бесшовной» системной интеграции информационных технологий и инноваций с целью создания информационно-аналитических ресурсов для внедрения в бизнес-процессы организации;
- создание условий «ситуационной осведомленности» для всех заинтересованных категорий пользователей с многоаспектным анализом массивов документов, их анализом, сравнением, рейтингованием с выводом отчетов и результатов анализа;
- обеспечение онтологического управления информационными массивами, которые объединяются в единое корпоративное информационное пространство – онтолого-управляемую систему корпоративных знаний;
- поиск в сети Интернет и в файловых электронных коллекциях текстовых документов, релевантных тематике исследований и экспертизы;

- автоматическая обработка естественно-языковых текстов с выделением поверхностных семантических отношений для дальнейшего их анализа;
- извлечение из множества документов знаний, релевантных выбранной предметной области, их системно-онтологическая структуризация и формально-логическое представление, а также построение, визуализация и верификация семантических структур синтаксических единиц текстовых документов и категориальных знаний заданной предметной области в виде онтологического графа;
- автоматизированное построение онтологий и тезаурусов предметных областей для организации системы управления знаниями;
- автоматизированный анализ и создание системы рейтингов объектов исследования и процессов с ними связанных с учетом всего множества факторов, влияющих на соответствующие объекты и процессы;
- обеспечение многовекторного исследования объектов и процессов с целью выявления параметров влияющих на их состояние, развитие и принятия соответствующего объективного решения.

ТОДОС – это:

- Модуль КОНСПЕКТ – контекстно-семантический анализ естественно-языкового текста и построение таксономии документов;
- Модуль КОНФОР – классификация и генерация онтологий предметной области;
- Модуль ЭДИТОР – конструирование трансдисциплинарных онтологий;
- Модуль АЛЬТЕРНАТИВА – онтология задачи выбора для информационно-аналитической поддержки принятия решений и обеспечения процессов многофакторного анализа и рейтингования;
- ПОИСКОВАЯ МАШИНА – поиск лексических структур на основе лингвистической обработки большого количества распределенных сетевых текстовых массивов;
- ЛИНГВИСТИЧЕСКИЙ КОРПУС – электронная библиотека со средствами ассоциативного поиска семантически связанных информационных массивов;
- Модуль КРИПТО – защита информационных массивов от несанкционированного доступа.

Модуль КОНСПЕКТ представляет собой лингвистический процессор [Величко, 2009], который обеспечивает первичное формирование лингвистического корпуса [Широков, 2005] и позволяет решать следующие практические задачи:

- 1) Выделение однословных и многословных терминов документа.
- 2) Формирование списка терминов с подсчетом их количества в тексте.
- 3) Навигация по списку терминов с отбором и отражением контекстов предложений из текста с выделенными терминами.
- 4) Автоматизированное построение тематического словаря по коллекции документов для использования в локальной полнотекстовой поисковой системе с целью повышения релевантности поиска.

5) Конспектирование текста по заданной теме. Конспект – это краткое изложение содержания статьи или книги. Система автоматически создает текстуальный конспект, где основные положения произведения, доказательства и выводы передаются словами автора, то есть цитатами. Тема конспекта задается пользователем программы с помощью ключевого слова или словосочетания. Из исходного текста в конспект отбираются предложения, содержащие тему, и автоматически отбираются термины текста, ассоциативно связанные с заданной темой. Пример конспекта, который автоматически построен в среде модуля КОНСПЕКТ, приведен на рис. 3.

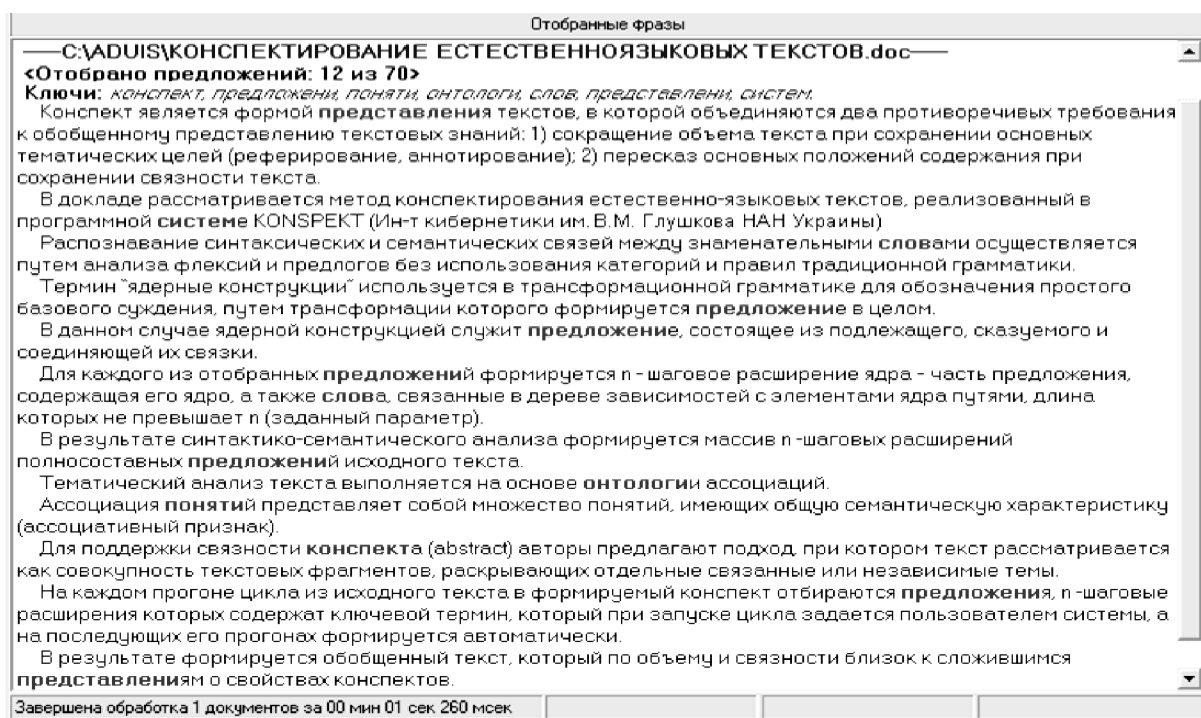


Рис. 3. Фрагмент среды модуля КОНСПЕКТ

Результаты, которые можно получить в ближайшее время при условии доработки разработанной программы:

- 1) Повышение качества обработки текстов на украинском и русском языках за счет увеличения объема словаря системы;
- 2) Автоматическое определение тематических направлений документа;
- 3) Сортировка документов по тематическим направлениям;
- 4) Построение семантической сети терминов документа;
- 5) Объединение семантических сетей терминов для нескольких документов;
- 6) Поиск информации в сети Интернет с помощью поискового модуля системы ТОДОС с дополнительной автоматической семантической фильтрацией результатов поиска. Для дополнительной фильтрации используются алгоритмы формирования семантических категорий заданной тематики, извлеченных из

текста. Список документов, которые получает пользователь после дополнительной фильтрации, сокращается в несколько раз;

7) Конспектирование текста по заданной теме (для украинского языка).

Модуль КОНФОР предназначен для построения семантических связей между объектами предметной области на основе индуцированной обработки отношений между объектами тематических классов ПрО. Для этого тематические классы описываются пользователем в соответствии с нижеприведённой схемой:

(имя класса объектов 1),(имя связи), (имя объекта 1),..., (имя объекта n)

(имя класса объектов 2),(имя связи), (имя объекта 1),..., (имя объекта j)

.....

(имя класса объектов m),(имя связи), (имя объекта 1),..., (имя объекта k)

Согласно схемы заполняется таблица, например MS Excel, ячейки столбца А которой содержат имена материнских вершин графа, ячейки столбца В – имена связей между вершинами, ячейки столбцов от С до Z – имена дочерних вершин. Выходной структурой модуля КОНФОР является онтограф ПрО [Палагин, 2012], [Довгий, 2013], вершины которого носят имена объектов.

Модуль ЭДИТОР предназначен для создания, редактирования, просмотра и анализа сетей понятий и формирования закономерностей, представленных в виде набора значений признаков, которыми описываются начальные понятия предметной области. Таким образом, описание предметной области является наглядным и легко интерпретируется пользователем.

Выделение закономерностей происходит методом индуктивного формирования понятий на основе пирамидальной сети.

Пирамидальной сетью называется ациклический ориентированный граф, вершинами которого выступают термины (понятия) предметной области, а ребрами (дугами) – семантические связи между терминами [Гладун, 1994].

Представление информации в виде графа позволяет изучать не только отдельный термин (понятие), но и получать все его семантические связи с другими понятиями, тем самым осмысливая его роль в данной системе знаний или в ходе решения задачи.

Однако сетевой граф может выступать не только средством организации знаний. Расширяя традиционные функции, граф можно превратить в среду, в которой обеспечивается активная работа со знаниями, а также оригинальным образом решаются учебные задачи.

Концептуальная модель предметной области или онтология состоит из графа (иерархии понятий) предметной области, связей между ними и законов, которые действуют в рамках этой модели.

Модуль АЛЬТЕРНАТИВА обеспечивает упорядочивание объектов-концептов онтологии, на основе интегрированной обработки характеризующих их свойств. Для этого используются весовая, бальная и лингвистическая шкалы. Каждая такая шкала определяет значение критериев, характеризующих свойства объектов тематической онтологии ПрО. В общем случае свойства-критерии характеризуются различными

степенями важности, которые при решении задачи выбора задаются некоторыми действительными числами – весовыми коэффициентами. Перед решением задачи для каждого критерия необходимо сформировать его значение для каждой альтернативы [Саати, 1989].

Модули ПОИСКОВАЯ МАШИНА и ЛИНГВИСТИЧЕСКИЙ КОРПУС, обеспечивают маркировку и индексирование семантических единиц, определяющих и описывающих контексты объектов тематических онтологий ПрО. Контексты семантических единиц составляют электронную библиотеку.

Модуль КРИПТО обеспечивает защиту информационных ресурсов электронной библиотеки системы. Защита реализуется на основе гомоморфного кодирования. Для создания эффективных методов защиты системы реализуются механизмы инфраструктуры публичных ключей (PKI). Инфраструктура публичных ключей позволяет объединить электронные подписи, публичные и частные ключи и пирамиду виртуальных лиц или организаций, которые могут выполнять действия по управлению тематическими онтологиями и отображающими их документами.

На основе анализа сопроводительной документации, справочных материалов, доступных в открытых источниках информации (интернет, научные труды), и, в отдельных случаях, практической работы с программными системами конструирования онтологий (ИКОН, ТОДОС, Protégé), построена сводная таблица, обобщающая сравнительные технические характеристики некоторых специализированных инструментальных средств инженерии онтологий (таблица 1).

Заключение

В статье представлен обзор специализированных инструментальных средств инженерии онтологий. Как видно из сравнительной таблицы, в которой представлены различные системы конструирования онтологий, стратегическим направлением развития таких систем является интероперабельность. Эффективность использования онтологических описаний в основном зависит от возможностей применения форматов их описаний при решении различных сложных прикладных задач. Наиболее конструктивным методом решения этой проблемы может быть разработка онтологий задач, которые смогут обеспечить семантическую синхронизацию взаимодействий различных по тематикам информационных систем на основе установления отношений между свойствами их концептов.

Библиография

- [Клини, 1957] Клини С. К. Введение в метаматематику [Текст] / С. К. Клини. – М. : Иностранная литература, 1957. – 526 с.
- [Палагин, 2012] Палагин А. В. Онтологические методы и средства обработки предметных знаний / А. В. Палагин, С. Л. Крытый, Н. Г. Петренко. – [монография] – Луганск : изд-во ВНУ им. В. Даля, 2012. – 323 с.
- [Гаврилова, 2000] Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем / Учебник для вузов. – СПб, Изд-во «Питер», 2000. – 384 с.
- [Палагин, 2009] Палагин А. В., Петренко Н. Г. Системно-онтологический анализ предметной области // УСиМ. – 2009. – № 4. – С. 3–14.
- [Овдей, 2006] Овдей О. М. Обзор инструментов инженерии онтологий / О. М. Овдей, Г. Ю. Проскудина. – Российский научный электронный журнал «Электронные библиотеки», 2004. – т. 7. – Вып. 4, ISSN 1562-5419. – Режим

- доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2004/part4/op>. – Дата доступа: 19.10.2006. – Название с экрана.
- [Noy, 1999] Noy N. SMART: Automated Support for Ontology Merging and Alignment / N. Noy, M. Musen. – Stanford Medical Informatics, Stanford Univ. – 1999. – 24 p. Режим доступа: <http://ais-portal.ru/2009/03>. – Дата доступа: 17.05.2010.
- [Calvanese, 2007] Software Tools for Ontology Design and Maintenance / [D. Calvanese, P. Dongilli, G. De Giacomo, A. Kaplunova, D. Lembo, M. Lenzerini, R. Möller, R. Rosati, S. Tessaris, M. Wessel, I. Zorzi] Project Deliverable D21,TONES, 2007. http://www.researchgate.net/publication/240754938_Software_Tools_for_Ontology_Design_and_Maintenance
- [Артемьева, 2008] Артемьева И. Л., Рештаненко Н. В. Интеллектуальная система, основанная на многоуровневой онтологии химии // Программные продукты и системы. – 2008. – № 1. – С. 84-87.
- [Филатов, 2007] Филатов В. А. Разработка высокоэффективных средств создания и обработки онтологических баз знаний / Филатов В. А., Щербак С. С., Хайрова А. А. – Системи обробки інформації, випуск 8 (66), 2007. – С. 120–124. – Режим доступа: www.nbu.gov.ua/portal/natural/soi/2007_8/Filatov.pdf. – Дата доступа: 17.11.2010. – Название с экрана.
- [Невзорова, 2001] Невзорова О. А., Федун Б. Е. Система анализа технических текстов «Лота»: основные концепции и проектные решения. // Изв. РАН. Теория и системы управления.– 2001. № 3. – С. 138-149.
- [Лукашевич, 2004] Лукашевич Н.В., Невзорова О. А., АвиаОнтология: анализ современного состояния ресурса // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2004. – Москва, Наука, 2004– т.2 - с. 424-430.
- [Кормалев, 2006] Кормалев Д. А., Куршев Е. П. Развитие языка правил извлечения информации в системе ИСИДА-Т // Труды международной конференции «Программные системы: теория и приложения». — Т. 2. — М. : Физматлит, 2006. — С. 365-377.
- [Киселев, 2004] Киселев С. Л., Ермаков А. Е., Плешко В. В. Поиск фактов в тексте естественного языка на основе сетевых описаний // Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2004. – Москва, Наука, 2004.
- [Chaudhri, 1998] OKBC: A Programmatic Foundation for Knowledge Base Interoperability. V. Chaudhri, A. Farquhar, R. Fikes P. Karp J. Rice // Fifteenth National Conf. on Artificial Intelligence. AAAIPres/The MIT Press, Madison, P.600-607, 1998.<http://www.oracle.com/technetwork/java/javase/index.html>
- [Стрижак, 2013] Стрижак О.Є. Засоби онтологічної інтеграції і супроводу розподілених просторових та семантичних інформаційних ресурсів [Текст] / Екологічна безпека та природокористування: Зб. наук, праць / М-во освіти і науки України, Київ. нац. ун-т буд-ва і архіт., НАН України, Ін-т телекомунікацій і гло-бал. інформ. простору; редкол.: О.С. Волошкіна, О.М. Трофимчук (голов. ред.) [та ін.]. - К., 2013. - Вип. 12. - с.166-177]
- [Величко, 2009] Величко В. Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем / В. Величко, П. Волошин, С. Свитла // "Knowledge – Dialogue – Solution" International Book Series "INFORMATION SCIENCE & COMPUTING", Number 15. – FOI ITHEA Sofia, Bulgaria. - 2009. – pp.24-31.
- [Широков, 2005] Корпусна лінгвістика / В.А. Широков, О.В. Булгаков, Т.О. Грязнухина та ін.. – К.: Довіра, 2005. – 471 с.
- [Довгий, 2013] Комп'ютерні онтології та їх використання у навчальному процесі. Теорія і практика. : Монографія / С. О. Довгий, В. Ю. Величко, Л. С. Глоба, О. Є. Стрижак та ін. – К. : Інститут обдарованої дитини, 2013. – 310 с.
- [Стрижак, 2014] Онтологія задачі вибору та її застосування при аналізі лімнологічних систем [Текст] / [О.Є. Стрижак, В.В. Горборуков, О.В. Франчук, М.А. Попова] // Екологічна безпека та природокористування: Зб. наук, праць / М-во освіти і науки України, Київ. нац. ун-т буд-ва і архіт., НАН України, Ін-т телекомунікацій і гло-бал. інформ. простору; редкол.: О.С. Волошкіна, О.М. Трофимчук (голов. ред.) [та ін.]. - К., 2014. - Вип. 15. – С. – 172-183.
- [Гладун, 1994] Гладун В.П. Процессы формирования новых знаний [Текст] / В. П. Гладун. – София : СД «Педагог 6», 1994. – 192 с.
- [Саати, 1989] Саати Т. Принятие решений. Метод анализ иерархий: Пер. с англ. / Т. Саати.-М.: Радио и связь, 1989.- 316 с.

Таблица 1 Системы конструирования онтологий

	ИКОН	ЛОТА	SIMER+MIR	ИСИДА-Т	Protégé	Ontolingua	InTez	OntoSTUDIO	OntoEdit	ТОДОС
Общая информация										
Разработчик	ИК НАНУ	?	РАН ИПС ДЦШ	РАН ИПС ДЦШ	SMI, Stanford University	KSL, Stanford University	С-ПТ ДУ	Semafora systems GmbH	Ontoprize	НЦ НАНУ, ИК НАНУ, ИТТИП НАНУ, НТУ КПИ, НАУКМА
Версия	1.0	?	?	?	4.1	4.0	?	3.1		3.0
URL	http://aduis.com.ua/	—	—	http://isida-t.ru/	http://protege.stanford.edu/	http://www.ksi.stanford.edu/software/ontolingua/	http://intez.ru	http://www.semafora-systems.com/en/products/ontostudio	http://www.semafora-systems.com/products/ontoeedit	http://editor3.inhost.com.ua
Доступность	Закранный код, ограниченный доступ	Закранный доступ	Закранный доступ	Закранный доступ	Закранный код, свободный доступ	Свободный доступ	Закранный код, свободный доступ	Закранный код, ограниченный доступ	Свободная лицензия	Открытый код, свободный доступ с разделением прав
Интерфейс пользователя	Локальное приложение	?	Локальное приложение	Локальное приложение	Локальное приложение, Веб-приложение	Веб-приложение	Локальное приложение	Локальное приложение, Веб-приложение	Локальное приложение	Веб-приложение
Язык интерфейса ПО	Русский, Украинский	Русский	Русский	Русский, Английский	Английский	Английский	Русский	Английский	Английский	Украинский, Русский, Английский, Арабский
Архитектура программного обеспечения										
Расширение	Плагины	—	—	—	Плагины	Плагины	—	Плагины	Плагины	Плагины
Язык ПО (платформа)	Java, Delphi	C++	?	?	Java	Lisp	?	Построенный на платформе Eclipse	Java	Java, Delphi, C++, HTML 5
Сохранение онтологий	файлы, СУБД	СУБД (формат TPS)	файлы	файлы	файлы, СУБД	файлы	файлы, СУБД	файлы, СУБД	файлы	файлы, СУБД
Архитектура приложения	n-уровневая	Клиент-сервер	3-х уровневая	3-х уровневая	3-х уровневая	Клиент-сервер	Клиент-сервер	3-х уровневая	3-х уровневая	Клиент-сервер
Функционал инструментального средства										
Язык обрабатываемых ЛКТ	Русский, Украинский	Русский	Русский	Русский	Английский	Английский	Русский	—	—	Украинский, Русский, Английский

	ИКОИ	ЛоТА	SIMER+MIR	ИСИДА-Т	Protégé	Ontolingua	InTez	OntoSTUDIO	OntoEdit	ТОДОС
Управляющая графическая оболочка системы	+	+	+	+	+	+	+	+	+	+
Системная интеграция онтологий	+	-		-	+	-	-	+	+	+
Система поиска ТД (Поиск ТД во внешних и внутренних информационных ресурсах)	+	-		-	-	+			-	+
Визуальное проектирование, редактирование онтологических структур	+	+		+	+	+		+	+	+
Автоматизированное построение онтологий ПДО	+-	-		-	-	-		-	-	+
Ручное построение онтологий ПДО	+	-		+	+	+		+	+	+
Автоматический лингвистический анализ ТД, описывающих заданную ПДО (синтаксический, семантический, морфосинтаксический анализ)	+	+		+	-	-		-	-	+

Интегрируемость									
Икон	Лота	SIMER-MIR	ИСИДА-Т	Protégé	Ontolingua	InTez	OntoSTUDIO	OntoEdit	ТОДОС
Конвертация файлов описания онтологий	-	-	-	+	KIF, Prolog, Ontolingua, LOOM, CLIPS, IDL	Экспорт в OWL	Импорт из UML 2.0, Database schemas (Oracle, MS-SQL, DB2, MySQL), Excel tables, Outlook E-Mails, Folder structures of the file system	Импорт из OWL, RDF(S), DAML+OIL, Flogis Экспорт в OWL, RDF(S)	Импорт из OWL, OXML, RDF(S), FLogic, Excel tables, Экспорт в OWL, OXML, RDF(S), Excel tables, Prolog, CLIPS, IDL
Интеграция с другими инструментальным средствами	?	?	?	Возможна средствами плагинов	ОКВС, Chimera		Protégé	OntoAnnotat, OntoMat, Semantic-Miner	Protégé - возможна средствами плагинов
Интеграция с базой знаний WolframAlpha	-	-	-	-					+
Интеграция с поисковыми системами : Google, Bing, Exalead, Solvera, Галактика Zoot, Информбюро и т.д.	-	-	-	-					+
Интеграция с пространственно-распределёнными системами управления (ГИС-системы)	-	-	-	-					+
Параллельная обработка данных	-	-	+	-					+

Информация об авторах



Величко Виталий Юрьевич – к.т.н., старший научный сотрудник, Институт кибернетики им. В. М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: velychko@aduis.com.ua

Основные области научных исследований: индуктивный логический вывод, обработка естественно-языковых текстов



Малахов Кирилл Сергеевич – младший научный сотрудник, Институт кибернетики им. В. М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: malahov@live.com

Основные области научных исследований: онтологический инжиниринг



Семенов Виталий Васильевич – аспирант, Луганський національний університет імені Тараса Шевченка

Основные области научных исследований: онтологический инжиниринг



Стрижак Александр Евгеньевич – к.т.н., старший научный сотрудник, Институт телекоммуникаций и глобального информационного пространства НАН Украины, Киев-186, 03186, Чоколовский бульвар, 13; e-mail: sae953@gmail.com

Основные области научных исследований: корпоративные интеллектуальные системы, поддержка принятия решений

Integrated Tools for Engineering Ontologies

Velychko V.Yu., Malahov K.S., Semenov V.V., Strizhak A.E.

Abstract: The article presents an overview of current specialized ontology engineering tools, as well as texts' annotation tools based on ontologies. The main functions and features of these tools, their advantages and disadvantages are discussed. A systematic comparative analysis of means for engineering ontologies is presented.

ACM Classification Keywords: I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods, H. Information Systems – H.2 DATABASE MANAGEMENT – H.2.4 Systems

НАУЧНЫЕ ПОДХОДЫ СИСТЕМЫ ФОРМИРОВАНИЯ КОНКУРЕНТОСПОСОБНОГО УЧИТЕЛЯ

Юрий Иванович Завалевский

Аннотация: В статье обосновываются научные подходы, в соответствии с которыми происходит формирование конкурентоспособности учителя. Доказывается необходимость комплексного применения мировоззренческих идей обозначенных подходов. Анализируются авторские концепции трактовки содержания данных подходов, на основе которых выделяются характерные черты в соответствии с проблемами конкурентоспособности личности, учителя.

Ключевые слова: методологические основы исследования, научные подходы, системный подход, синергетический подход, личностно-деятельностный подход, аксиологический подход, ресурсный подход, компетентностный подход.

Актуальность проблемы

Исследование вопросов конкурентоспособности учителя, в частности в его инновационной деятельности, опирается на получение новых знаний о сути, функции, структуре, содержании, механизмах стимулирования как конкурентоспособности, так и инноватики, которые обеспечивают раскрытие методологии как учения о научном методе познания или как системы научных принципов, на основании которых базируется исследование и осуществляется выбор совокупности познавательных средств, методов, приемов исследования.

Методология исследования основывается на комплексе научных идей, отражающих не только уровень владения ученым фундаментальными истинами, теориями, принципами, но и исповедуемую им глобальную философско-мировоззренческую позицию.

Цель статьи заключается в обосновании научных подходов в соответствии с проблемой формирования конкурентоспособности учителя.

Введение

В исследованиях М. Васильевой, В. Гринев, Т. Роговой внимание акцентируется на многоуровневом, иерархическом характере методологии, которая выходит из концептуальных положений И. Блауберга, Э. Юдина [Васильева, 2003], [Гринева, 1970], [Рогова, 2006], [Юдин, 1973].

В нашем исследовании методологические основы раскрываются не на основании характеристики уровней методологии по Э. Юдину, поскольку они часто содержательно пересекаются, а путем раскрытия сути методологических подходов.

Основная часть

Методологическими основами исследования вопросов конкурентоспособности личности учителя считаем основные положения системного, синергетического, личностно-деятельностного, аксиологического, ресурсного, компетентностного подходов. Рассмотрение подходов как методологических основ исследования вопросов конкурентоспособности личности учителя позволило выделить в исследовании три аспекта:

- 1) положения, касающиеся развития личности учителя;
- 2) положения, касающиеся самореализации личности, в частности профессионально-педагогической самореализации учителя;
- 3) интегрированность подходов в системе формирования конкурентоспособности учителя на основе его инновационной деятельности.

Особенностями системного подхода является изучение объекта познания или его преобразования как системы; восприятие системы как единства отношений и связей отдельных частей (компонентов), которые обуславливают выполнение функций; выявление различных типов связей между частями системы и сведение их в единую системную теоретическую картину; раскрытие частей системы в логической последовательности – цель, задачи, стратегия, структура, функции, принципы, содержание, средства (приемы, методы, формы, технологии), системообразующий фактор, условия, критерии, показатели, уровни, результаты, направления развития системы; реализация в практической деятельности существенных признаков системы – целеустремленности, целостности, структурности, функциональности, иерархичности, самоорганизации, взаимосвязи с внешней средой.

Формирование конкурентоспособности личности учителя с позиции системного подхода как процесс количественных и качественных изменений, которые обуславливают приобретение человеком новых интегративных качеств, позволяет рассматривать профессионально-педагогический рост учителя, который осуществляется в соответствии с внешними требованиями субъектов, включенных в эту систему, и предусматривает:

- изучение конкурентоспособности учителя с позиции более сложного целого – формирования и развития личности учителя, его самореализации;
- определение цели конкурентоспособности учителя, которая положительно влияет на его успешность на рынке труда;
- обоснование компонентов конкурентоспособности учителя (личностного, профессионального, поведенческого), которые подчинены общей цели – достижению уровня конкурентоспособности на рынке труда;
- установление взаимосвязей компонентов конкурентоспособности, их значимости в процессе профессиональной реализации учителя, на основе которых можно изменять, координировать, упорядочивать связи между частями, создавая тем самым условия для достижения положительных результатов;

- установление взаимосвязи конкурентоспособности учителя с внешней средой, которая может влиять положительно или отрицательно на процесс профессиональной реализации;
- выявление новых интегративных качеств по достижению уровня конкурентоспособности учителя, в частности его активность в отношении постоянного саморазвития, использования инноваций.

Акмеологический подход позволяет рассматривать приобретение учителем конкурентоспособности как процесс восхождения к вершинам акме – самой высокой точки в развитии, самовыражении.

Характеризуя ценность акмеологического подхода для процесса самореализации будущего учителя, Л. Рыбалко определяет, что **многоуровневую** модель акме определяют или равноценно высокие результаты, полученные человеком в различных направлениях деятельности, или один очень высокий результат среди других, более слабых по значимости [Рыбалко, 2007]. Действительно, чтобы стать конкурентоспособным, учитель должен пройти процесс достижения вершины профессионального мастерства, но при этом ориентироваться не на идеальные цели личного профессионального уровня, а на внешние ориентиры в оценке его труда. Таким образом, вершина акме (конкурентоспособность) будет четко обусловлена теми требованиями, которые определяют ожидания общества относительно результатов труда учителя.

Следует обратить внимание на то, что с позиции связи акмеологического и системного подходов процесс достижения конкурентоспособности является оптимальным развитием внутреннего потенциала учителя, его восхождением к акме.

Новым междисциплинарным научным направлением, основанным на идеях системности и целостности мира, научных представлениях человека о нем и самого себя в этом мире, является синергетика – теория самоорганизации. Использование синергетического подхода в педагогических исследованиях среди ученых и практиков имеет как сторонников, так и противников проникновения его в педагогику. Считаем, что синергетика является альтернативной формой теоретического освоения действительности, а ее положения вполне приемлемы в изучении вопросов развития личности, что, в частности в рамках нашего исследования касается проблемы самоорганизации личности в условиях динамично меняющейся профессиональной среды.

Проблема самореализации личности на основе идей синергетического подхода отражена в исследованиях А. Ковалева [Ковалев, 1970], который утверждает, что синергетический подход к процессу самореализации личности позволяет выработать новую стратегию в отношении безболезненного перехода к новым, актуализированным современным социокультурным условиям, к качествам личности выпускника, который вступает в самостоятельную жизнь; Н. Лосевой [Лосева, 2007] (доказывает, что процесс самореализации личности является синергетическим феноменом, а синергетика позволяет наиболее адекватно выявить его суть через понятие «открытая система») [Нестеренко, 2004] (рассматривает синергетическое измерение самореализации личности в условиях трансформации общества, доказывает значимость синергетики для изучения вопросов самореализации будущего специалиста).

Развитие личности в синергетике объясняется с позиций открытости, нелинейности, неравнозначности, неустойчивости личности, хаоса как фактора создания, конструктивного механизма эволюции. В процессе взаимодействия личности с окружающей средой происходит опредмечивание – распределмечивание ее сущностных сил, развитие потенциальных возможностей. Потребности развития могут реализовываться только открытой системой. Нелинейность личности обеспечивает оптимальный выбор способов развития. Развитие нелинейной системы почти не предсказуемо, количество вариантов больше, чем в линейных системах. Развитие открытых систем, к которым относится личность, происходит благодаря **неравнозначности**, неустойчивости, случайности, которые порождают хаос: «Только системы, далекие от равновесия, системы в состоянии неустойчивости, способны спонтанно организовывать себя и развиваться» [Цуканова, 1985]. Заметим, что случайность побуждает личность к необходимости выбора, а из хаоса собственными силами может развиваться новая организация. Самоорганизация в синергетическом измерении влияет на развитие личности.

Исходя из изложенного, можно сделать вывод, что развитие личности учителя обусловлено его способностью к самоорганизации, которая влияет на выбор направлений и способов самореализации, достижения конкурентоспособности. Итак, существует множество направлений и способов самореализации, саморазвития и учитель имеет право выбирать лучшие, оптимальные для него. Ценным в профессиональной подготовке будущего учителя является знание о самоограничении в выборе способов самореализации, ибо, если он будет знать, на что способен, то не будет попусту тратить время, собственные усилия, материальные средства. Учитель может описать, рассчитать желаемые, оптимальные для себя «сценарии» развертывания собственного внутреннего потенциала, и, в случае их соблюдения, заведомо будет ускорять время выхода на **аттрактор**.

Ведущей синергетической идеей в плане исследования является выяснение взаимосвязи самореализации и самоорганизации учителя.

Конкурентоспособность учителя с позиции синергетики можно определить как самоорганизующуюся объективацию сущностных сил в процессе педагогической деятельности, успешность которой зависит от взаимодействия с другими подсистемами педагогического процесса. Инноватика является фактором конкурентоспособности учителя, той побудительной силой, которая упорядочивает внутренние сущностные потенции педагога для максимального их воплощения в жизнь в форме самовыражения и самоутверждения. Реализация внутреннего потенциала требует самокоррекции, самоизменения, которые могут осуществляться в процессе самообразования и самовоспитания.

Суть личностно-деятельностного подхода заключается в признании того, что формирование личности как продукта социально-исторического развития, носителя культуры, ее уникальности, интеллектуальной и нравственной свободы, права на уважение, предполагает опору на естественный процесс саморазвития способностей, самоопределения, самореализации, самоутверждения, который происходит в процессе деятельности, самодеятельности и зависит от определенных условий.

В экспериментальном педагогическом исследовании исключительно важным является сочетание личностного и деятельностного подходов, поскольку именно таким образом достигается необходимая

целостность в формировании личности. Развитие личности происходит в деятельности. Оно обусловлено тем, что, с одной стороны, деятельность в любых видах, формах непосредственно и опосредованно влияет на изменения в структуре личности, а с другой, - личность имеет право выбирать те виды и формы деятельности, к которым она склонна и которые удовлетворяют потребности личностного развития.

Особенностями личностно-деятельностного подхода к вопросам конкурентоспособности личности является признание ее как субъекта, а деятельности как средства становления и реализации в определенном общественном пространстве; направленность на развитие личности, способной выбирать, оценивать, программировать, конструировать те виды деятельности, которые являются адекватными ее природе и удовлетворяют ее потребности; ориентация на развитие личности, способной к самоизменению, самосовершенствованию.

В рамках исследования важным является признание личности учителя как субъекта педагогической деятельности, с позиции которой он полноценно будет реализовывать собственный потенциал, постепенно достигая конкурентоспособности в профессиональной среде. С позиции субъекта педагогической деятельности учитель характеризуется как человек активный, способный к продуктивной деятельности, самоорганизации, саморегуляции, саморазвитию, самосовершенствованию. Он имеет право на обращение к собственному «Я», на реализацию личностно-профессиональных способностей, на выбор тех средств, которые нужны ему для полной самореализации в процессе профессиональной подготовки. Внутренним условием его существования является понимание, восприятие себя как субъекта педагогической деятельности, а внешним - привлечение к процессу педагогической деятельности. Учителя можно признать субъектом педагогической деятельности, когда он активно овладевает различными способами раскрытия внутреннего потенциала в воспитательной, исследовательской, учебно-познавательной, самообразовательной деятельности, реализует способности к самосовершенствованию, саморазвитию, то есть полностью вовлечен в инновационную деятельность. В рамках этого требования субъектом деятельности является новое, преобразованное качество личности, которое свидетельствует о полной перестройке всей системы психологической организации личности, а не только о реализации способностей, целей в соответствии с потребностями.

Развитие субъектности учителя является показателем его профессионально-педагогической самореализации. Субъектность как осознание, реализация человеком своей сути характеризует определенную собранность человека, освоение собственных сущностных сил. Это качество личности проявляется в осознанном стремлении изменить не только себя, но и положительно повлиять на окружающий мир. В понимании субъектности как качества важным представляется единство мотива и нормы поведения, формирование устойчивой жизненной позиции, отношения человека к другим людям с различным общественным статусом – к высшим и низшим, к старшим и младшим, к лицам противоположного пола. А. Волкова называет и свойство, в основе которого лежит отношение человека к себе как к деятелю [Волкова, 1998]. К тому же, субъектность личности способствует достижению вершин личностно-профессионального развития. Благодаря субъектности человек является организатором собственного развития, оптимально использует внутренние и внешние ресурсы, что сказывается на его конкурентоспособности.

В процессе проведенного анализа мы пришли к выводу, что с позиции личностно-деятельностного подхода развитие личности необходимо рассматривать как развитие ее субъектности, развитие процессов самости, в частности профессионально-педагогическую реализацию учителя - субъектную деятельность, которая способствует его конкурентоспособности.

Особенностью аксиологического подхода является изучение значимой роли ценностей в жизни человека, механизмов преобразования общечеловеческих, общественно значимых ценностей в собственные, личностные. Отметим, что на рубеже веков меняются ценности, которыми человек жил, творил, воспитывался, учился. Пришло время осознания высшей ценности – это бытие человека, переоценка отношения к нему, создание условий для выбора им правильных ценностных ориентиров в жизни, закалывание жизненной позиции, что будет способствовать потенциальному самораскрытию, самореализации. В условиях перестройки рынка труда, роста спроса на конкурентоспособного, мобильного специалиста человек начинает осознавать значимость максимальной реализации своих личностных возможностей, достижения определенных вершин, совершенствования способностей, построения карьеры на современном этапе перестройки и обновления общественного государственного строя в Украине.

В профессиональной реализации учителя под понятием «аксиологический подход» понимаем направленность педагогического процесса в школах на реализацию ведущих положений аксиологии, педагогической аксиологии, усвоение которых способствует переводу общечеловеческих, профессионально-педагогических ценностей на уровень личностных, формирование ценностей самости. Ведущими понятиями аксиологического подхода есть ценность, аксиологическая характеристика личности как субъекта ценностных отношений, общие аксиологические категории – значимость, идеал, смысл, социальные нормы, оценка, ценностные ориентации и отношение, установка, а также личностные ценности – потребности, мотивация, цель, развитие, самореализация, профессионализм и тому подобное.

Итак, аксиологический аспект достижения конкурентоспособности учителя предусматривает:

- направленность цели профессионально-педагогической самореализации на приобретение учителем смысла жизни в процессе приобретения опыта педагогической деятельности, которая определяется функциями смысла жизни, а именно: интеграцией системы ценностей, целевой организацией жизни, прогнозированием будущего, формированием нравственной оценки поведения и деятельности, осознанием собственной значимости, решением противоречий между «Я»-реальным и «Я»-идеальным, а также адаптивной, защитной, экспрессивной, координационной, стимулирующей, организующей, познавательной, регулирующей функциями;
- усвоение учителем базовых общечеловеческих и профессионально-педагогических ценностей, реализация собственно созданных ценностей в процессе профессиональной деятельности на основе таких аксиологических требований как признание человека в качестве высшей ценности; предоставление права учителю на свободное самоопределение и самореализацию; овладение учителем личностной системой знаний и умений; сформированности профессионально-педагогических ценностей в процессе интериоризации и экстериоризации; интеграция социокультурных, образовательных, профессионально-

педагогических ценностей в процессе формирования профессионализма как акмеологической ценности; поддержка развития внутреннего потенциала учителя как ценности и установка на самовыражение в процессе профессионально-педагогической самореализации;

- выявление и формирование ценностных ориентаций учителя, а именно развитие у него ценностных отношений, то есть его внутренней профессиональной позиции как совокупности отношений к себе, педагогической деятельности, ученикам, коллегам, родителям, общественности; формирование ценностного поведения как целенаправленных действий, которые направлены на реализацию усвоенных ценностей;
- создание условий в процессе профессиональной деятельности для содействия построению субъект-субъектных отношений между участниками педагогического взаимодействия; стимулирование развития процессов самости, которые имеют первоочередное значение в формировании педагогического мастерства и профессионализма.

Особенностями ресурсного подхода является изучение ресурсов личности и выявление путей их раскрытия и реализации в различных видах деятельности в соответствии с природными наклонностями человека.

Ведущие положения ресурсного подхода в аспекте исследования конкурентоспособности обуславливаются взаимосвязью таких понятий, как «ресурс», «резерв», «потенциал». Уточним, что под понятием «ресурс» подразумевается то, что человек, общество имеют вообще, а понятие «резерв» – это запас чего-либо, что специально сохраняется; возможность, которая еще не использована в действии.

К ресурсам личности принадлежат сущностные силы личности (цель, потребности, мотивы, интересы, способности), которые отражают сознательное, бессознательное, переходное состояние, то есть являются запасом того, чем человек может воспользоваться в стандартных и нестандартных ситуациях.

Развитие ресурсов личности происходит в процессе достижения конкурентоспособности. Человек как биосоциальное существо *интериоризирует* из среды общественные запросы по профессии, которые накладываются на его природные задатки и потенции. Ресурсы как совокупность всего того, чем владеет личность, таким образом пополняются, восстанавливаются, и при мотивации, активизируются в процессе профессиональной конкуренции. Часть из них реализуется постоянно или частично, в определенной степени, то есть выступает как *актуальное личности*. Другая часть ресурсов находится в нереализованном состоянии, то есть в потенциальном состоянии, и к тому же залегает в человеке в форме двух пластов. Один из них является запасом незадействованного и «Я» о нем не знает, а другое «Я» знает, но специально скрывает, есть резерв потенциала. В первом случае можно говорить в целом о собственном потенциале, а во втором - о резервном потенциале. При разных обстоятельствах потенциальное человека переходит в актуальное. Такие обстоятельства могут быть субъективными - человек хочет изменений, реализации скрытых ресурсов, а также объективными – как того требует профессиональная подготовка.

Раскрытие и использование учителем своего ресурсного потенциала способствует его продвижению к самосовершенствованию в профессии, соответственно вовлечению в процесс инновационной

деятельности. Это можно объяснить тем, что в процессе инноватики происходит наращивание потенциала, человек обогащается новыми смыслами, ценностями, изменяется система потребностей, а в процессе реализации трансформированных сущностных сил расширяются возможности проявить, выразить собственные новообразования, связанные с развитием способностей, наклонностей и тому подобное. Перевод потенциального в актуальное позволяет человеку посмотреть на себя с позиции другого человека, заметить то, чего не было выявлено и реализовано ранее. Нарращивая и реализуя свой потенциал, человек совершенствуется, развивается. Особенно это заметно в процессе приобретения учителем профессионального опыта при его активности в использовании новейших технологий. Интериоризуя педагогический опыт, учитель пополняет собственные ресурсы, создает педагогический потенциал.

Сегодня в научно-педагогической литературе и педагогической практике активно обсуждается проблема модернизации образования на основе компетентностного подхода, который является весьма актуальным, поскольку подразумевает целостный опыт решения жизненных проблем, выполнения ключевых функций, социальных ролей, компетенций современного учителя.

Компетентностный подход в образовании шире, чем подход с позиций предметных знаний, умений, навыков, и включает в себя широкие гуманистические, морально-этические, культурные, эстетические, мотивационные и другие компоненты, нацеленные на творчество, действие, инициативу, выполнение, результат.

Главным стимулом для развития усилий в области компетентностного подхода в образовании стали требования бизнеса и предпринимательства к квалифицированным работникам принимать ответственные решения на основе полученных знаний и их интеграции. Следует учитывать, что на рынке труда необходимы не сами по себе знания, а способности специалиста выполнять соответствующие профессиональные и социальные функции, практически решать различные виды производственных задач [Бондаренко, 2005].

Компетентностный подход предполагает смещение акцента с накопления нормативно-определенных знаний, умений и навыков к развитию у учителей способности практически действовать, применять навыки и опыт успешных действий в ситуациях профессиональной деятельности и социальной практики.

Анализ идей компетентностного образования позволяет уточнить отдельные положения по исследуемой теме:

- компетентность учителя определяется его способностью выполнять профессионально-педагогические задачи в соответствии с требованиями учебных программ; запросов учеников и их родителей по развитию ребенка во внеклассной деятельности; к профессиональной реализации за пределами школы;
- компетентность вызывает способность к актуальному выполнению профессиональной деятельности, что проявляется в постоянном обновлении знаний, владении новой информацией для успешного решения профессиональных задач в настоящее время и в данных условиях;
- компетентность включает в себя как содержательный (знания), так и процессуальный (умение) компоненты. Это означает, что компетентный человек должен не только понимать суть проблемы, но и

уметь практически ее решать. Компетентный специалист в зависимости от конкретных условий может применить определенный метод решения проблемы.

Выводы

Итак, методолого-теоретическую базу проблемы конкурентоспособности учителя рассматриваем как систему, в которой интегрированы фундаментальные философские идеи, разные подходы к изучению педагогических явлений и технологических приемов, стимулирующих к активности в процессах самосовершенствования.

Библиография

- [Бондаренко, 2005] Бондаренко Н. В. Запросы работодателей к качеству профессиональной подготовки работников / В. Бондаренко // Вестник общественного мнения. - 2005. - № 3. - С. 41-49.
- [Васильева, 2003] Васильева М. П. Теория педагогической деонтологии: монография / М. П. Васильева; Харьк. гос. пед. ун-т им. С. Сковороды. - Х.: Новое слово, 2003. - 216 с.
- [Волкова, 1998] Волкова Е. Н. субъектности педагога: теория и практика: Автореф. дис. на соиск. науч. степени д-ра психол. наук: спец. 19.00.01 «Общая психология, психология личности, история психологии» / Е. Н. Волкова. - М., 1998. - 36 с.
- [Гринева, 2001] Гринева В.М. Формирование педагогической культуры будущего учителя (теоретический и методический аспекты): Автореф. дис. д-ра пед. н. 13.00.04 В.М. Гринева; Ин-т педагогики и психологии проф. образования АПН Украины. - К., 2001. - 45 с.
- [Ковалев, 1970] Ковалев А. Г. Психология личности / А. Г. Ковалев. - 3-е изд., Перераб. и дополн. - М.: Просвещение, 1970. - 391 с.
- [Лосева, 2007] Лосева Н. М. Самореализация личности преподавателя вуза как общепедагогическая проблема: Автореф. дис. на соискание ученой. степени д-ра пед. наук: спец. 13.00.04 «Теория и методика профессионального образования» / Н. М. Лосева. - Х., 2007. - 43 с.
- [Нестеренко, 2004] Нестеренко Г. Возможности самореализации личности в контексте синергетической модели высшего образования / Г. Нестеренко // Высшее образование Украины. - 2004. - № 1. - С. 20-26.
- [Рыбалко, 2007] Рыбалко Л. С. Методолого-теоретические основы профессионально-педагогической самореализации будущего учителя (акмеологический аспект): монография / Л.С.Рибалко. - Запорожье: ЗГМУ, 2007. - 443 с.
- [Рогова, 2006] Рогова Т. В. Теоретические и методические основы персонализированного подхода в управлении педагогическим коллективом средней общеобразовательной школы: Автореф. дис. на соискание ученой. степени д-ра пед. наук: спец. 13.00.04 «Теория и методика профессионального образования» / Т. В. Рогова. - Х., 2006. - 36 с.
- [Цуканова, 1985] Цуканова Е. В. Деструктивные параметры взаимодействия в ситуации временного дефицита / Е. В. Цуканова // Психологические исследования общения. - М.: 1985. - С. 285-299.
- [Юдин, 1973] Юдин Э. Г. Становление и сущность системного подхода / Э. Г. Юдин, И. В. Блауберг. - М.: Наука, 1973. - 270 с.

Информация об авторе



Завалевский Юрий Иванович – заместитель директора Института инновационных технологий и содержания образования, доцент, кандидат педагогических наук, 03035 ул. Митрополита В. Липкивского 36, г. Киев, Украина

Scientific Approaches of the System for Formation of the Teachers Competitiveness**Yuriy Ivanovich Zavalevskiy**

Abstract: *The scientific approaches, according to which teacher's competitiveness are formed, are discussed in this article.*

Keywords: *methodological bases of research, scientific approach, system approach, synergetic approach, student-activity approach, the axiological approach, resource-based approach, competence approach.*

ON THE TRAINING OF CONSOLIDATED INFORMATION ANALYSTS

**Kateryna Solovyova, Andrey Danilov, Olga Pishchukhina, Panasovska Yuliia,
Olga Lashyna, Maria Kozlova, Maryna Khoroshenko, Maksym Voroniy**

Abstract: The current development of society requires the use of modern information technologies of business administration, regardless the field of activity. For automation and informatization an enterprise requires the introduction of innovative technologies, including technologies for knowledge management. To ensure the integration of modern information technology corresponding to the noospheric stage of scientific development organizations require highly qualified analysts for consolidated information (business intelligence analysts, information analysts, knowledge managers). The Social Informatics Department of the Kharkiv National University of Radio Electronics is engaged in the training of such specialists, its workers developed a standard for the preparation of consolidated information analysts in Ukraine. This standard is largely in line with the preparation of Competitive Intelligence Professionals (Competitive Intelligence)[1, 2].

Special difficulties in the training of consolidated information analysts in a lack of teaching hours are related to the need of development of a special kind of systematic analytical and creative thinking, as well as the ability to perceive and implement modern world standards of information and organizational culture. The training program is formed in a manner to develop thinking and understanding of the relevance, necessity and features of work with knowledge at the graduates from various universities and bachelor systems. An important role in this process belongs to the integration of educational and scientific process at the department that allows us to prepare highly qualified specialists in a short time, but the main feature is the use of training in the preparation of a noospheric system methodology - systemology and modern systemological method – the systemological classification analysis [3].

The article deals with the relevance of the training and development of the specialty through the use of modern tools of knowledge management, some aspects of the training of consolidated information analysts, methods and tools to improve the quality of their education in terms of reducing academic load and reduce the direct interaction between teachers and students.

Keywords: Knowledge management, consolidated information analyst, systemological classification analysis, information technology, classification, social network on the Internet, model, e-learning, UML diagrams, learning organization, software testing, business process analysis and modeling requirements.

Introduction

Enhancing of the role of information technology in society is a natural process that has accelerated within recent years. The emergence of new methods, techniques and approaches to work in various subject areas requires the use of modern information tools, which will significantly improve the efficiency of employees.

A distinctive feature of the modern higher education in business analysts' training is growing volume of variable information, skills, the mastery of which is necessary for graduates to prove their competence and create a competitive advantage in the labour market.

To improve the quality of information analysts' training systemology and the modern method corresponding to the noospheric stage of scientific development should be used – the systemological classification analysis which will effectively organize and consolidate knowledge and information improving the quality of information processing and the development of effective managerial decisions that will ensure the sustainable development of the organization.

Another area of improvement of the educational process in terms of reducing teaching hours and academic load due to the introduction of the credit-modular system is the development of e-learning (e-learning systems) used by students for self-development material or for enhancing of their skills. Information and computer technologies are an integral part of the higher education development in Ukraine, which is currently under the influence of the European integration process. The learning process of business analysts in the modern university requires the use of e-learning tools within the organization of independent work of masters, and the creation of computer systems, allowing partially to master skills without the direct involvement of the teacher is the main direction of development and improvement of learning while using information technology [5].

It should be pointed out that while working with the knowledge and information semantics the direct communication between a master and a teacher in the dialogue is vital. Therefore, despite the active implementation of distance education in the educational process, training of highly skilled analysts is not possible without a direct interaction between teachers and students in learning and knowledge transfer. To partially solve the tasks the authors suggest the use of knowledge management in information systems – particularly the knowledge-oriented social networks.

The Purpose of the Work

The purpose of the work is to present the research to improve the quality of training in the field of information analysts – namely consolidated information analysts. The proposed methods and approaches are aimed at improving the quality and level of intellectual capital of the organization through the use of systemological approach.

The goals of this research are:

- the study of relevance of knowledge systematization in the training of the analysts using the method of systemological classification analysis at the example of developing of the testing methods classification useful for learning;
- the study of the methods for optimization of analysts' training through the use of e-learning technology;
- the study of the methods to develop knowledge-oriented social networking in the training of highly qualified analysts.

The Method of the Research

As the method of the research and solving the set tasks the method of systemological classification analysis which refers to the noospheric stage of scientific development and helps to solve complex ill-structured problems effectively in non-formalized qualitative semi-structured domains [3, 4, 5 etc.] was employed.

On the Relevance of Knowledge Management Competencies in Consolidated Information Analysts' Work

In today's economy, the engine of which is information and knowledge, the organization sees more value in its intellectual assets than tangible ones [6]. Firms applying knowledge management are increasingly winning in the market, as the use of knowledge management helps to increase the competitiveness of an organization.

Knowledge management helps to organizations: it creates a competitive advantage, increases productivity, stimulates innovation, promotes cooperation, encourages and harnesses the power of education to increase social capital, attracts and retains human capital to create and use structural capital, shares best practices and processes, and provides leadership and effective decision-making, increases the level of customers' satisfaction [6].

Knowledge management is a means for performing multiple tasks meaningful information processing, including: management, acquisition, storage, presentation and knowledge transfer. To solve these problems it is necessary to form the analysts' organization knowledge management competence. These competencies are important knowledge and skills in the modern consolidated information analysts; they reflect the ability to work with the knowledge, to manage knowledge and information flows, processes, and contribute to training and the development of organization.

For an analyst of consolidated information to have full knowledge management competencies, he or she must possess the knowledge, skills and abilities, with which it is possible to solve the problems of knowledge management. The most important one is the application of systemological classification analysis as the noospheric method of systematization, classification, presentation, acquisition, extraction and production of new deep knowledge. Using this method will allow to visualize and formalize knowledge of almost every industry, including ill-structured ones.

It is necessary to define the requirements for skills and knowledge for consolidated information analysts obtaining knowledge management competencies. To solve this problem it is advisable to carry out a study of these competences and formalize their requirements, for example, in the form of using case diagrams, UML language and Rational Rose tools. Figure 1 shows a fragment of the diagram for requirements for knowledge and skills of the consolidated information analysts who possess the competencies of knowledge management.

Diagram requirements for skills and knowledge of analysts who have competencies in knowledge management identifies and visualizes the basic requirements for skills and knowledge of consolidated information analysts who have knowledge management competencies and some links arising between these knowledge and skills.

The studies in this field allow to identify the skills which should be developed and the knowledge required for efficient work of the consolidated information analysts.

On the Development of a Classification Model for Testing Types to Improve the Quality of Information Analysts

Software testing is a young and developing science. At present there is no single source of information, which would clearly describe the process of software testing or systematize experience of people working in this field.

As a rule, young professional testers get their knowledge from many different sources – and often it's inaccurate and contradictory. As a result of the lack of models describing the actual testing area is not completely clear what

the software testing process is, according to what uniform rules and how it is carried out. In each case, their own unique system is being produced, and this system may be unsuitable for any other situation.

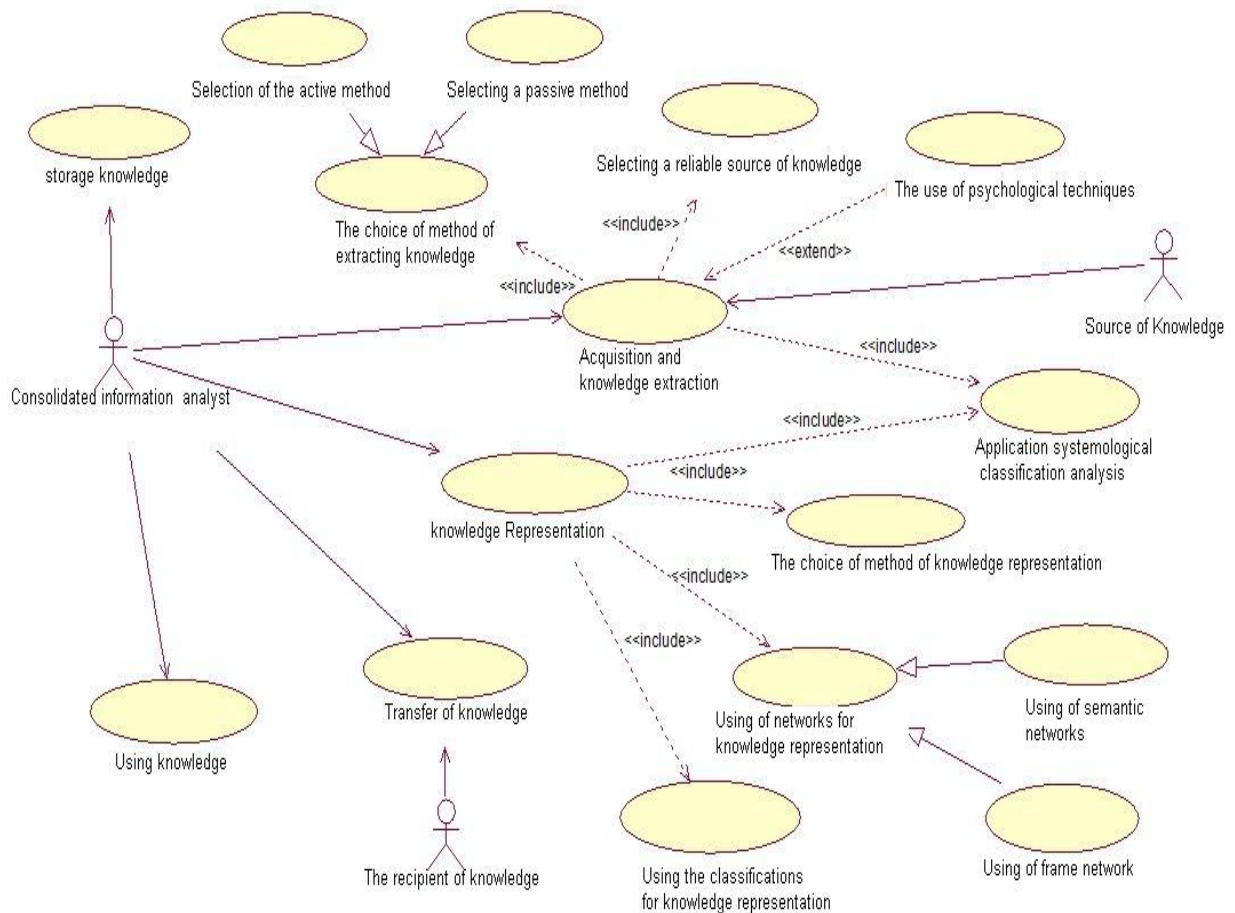


Figure 1. Diagram of the Requirements for Knowledge and Skills of Consolidated Information Analysts Having Knowledge Management Competencies

Nowadays there are many different classifications of software testing on different grounds (not only by testing types). A distinctive feature of all these classifications is their being contradictory and incomplete. This situation regularly baffles even the experts with experience.

The most common types of classification models of software testing is divided into functional, non-functional and regression testing types. This model includes a main field of testing, but does not take into consideration the time factor and the degree of being prepared for the software testing process. Functional and non-functional test types are "independent" and used in clearly defined within the scope of application, depending on their purpose. Regression testing is applicable to all areas, it covers is any type of testing, as is, in fact, screening. Regression testing, as opposed to a functional or non-functional ones, can only be applied after the first cycle of testing has been carried out when a portion of errors and defects have been already found, recorded and eliminated.

Functional and non-functional testing forms may be performed independently at any stage of the software development. Regression testing depends on the functional and non-functional ones, as it's being carried out on their basis, and often uses their tests results repeating the test cycles.

Based on the discrepancy identified above (as for the time and procedures), it is logical to assume that a certain distinction is necessary because the regression testing has properties that cannot be on the same level as the functional and non-functional test types have.

Functional and non-functional types of testing can be attributed to basic testing, as the test cases related to these two types, are the basis for software testing process.

Regression types of testing should be attributed to screening test types, which are not always independent, often repeat functional and nonfunctional tests which have been already carried out, being in some cases aimed only at test bug fixes.

Besides, the least explored area of software testing is the test of the product's documentation, which is not mentioned in most of the existing classifications of today.

Documentation testing is hardly paid with attention, because in the process of development of this type of testing they always lack time for it. However, the time spent on checking and testing of technical documents could save for a development team more than one working day in the future. Testing of the documents as well as the regression testing cannot be placed in one line with non-functional and functional types as there is a difference in time and area of application again.

Testing of documentation is a preparatory kind of testing that is supposed to identify errors, inaccuracies and contradictions in the documents according to which the project is being developed. It is necessary as well as the regression, as many mistakes are repeated more than once and lead to even a greater number of deviations from the requirements in the development and implementation process.

Thus, if we consider the classification of software documents testing not in terms of functional, non-functional and regression tests, but in terms of the time scale and purpose and include them into a system of preparatory, primary and verification kinds, the controversy related to the different start time of the tests and different test facilities are eliminated. Such a classification model of the software testing types can be especially useful both for those who are now just trying to learn how to become a professional tester and are involved into the collection of information about the subject domain, as well as for experienced professionals who have been working for many years and have decided to be engaged into scientific activities and organize their accumulated experience.

The Research and Modeling of E-Learning Systems as a Tool to Support the Educational Process In Terms of Academic Load Reduction

Peculiarities of consolidated information analysts' training related to meaningful description of the processes being studied and the analysis of qualitative information restrict the use of e-learning aimed mainly at the formation of the solution search skills requiring the implementation of the unambiguous algorithm. However, some aspects of analysts training in the field of decision-making theory, social processes modeling with the use of formal methods for their description, etc. allow partial application of computer training programs due to the parametric representation of problems in these knowledge areas. E-learning systems are considered as auxiliary means offering only additional ways of learning to the basic functions of a teacher related to the training of specialists in this field.

E-learning should not only perform functions related to the provision of theoretical material, but also to control the functions of training, testing, problem solving skills development and assessment of students with the search for "white" spots in their responses, and providing recommendations to return to the part of the theoretical material

which caused difficulties, or for additional training in the practical development of the material. Besides, the guidelines should contain information about the errors, the ways of their elimination and the level of student's knowledge that forms the distinctive feature of intelligent e-learning is the presence of feedback to the student, supported by special interactive modules that are responsible for the assessment of the student responses and errors place and type diagnosis which is an ill-structured area.

The analysis of existing e-learning systems shows that they are presented in the form of testing or information systems without feedback in the learning process. One of the major disadvantages of such systems is that they cannot be used during the learning cycle and restricted to a certain set of primitive functions (electronic books, databases and learning, etc.). The analysis of the existing development of computer training programs in the preparation of analysts revealed that it is the formation of feedback is most difficult in these developments, as it includes quite a number of formalized criteria for assessment and diagnosis of the errors causes. The proposed criteria do not always fully reflect the nature and function of the feedback realized by the teacher as the studied developments did not take into account features and specificity of analysts' training. This determines the urgency of the problem and the need to develop e-learning systems with feedback for the consolidated information analysts' training that will make the process of learning more effective and aimed at the students' self-development.

In the study of feedback nature of the educational process its main components implemented by a teacher have been revealed: assessment function, the function of the search of the location where an error occurred in the case of an incorrect answer and recommendatory functions of re-examination of the material according to the results determined by the location of the error. The last two functions are combined in a block of errors intellectual analysis that performs error detection in case of the presence of solution results mismatching to set values. Next the location of the error is detected at any stage of problem solving process and the status of the error is referred to a particular lecture topic or type of tasks. Then the problem (the gap) in the student's knowledge is diagnosed and the recommendations to improve the result are formed, for example a student can be offered to explore the theoretical material with the topics or chapters, to learn an algorithm for solving the proposed problems, to try the tasks of a simpler level, etc.

The problem of intelligent e-learning systems development is considered as the creation of complex computer projects and is closely linked with the development of requirements for their structure, functions and software content. Requirements describe the behavior of the system, the properties of the system, its classes and attributes and limit the computer system development process [7].

Analysis and modeling requirements require special formalized approaches that can accommodate the interests of all those involved in the process of e-learning: students with various psycho-types, basic knowledge levels, information processing perception and speed and teachers with their experience of "how to do", "how to teach", "how to assess" and "how to identify the true causes of the error". Moreover, other external constraints such as training programs standards, curricula, etc., should also be taken into account.

The study of this ill-structured domain proves that the development of a complex computer system is not possible without the development of information support requirements. Information support for the requirements development for an intelligent e-learning system must contain a formal description of its functions, conditions and transitions, logical aspects of the sequence of actions and interactions between its members in the form of adequate visual models of functional and user's requirements.

User requirements describe the goals and objectives that e-learning can solve and indicate what students can do with the system. The ways of presenting these options are to use the scripting requirements and an "answer-set" approach. Functional requirements define the functionality of the software which developers must generate and users should be able to perform. To cover all the requirements it is necessary to develop models that represent them at a certain level of abstraction and can detect incorrect, inconsistent, missing or redundant requirements. Data flow diagrams, state-transition diagrams, cooperation and sequence diagrams relate to such a conceptual representation of an intellectual computer system. Each model describes a particular aspect of the system, uses a set of diagrams or formal description as well as reflects the point of view and is the subject of a variety of people with specific interests, roles or tasks. These models provide a useful tool for the analysis of software development problems and information exchange between the developed modules. Modeling extends the coverage and treatment of functional and user requirements, and improves the quality of the system [8].

The most effective way to create visual models is a special standard object-oriented modeling language - Unified Modeling Language (UML). At the level of abstraction of separation, which is suitable for the requirements analysis, UML-notation is used for all types of diagrams for intelligent e-learning systems. Analysis modeling techniques are supported by various business automation software tools known as computer-aided software (CASE-tools). This choice is due to a number of advantages if compared with conventional methods: 1) they are easy to improve the diagrams accuracy in case of requirements' repetition, since it is impossible to create a model with a high degree of adequacy of the actual processes immediately due to external and internal changes during the simulation; 2) CASE- tools support rules for each method of modeling and can identify syntax errors, inconsistencies and detect semantic errors in the scheme , and that is not always possible without the use of such tools.

Figure 2 and Figure 3 show the sequence diagram and the cooperation diagram which reflect transactions between the separate modules of the intelligent e-learning system.

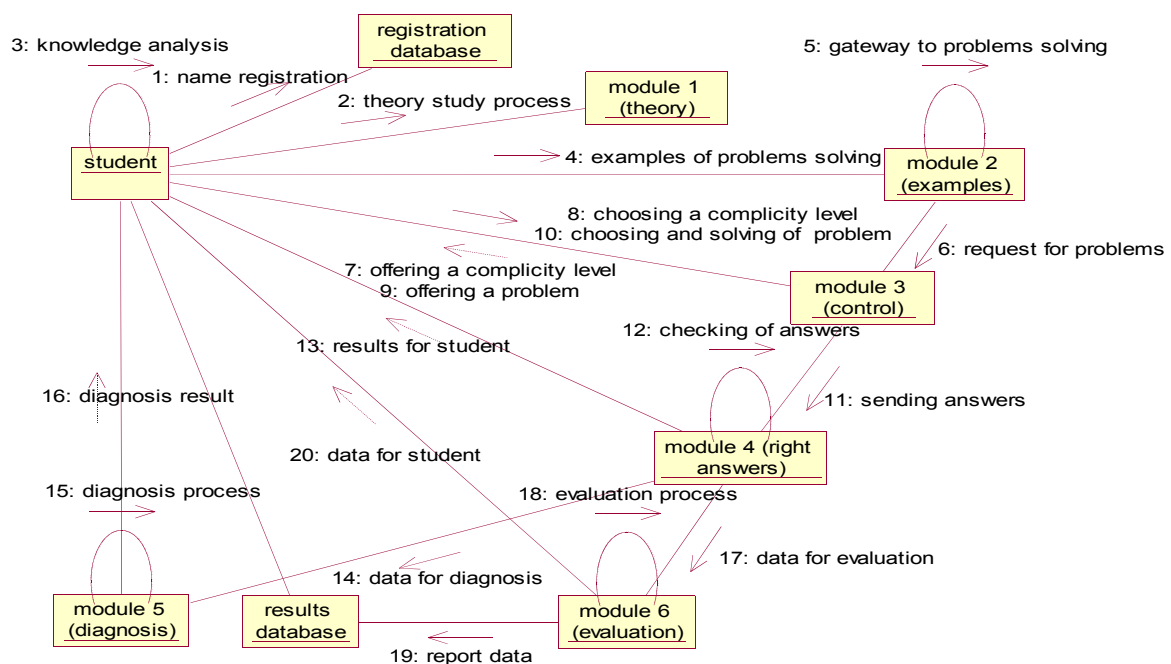


Figure 2. Cooperation Diagram of an E-Learning System

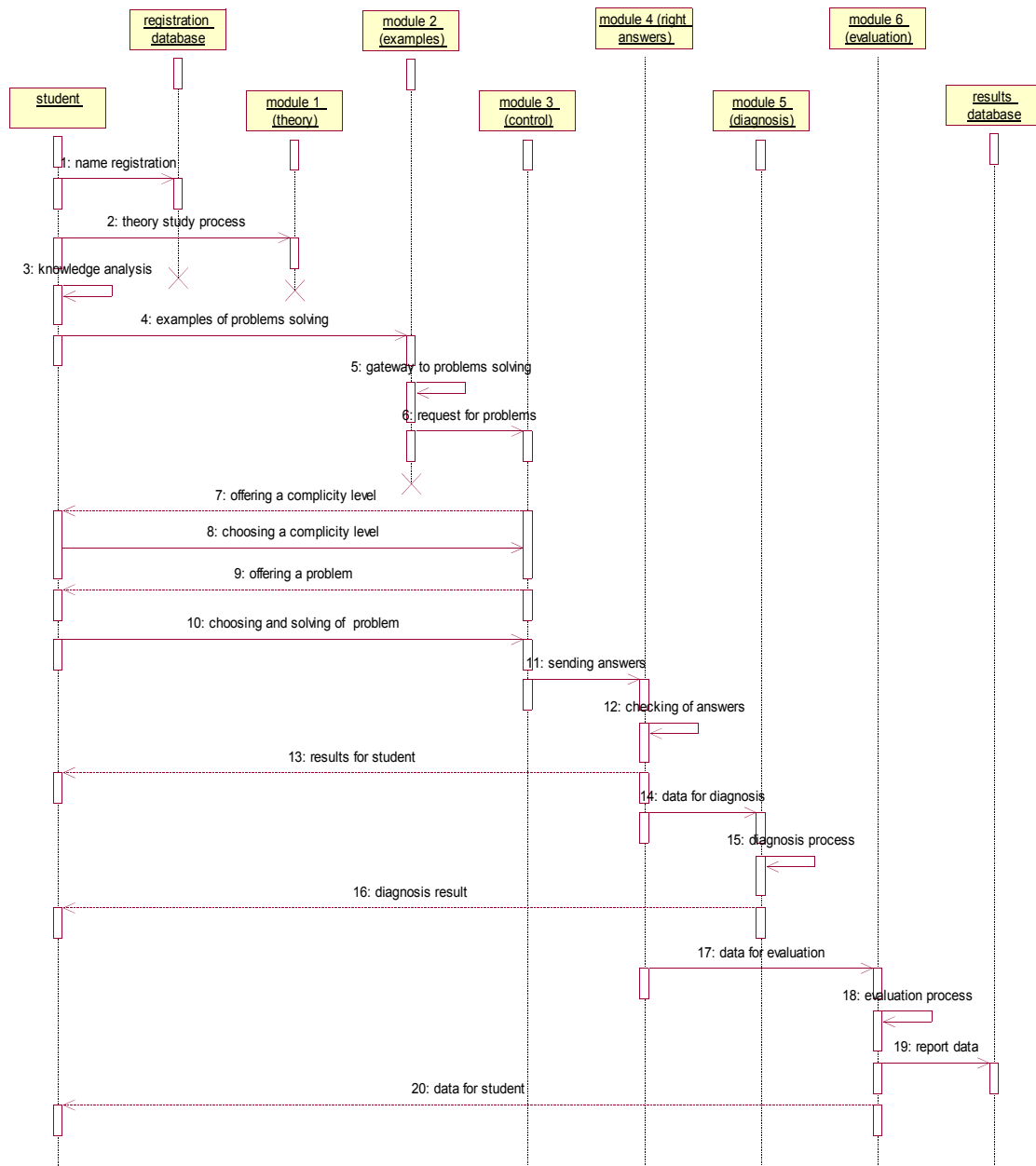


Figure 3. Sequence Diagram of the Interaction of the Modules of an E-Learning System

The presented diagrams reflect the way the system modules communicate in dynamics and show the interaction of classes and transmitted messages for a better understanding of user's requirements. A part of the transition state diagram (the receipt of a student's answer and the comparison with the correct variant) is shown in Figure 4 and forms a complete and clear understanding of the mechanism of a finite number of states.

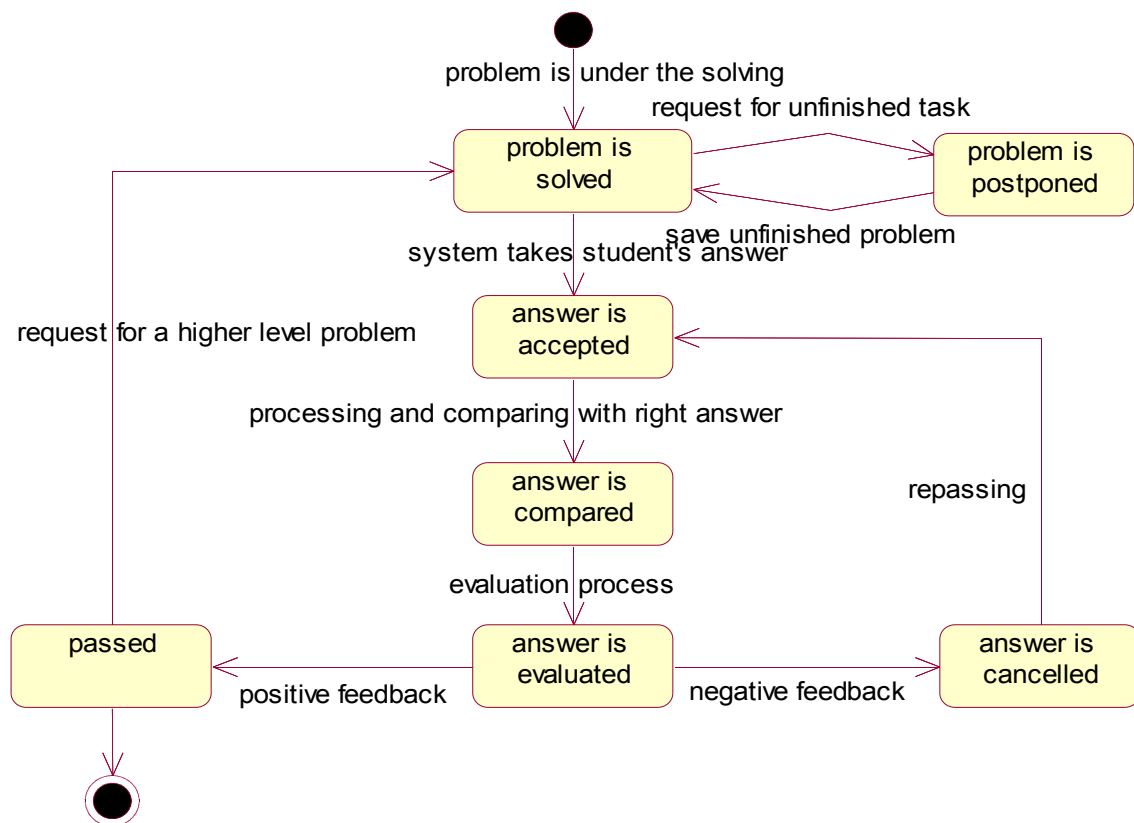


Figure 4. Part of a State Diagram of an E-Learning System

The state diagram which contains three types of elements – initial states, final states and transitions - shows the transition between the states, not as part of the processes performed by the system, but as the only possible state of changes associated with these processes. The transition diagram helps to understand the expected behavior of the system, check whether all the necessary state transitions are described correctly and completely as for the functional requirements.

The presented models are different diagrams related to each other and with their common data definitions in the data dictionary that allows you to maintain a model in a consistent state, and in accordance with the functional and user's requirements for filling software content. A visual modeling intelligent e-learning system is a way of perception of the problem with the help of visual abstractions, concepts and reproduces the behavior of objects in the subject area.

Thus, the problems of analysis and modeling requirements in the development of e-learning systems with feedback during the business analysts' training have been defined, the subject area requirements have been determined; the choice of CASE-tools to create models for the generated domain has been proved. The visual model of functional requirements and the requirements of users in the form of graphs showing the development of behavior and logical structure of the system, which allows to identify incorrect, inconsistent, missing or redundant requirements has been developed. Visual models of user and functional requirements of the system of e-learning have been developed on the basis of CASE- tools and provide information support, explaining the events and the corresponding response. It is the interaction of modules at a higher level of understanding than the text

specification requirements give, but it cannot completely replace the technical task in natural language, as each type of modeling has its advantages and its limitations. E-learning system modelling is an effective procedure to reflect changes in functional and user's requirements which resulted in the optimal sequence determined by its design which improves the quality of the development system by the criterion of coverage and treatment of the considered types of claims and accordingly improve the efficiency of its use by students.

On the Application of Knowledge-Oriented Social Networking Web-Sites on the Internet in Competitive Specialists' Training

With the rapid development of information technology and weave of all spheres of social life with the Internet technologies the importance of research in the field of education and improving the training quality has enhanced. Given the rapid pace of innovation in most organizations, a modern person must not only know and successfully fulfill all his or her responsibilities, but also to be capable to permanent self-learning.

The employees who have been working in organizations for a long time, have the knowledge necessary for young professionals, but unfortunately most of them do not pay enough time to self-education and further researches. Thus valuable knowledge of experts (their intellectual capital) is a load of knowledge, which is not being improved over time, becomes obsolete and non-relevant. Young professionals have the opposite problem, with basic skills and desire to learn they cannot get work experience and acquire the necessary social capital before the actual start of work in the organization.

To effectively solve this problem it is advisable to use modern methods and technologies of knowledge management, in particular the use of knowledge-oriented social networking sites on the Internet as an essential element of the educational process. For their effective functioning, consider the following factors:

1. Menu and structure of a social network should be user-oriented (the method of systemological classification analysis is strongly recommended).
2. A framework of motivational techniques to engage the students, academics and experts into work should be developed.
3. The interaction between the specialists should be based on interest in communication and solving the set problems.
4. The system of interaction between students of junior and senior courses, graduates of previous years and teachers should be implemented at the universities.
5. The use of the proposed social networks in the work of practice communities is reasonable.

The use of a social network on the Internet to solve the problems of education and its integration into the real economy will:

- Increase the competitiveness of graduates, adapting their skills to the needs of business and industry of their activities.
- Improve the social capital of graduates and employees of organizations.
- Find new ideas and valuable employees for employers.
- Accelerate the exchange of knowledge and the solution of typical problems in the subject area as a result.
- Quickly introduce modern technology in the educational process and the work of the organization.
- Find sources of funding for researches and the practical implementation of scientific developments.

- Consolidate the knowledge and efforts of specialists from different regions and countries to meet the challenges of science, business and education, thus ensuring the sustainable development of the state and society.

Implementation of the proposed project on the practice will consolidate the efforts of specialists of subject areas and to further analyze the efficiency and quality of the training system in the Internet space with a view to making recommendations for its improvement.

On the Transformation of an Organization into a Learning Organization

One of the priorities of any organization in a competitive environment is the development and training of employees, the permanent being of the staff at the same level is not profitable for the company. The most appropriate form of development is a learning organization.

The emergence of a learning organization is seen as a natural reaction to the conditions of the transition from a knowledge-based economy to a knowledge economy, in which there is a need for a new organizational form, structure and instruments of coordination. A unique distinguishing feature of such an organization is not just staff combined with his experience and knowledge, but the culture of organizational learning and development, and staff perceives work as a process open to continuous improvement.

In learning organizations people are focused on the development of their potential to achieve common results; a new system of thought is formed, collective desire is expressed freely, and people are continuously learning how to learn together. A learning organization is an organization that creates, acquires, preserves and transmits knowledge. It is able to successfully change the type of their behavior to reflect new knowledge or projects that support the learning process in all fields, promoting training on how to learn [10].

The idea of a learning organization, continuous self-improvement and self-learning of an employee most closely matches the needs of the development of IT-sphere in modern conditions.

The concept of the learning organization becomes widespread in the knowledge management theory and practice. The knowledge production process is reduced to the study and development of latent characteristics and abilities of their staff by continuous training during daily activities.

The key provision of the concept of the learning organization is that any organization can be called learning, the only question is how big and recognized her ability to learn and self-development is. That is why non-learning organizations do not exist, and at the same time, there is no organization that can be called learning ones understanding this kind of a high degree of perfection and excellence. Thus, a learning organization is not a static condition which the company will sooner or later reach, but the process of continuous learning, self-improvement and development.

Learning organization is a team united by a common vision and values of employees who are able to develop and improve the work process (and the result of this process), relations arising in it, as well as their own understanding of the situation through constant feedback from your colleagues, clients, partners, heads, namely from the internal and external environment [10].

Distinctive features of a learning organization are system thinking, which allows to make holistic view of the processes and phenomena and their most effective change; encouraging employees to develop, determines the identification of values workers and their consistency with the objectives of the organization and society; creating a common vision to reach the common goal [10].

It should be pointed out that a single model of a learning organization does not exist - it is rather a philosophical attitude to the fact that there is such an organization and there are some certain roles of its employees. In a learning organization, each of them can participate in the identification and resolution of problems based on knowledge, intelligence and lessons organizational values and culture, and that allows the company to experiment continuously, to use the newly emerging opportunities. In this case, turn the heads of his subordinates in the partners, and their main task is to create conditions for the development of learning abilities and make fuller use of energy and knowledge workers.

Education in some ways improves the efficiency of the company, but no significant impact on the business process in the company and its financial performance. The consequence of this is the lack of loyalty of the company and its employees. In a stable industrial environment these are the normal conditions of profitable functioning of "economic companies." But these companies are unlikely going to be the leaders in the global market and allow our country to make a breakthrough to the "knowledge economy".

We consider the transformation of a learning organization within the information technology industry, which is one of the fastest growing industries. According to Datamonitor Research Company, the information technologies industry consists of two main markets:

- 1) data processing and outsourcing services (divided into electronic data processing, IT outsourcing and business process outsourcing services);
- 2) IT consulting services (service integration and development, support for hardware and software infrastructure and proper IT consulting). Market BPO brings 43% (239.3 billion dollars) of industry profits, giving the way to IT outsourcing (50.4%, 280.3 billion dollars). The industry of data processing services and outsourcing has steadily grown by an average of 11.1% per year companies for 5 years.

Today information technologies can make a decisive contribution to strengthening the relationship between labour productivity growth, investment and employment. New types of services distributed over networks are able to create a lot of jobs, and that is confirmed by the practice of recent years.

A number of stages in the transformation of a learning organization with systemological analysis employed have been investigated:

1. Formation of a highly qualified staff as an intellectual and professional basis of a company.
2. Development and implementation of quality system standards to monitor their compliance to them.
3. Formation of the company's development strategy, including staff development.
4. The introduction of instant availability of necessary information.
5. Improvement of non-financial motivation of employees and the strengthening of staff.

Effective training of an organization is the constant and purposeful process of obtaining a competitive advantage being held and quality as a key resource of the enterprise which allows to realize the concept of a learning organization through its constant transformation.

Findings

The paper brings up some solutions proposed to improve the training of highly qualified analysts for consolidated information having noospheric system thinking, new information and organizational culture and use of innovative knowledge-oriented tools and methods in their work (for instance, the systemological classification analysis) aimed at increasing the intellectual capital and sustainable development of the organization.

Features of high-quality training of consolidated information analysts define a key role of the teacher in teaching students due to the need to analyze the qualitative aspect of the content of information and knowledge, describing the studied subject area. E-learning system is an auxiliary tool in the training of highly qualified analysts and is considered to be an additional opportunity to provide an interactive media student. The proposed approach allows to partially implement the most significant part of the training – a feedback by which the gradual control of the learning process is to determine the place of error in some cases to recognize their causes, to form the recommendations, to address them and to improve the results.

The application of systemological cognitive methods and knowledge management is useful in various fields and can improve the problem-solving education and improve its quality. In some companies, as yet, the knowledge management system is a sphere of activity of the professionals working with staff and IT-departments (technical ability to provide knowledge management for the system activities). The potential of a knowledge management system is not used to the full; knowledge, human factors, cognitive features of employees are not being taken into account to the proper extent. Therefore, the actual problem of improving knowledge management in organizations and the formation of learning organizations that are designed to perform analytics experts - analysts for consolidated information [11, 12].

The authors express their sincere gratitude to the Krassimir Markov for his continued long-term support and useful and creative discussion of the new results aquired.

Acknowledgements

The paper is partially financed by the project ITHEA XXI of the Institute of Information Theories and Applications FOI ITHEA and the Consortium FOI Bulgaria (www.ithea.org, www.foibg.com).

References

1. Бондаренко М.Ф., Соловьева Е.А, Маторин С.И. Информационному обществу – профессионалов в области информации: Business Intelligence, Knowledge Management на службе государства - [Электронный ресурс] - Режим доступа: [www/URL: http://it2b.ru/it2b2.view3.page123.html](http://it2b.ru/it2b2.view3.page123.html) – Загл. с экрана.
2. Бондаренко М.Ф., Соловьева Е.А, Маторин С.И. Підготовка професіоналів у галузі інформації для державної служби України. - Вісник державної служби України №1. 2004
3. Соловьева Е.А. Естественная классификация: системологические основания / Е.А. Соловьева. [Текст] – Харьков: ХНУРЭ, 1999. – 222с.
4. Solovyova K. Mathematical and Systemological Foundations of Natural Classification Automatic Document and Mathematical Linguistics. Allertion Press. Inc., New York, 1997, V. 30, No. 4.
5. M. Bondarenko, S. Matorin and V. Matorin. Solovyova K. Embedding Conceptual Models of Knowledge in Technology of Systemological Analysis IFIP WG8.3 International Conference on Creativity and Innovation in Decision Making and Decision Support, vol. 2, 2006, London, UK.
6. Важность управления знаниями организации [Электронный ресурс]. – Режим доступа: [www/URL: http://www.businessbasis.ru/vazhnost-upravleniya-znaniyami-organizatsii/](http://www.businessbasis.ru/vazhnost-upravleniya-znaniyami-organizatsii/) – 25.04.14. – Загл. с экрана.
7. Pishchukhina O.A. Analytical support of requirements development for intelligent e-learning systems. - *Радіоелектроніка, інформатика, управління.*- Запоріжжя:ЗНТУ. - 2013. - №.2(29). - С.136-139.
8. Бондаренко М.Ф., Соловьева Е.А., Маторин С.И., Ельчанинов Д.Б. Объектная технология моделирования информационных и организационных систем: учеб.пособие. –Харьков:ХНУРЭ,2005.–160с.

9. Вигерс К. Разработка требований к программному обеспечению.: Пер. с англ.: - М.: Издательско-торговый дом «Русская редакция», 2004. - 576 с.
10. «Пятая дисциплина. Искусство и практика самообучающихся организаций», Питер Сендж, Харьков 2006г., - 384с.
11. Mikhail Bondarenko, Nikolay Slipchenko, Kateryna Solovyova, Viktoriia Bobrovska, Andrey Danilov. Systemological Classification Analysis In Conceptual Knowledge Modeling. // Information Book Series «INFORMATION SCIENCE & COMPUTING». Supplement to the International Journal «INFORMATION TECHNOLOGIES & KNOWLEDGE» 2010. Варна, Болгария. Pp. 169-176.
12. Mikhail Bondarenko, Nikolay Slipchenko, Kateryna Solovyova, Andriy Danylov, Ruslan Kovalchuk, Shcurenko Irina Conceptual Knowledge Modeling and Systematization on the Basis of Natural Classification / International Journal "Information Theories and Applications", Vol. 18, Number 2, 2011 Варна, Болгария. Pp. 151-171.

Authors' Information



Kateryna Solovyova - Chief of Social Informatics Department and Knowledge Management Center, Professor, Doctor of Technical Sciences, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: gt_ekasolo@yahoo.com, si@kture.kharkov.ua.

Major Fields of Scientific Research: Knowledge Classification, Systematization, Elicitation, Acquisition and Modeling, Knowledge Management, Systemological Analysis, Ontological Engineering, Decision Making Support, Knowledge-Based Systems and Technologies, Artificial Intelligence, Business Intelligence, Modern (e-) Learning, Learning Organization, Competitive Intelligence, Cognitive Modeling, Intellectual Capital

Andrey Danilov – Social Informatics Department, Kharkov National University of Radio Electronics, senior Lecturer, Lenin Ave.,14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua, Skil06@ukr.net.

Major Fields of Scientific Research: Social Networks, Ontological Engineering, Competitive Intelligence, Decision Making, Intelligence Technologies, Knowledge Research and Application, Knowledge Management, (e-) Learning, Artificial Intelligence, Systemological Analysis, Social Capital.

Olga Pishchukhina – Associate Professor, Social Informatics Department, Ph.D., Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: pishchukhina@gmail.com, si@kture.kharkov.ua.

Major Fields of Scientific Research: E-learning Systems, Artificial Intelligence, Decision Making, Uncertain Systems.

Yuliia Panasovska – graduate of Social Informatics Department, Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua.
Major Fields of Scientific Research: Ontological Engineering, Intelligence Technologies, Knowledge Research and Application, Knowledge Management, Systemological Analysis.

Olga Lashyna - Social Informatics Department, Kharkov National University of Radio Electronics, Assistant, Head of the Multimedia Laboratory, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua, ambologera@yandex.ru.

Major Fields of Scientific Research: Learning organizations, Technologies for Information

Management, Information Systems, Knowledge Management, Systemological Analysis.

Maria Kozlova - Social Informatics Department, Kharkov National University of Radio Electronics, Departmental Assistant, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: koslovamari@gmail.com, si@kture.kharkov.ua.

Major Fields of Scientific Research: Semantics, Knowledge Management.

Maryna Khoroshenko – MA Student of Social Informatics Department, Kharkiv National University of Radio Electronics, Master's degree student, Lenin Ave., 14, Kharkiv, 61166, Ukraine; e-mail: maryna.khoroshenko@gmail.com, si@kture.kharkov.ua.

Major Fields of Scientific Research: Technologies for Information Management, Systemological Analysis, Testing Process, Manual Testing.

Maksym Voroniy - Associate Professor, Social Informatics Department, Ph.D., Kharkov National University of Radio Electronics, Lenin Ave., 14, Kharkov, 61166, Ukraine; e-mail: si@kture.kharkov.ua.

Major Fields of Scientific Research: Requirements Analysis, Knowledge Classification, and Modeling, Knowledge Management, Systemological Analysis, Ontological Engineering, Knowledge-Based Systems and Technologies, Artificial Intelligence, Business Intelligence, Modern (e-) Learning, Competitive Intelligence, Cognitive Modeling

PHOTOGRAMMETRIC TECHNIQUE FOR THE DETERMINATION OF THE MOMENT IN TIME AND THE PLACE OF THE LOSS OF STABILITY DURING THE PLASTIC DEFORMATION OF SHEET METAL

Spilko Hubenov, Georgi Stoilov

Abstract: *An experimental study establishing a technique for the contactless optical determination of the local plastic formability limit on the loss of stability of sheet metal has been conducted. The possibilities for the recording of the time instant and the place of the localisation of the deformation based on the changes in the optical images of the surface have been examined. An optical criterion for the determination of the time instance and the place of the loss of stability derived through a correlation analysis of the images has been proposed. The experimentally obtained results have been presented and analysed.*

Keywords: *stable formability limit, loss of stability, photogrammetry, correlation analysis, digital image correlation.*

1. Introduction

Upon the implementation of sheet stamping operations with a variety of variable schemes can be observed an uneven distribution of stress, strain and strain rates in the volume of the metal during the entire technological process. The influence of a large array of factors greatly hampers the general summary of the formability and localisation of the plastic deformation in particular and is the reason for the lack of established unified methodologies for rational technological design of tools and technologies for sheet stamping.

For the manufacture of complex products with irregular shape with the realisation of combinations of various operations throughout the entire range of variation of the stress state, the use of marginal sustainable formability diagrams (MSFD) ensures the maximum usage of the plastic resource of the material. In these cases the registration of the exact moment of the beginning of the focused macrolocalisation is crucial, and given the associated technical difficulties, the possibility for the creation of the diagrams and the practical application of the existing criteria are insufficiently studied [Генов, 2004]. In this regard of particular interest is the study of the various aspects of focused localisation that leads to the loss of stability of the plastic formability.

In the laboratories "Plasticity and Fracture of Materials" and "Mechanical Testing of Materials" at the "Materials Science and Technology of Metals" Department at the Technical University – Sofia has been conducted years of research on the deformability of sheet metal [Генов, 2007; Stoilov, 2009; Пригоровский, 1983]. One of their main goals is the study of the options for control and management of the technological process and facilitation of the creation of marginal formability diagrams (MFD) and MSFD. The efforts are focused on the establishment of a simple methodology for determination of the moment in time and place of the loss of stability, based on contactless measuring of the deformations.

There are a number of practically established traditional contactless experimental methods for the study of displacements and deformations that differ in the complexity of the preparation and the study itself, the

experimental apparatus used, the nature of interaction with the object and the material from which it is made, the sensitivity, the accuracy, the reproducibility and the comparability of the results [Пригоровский, 1983; Heymann, 1986; Khan, 2001]. They are based on different optical effects or on the optical recording of images of real objects.

The experiments conducted so far for the determination of the moment in time and place of the loss of stability by optical means has used as optical criterion for the determination of the moment of the loss of stability of the sheet metal the integral intensity of the reflected light from the metal surface [Хубенов, 2008]. Its variation as a result of the changes during the course of the deformation process in the topography of the surface, when registered properly, can be used as an indicator of the occurrence of the focused localisation.

This work presents a new series of studies carried out jointly with the "Mechanics of the Deformable Solid Bodies" Division at the Institute of Mechanics at the Bulgarian Academy of Science, related to the contactless determination of the marginal sustainable formability of sheet metal based on an optical criterion for the loss of stability. They include the capture of photographic images of the surface with subsequent correlation analysis through the use of software for digital processing and analysis of the images recorded with the digital camera.

2. Methodologies

The sustainability of the formability of the sheet metal is a key condition for obtaining high quality stamped items. One of the most interesting, practically important and not fully explored to date issue that is inextricably linked to the study of plastic flow is its tendency to spontaneous localization, which in the limiting case (when exceeding a given critical value of deformation) leads to the coexistence in the material of two types of areas – involved and not involved in the deformation process [Рыбин, 1986].

One of the most important directions in the studies of the processes of plastic deformation is the determination of the marginal formability. Two approaches are applied for this purpose, based respectively on the plastic resource and the loss of sustainability of the deformation process. The former one determines the marginal formability by the disruption ensuing as a result of the depletion of the resource in terms of a given mechanical deformation scheme. In the latter approach the formability is limited by certain criteria for the loss of stability of the process and the associated localisation of the deformation.

There are two known macroscopic manifestations of the localisation mechanism after the loss of sustainability of the deformation process, in which the previously steady deformation localises in a given area. Accordingly, there is an initial scattered (diffused) deformation, followed by a concentrated local deformation. It is assumed [Hora, 2003] that the type of the localisation depends on the nature of the deformation state (as defined by the parameter of the deformation state $\beta = \varepsilon_1 / \varepsilon_2$, $\varepsilon_1 > \varepsilon_2$): in the range of $\beta \in [0; +1]$ occurs as scattered deformation, while in the range of $\beta \in [-1; 0]$ occurs as concentrated local deformation.

The approaches for the determination of the moment of the loss of stability are based on different sets of criteria: force criteria; criteria for limiting stress; criteria based on patterns of damage; bifurcation criteria. The classic force criterion for the loss of stability, the maximum force criterion (MFC), is proposed by Swift and is formulated as $dF_x = 0$ or $d\sigma_1 / d\varepsilon_1 \geq \sigma_1$, where σ_1 and $d\varepsilon_1$ are respectively the maximum principal stress and strain [Swift, 1952]. It applies to homogeneous state of stress (constant normal stress and neglected tangential stress) and determines the moment in time of the occurrence of the diffused localisation of the deformation.

The modified force criterion (MMFC) is formulated by Hora and Tong for three-dimensional stress state [Hora, 1994]. It leads to the emergence of an additional effect of hardenability which displaces the start of the localisation to higher values of deformation. The additional hardenability is represented as β and is not assumed as constant, but is included as a variable parameter in the formula that is describing the MMFC:

$$\delta\sigma_1 / \delta\varepsilon_1 + \delta\sigma_1 / \delta\beta \times \delta\beta / \delta\varepsilon_1 \geq \sigma_1 \quad (1)$$

Even more accurate is the enhanced modified force criterion (EMMFC), since its formulation includes not only the additional hardenability, but also takes into account the thickness of the material and the correlated probability for propagation of the crack.

The field of experimental mechanics uses as methods for the measurement of the deformation of solid bodies different variants of the so-called close range photogrammetry – analog, analytical and digital [Zeng, 1992].

In recent times the importance of digital photogrammetry is constantly growing. It records a point on the surface of the object on consistently captured analog or digital images (photographs) with a high resolution camera. Its spatial coordinates can be derived from those of the corresponding points in the plane of the images. The geometric data is subjected, including in real time, to mathematical processing through the use of specialized software [Rastogi, 1999]. Among the distinctive features of digital photogrammetry is the absence of specific mechanical modules, mechanically induced inaccuracies in the measurement and calibration, the geometrically stable storage of data, the possibility for combination of different sensors and the reconciliation of recording, processing, editing, storing and managing of the data in a given system [Ламбин, 1987].

Of increasing importance for the deformation studies are the modern automated systems for optical 3-D analysis based on photogrammetry, which combine into one all the stages of the information process. The optical systems ARAMIS, AutoGrid, ARGUS use photogrammetric algorithms to recognise the positions of the points of intersection of the reference grid placed on the object in space and calculate their movements from their initial positions. The results converted into deformations are presented most often in the form of graphics or 3-D visualizations [Feldmann, 2006; www.gom.com, 2009]. The accuracy of the measurement of the automated systems is dependent on the degree of heterogeneity and the scale of the deformation. The automated optical systems are most effective when seeking maximum data density with minimum base length of measurement. The deterrent for their mass practical application is their high price.

The data processing in photogrammetric methods is most commonly performed through the use of digital images correlation analysis (DIC). The digital (discrete) image represents a matrix of values that correspond to the average integral value of the vicinity around a given discrete point. The correlation methods compare the sequentially recorded images, while the evaluation and the measurement of the deformations are based on the creation of the fields of the vectors of displacement of the individual elementary areas (dots) and the subsequent determination of the components of the deformation [Напрушкин, 2008]. There are two approaches used for the determination of the marginal formability in terms of the localisation of the deformation: space dependent and time dependent. The various methods differ in the dimensionality of the determined deformations.

The 0D method is not distinguished with high accuracy. It does not consider the focused localisation and therefore the values of the deformations are adjusted by a factor based on empirical assumptions.

The 1D method determines the marginal formability in pre-selected sections using mathematical criteria. The method is based on the assumption that after the start of the localisation the deformations outside of its area remain constant, allowing the assessment to be carried out just before or just at the time of the occurrence of the

crack [ISO/DIS 12004-2, 2008]. One version of the classic sections based method is the time dependent one with its two varieties: gradient and differential. The former takes into account not only the distribution of the deformations, but also the temporal variation of their gradient. An advantage of this method is the ability to automate the calculation procedure [Eberle, 2008]. The differential sections based method determines the moment of the loss of stability on the basis of the temporal variation of the differentials of the main deformations.

The 2D method, unlike the sections based one, depending on the moment in time assesses not a single section, but all of the points from the raster. Based on the global distribution of the coordinates of all the images and under the assumption of an idealised grid, the deformations can be calculated through the function of Lagrange [Volk, 2006].

3. Experimental Procedure

For the testing purposes of the deformability of sheet metal an experimental rig is constructed, consisting of a hydraulic unit, a tool for carrying out the deformation process (Figure 1) and an electronic unit for recording and control. The latter includes a pressure transducer CPT 2500, 0/160 bar with adapter CPA 2500 and software USBRAM DDR2 by the company WIKA Alexander Wiegand GmbH & Co KG. This allows the moment of the loss of stability of the deformation process under the classic force criterion to be determined based on the analysis of the variation in the pressure of the fluid through the specialized software USBsoft 2500.

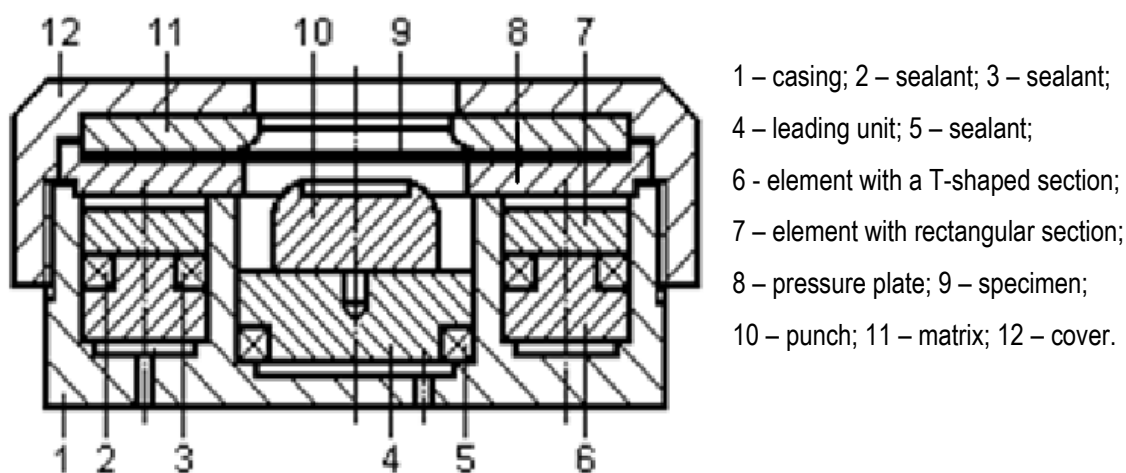


Figure 1. Deformation tool.

There is a possibility for the realisation of experiments using the two practically established methods – of Marciniak [Marciniak, 1967] and Nakazima [Nakazima, 1968], and in accordance with the respective punches (cylindrical with flat end or hemispherical) optical schemes with one or two cameras are used for the recording of the optical images. The data obtained through the rig can serve to build MFD – at the occurrence of a visible crack and MSFD – at the occurrence of concentrated macrolocalisation. There is also the option for direct control of the deformation process through the use of additional hardware and software means.

For carrying out the present study an experimental set has been constructed (Figure 2), comprising the aforesaid rig, an intense incoherent light source and a fast digital camera with various lenses and filters.

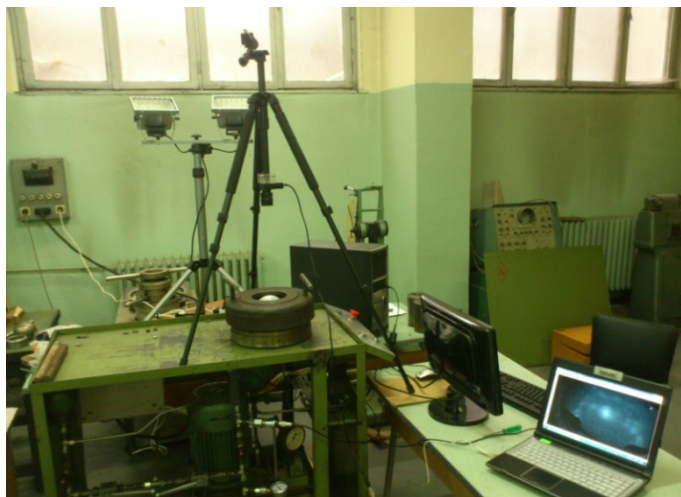


Figure 2. Experimental set.

The conducted experiments aim to establish the applicability of an optical criterion for the determination of the moment of the occurrence of a concentrated localisation. In line with this task a methodology has been developed, based on the photogrammetric determination of the deformations. Upon juxtaposition of the results with the data recorded under a force criterion for the loss of stability of the deformation process, it is possible to set an empirically established threshold value that determines the emergence of concentrated localisation.

During the experiments series of specimen of sheet metal are deformed at a constant rate, while the stress is discontinued following the final destruction of the specimen. Simultaneous with the start of the deformation process is launched the parallel capturing of sequences of optical images of the surface of the sheet metal in the area of deformation and the registration of the temporal variation of the pressure over the working fluid.

Since the study of formability through the use of coordinate grids is characterised by certain defects leading to a difficult and sometimes subjective assessment of the displacement of the nodes during the deformation process, its use in this case cannot guarantee the accuracy and reproducibility of the results. Therefore another option is preferred in order to quickly and inexpensively obtains the stochastic raster required for the photogrammetry that covers the specified requirements. Following the cleansing of the surface of the specimen it is sprayed uniformly with a coloring agent in the form of an aerosol. The fine colored dots serve to create a unique image in each area. This facilitates the registration of the vector field of the local displacements of a sufficiently large number of points from the area of deformation.

The deformations that correspond to the moment of the loss of stability are determined in three steps: specifying the sections containing the presumed localisation of the deformation; establishing the moment of the occurrence of the focused localisation by means of a criterion set as a threshold value that allows the determination of the internal limitation values for the creation of a distribution curve of the deformation; determining the external limitation values for which for the first time there is a reduction of the curvature of the established distribution.

The setting of the threshold value for the displacement is done in accordance with a suitable force criterion based on the parallel registration of the temporal variation of the pressure over the fluid. In this case the use of the classical criterion for the occurrence of a diffused localisation is justified in view of the precision of the available sensor. The so determined moment of the loss of stability, fixed with some delay following the attainment of the

maximum pressure value (allowing its random fluctuations to be neglected), can be considered as the start of the concentrated localisation.

The experiments performed during this study are using the Marciniak method. Round specimen are used with a diameter of 206 mm with bilateral notches with different radius r_m : 0; 43; 53; 60.5; 68; 75.5; 83; 93 mm. Each series of specimen (for a given radius of the bilateral notches) includes specimen with axes of symmetry at 0° , 45° and 90° to the direction of rolling. The numbering of the series is in ascending order from the smaller to the larger radii. The specimen tested are comprised of sheet steel 08kp with thickness $s = 0.7$ mm and indicators: $R_e = 190$ MPa, $R_m = 600$ MPa and a coefficient of deformation hardenability $n = 0.370$.

A camera UI-1490ME with a USB 2.0 interface and a CMOS sensor by the company Aptina with resolution of 10 megapixels (3840x2760) is used for capturing the images. It is fitted on a tripod above the specimen and the image is centered at the specimen. Thus, when capturing a specimen with a size of 200 mm is obtained a resolution of the camera of up to 50 μm . The camera is positioned so that the field of the lens covers the whole area of deformation in the zone of the specimen that is free of stress. For the series from 5 to 8, in which the stress state is of a biaxial tensile, the area of the expected thinning due to the localisation and respectively of the breaking, is at the center of the area of deformation under optimum conditions of deformation or in the vicinity of the radius of curvature under degraded conditions of friction. This requires the registered image to cover the entire observed surface and the capturing is done at a lower speed in order to ensure a sufficiently high resolution of the entire field of the image. Due to the limited bandwidth of the interface and the built-in computing power in the camera, the experiments with the capturing of the entire surface of the specimen are with a maximum speed of 2.5 frames per second. This requires the use of a lower capturing speed in order to ensure a sufficiently high resolution of the entire field of the image. Under a strained state of tension-stress (for series from 1 to 4) the thinning and breaking are expected to occur in the central area of the specimen, which allows for the increase in the speed of capturing to 6 frames per second at the expense of a reduction of the scope of the frames while preserving the image quality. Used is a lens with a depth of focus that is sufficient for operation over the entire range of variation of the distance during the deformation process due to the relative displacement of the surface of the specimen (the image plane).

In order to increase the speed of measurement the amount of data is reduced by capturing only the part of the specimen, in which the localisation of the deformation is expected to occur. For example, for the narrowest specimen (from series 1 and 2) a rectangle encompassing the central region for width and along their whole length is used, in order to determine with sufficient precision the correlation between the deformation in this area and the other areas. The analysis of the results can be based mainly on two orthogonal cross sections passing through the geometric center of the specimen, one of which is parallel to the notches of the specimen, meaning in the direction of the maximum principal stress σ_1 .

Software has been developed for the mathematical processing of the results based on the correlation analysis of the optical images using the classical and the gradient sections based methods. For the calculation of the deformation the displacement of the points from the raster over the entire surface of the specimen is monitored.

Pretreatment is performed following the registration of the input image in order to correct the distortions of the input converters. Digital filtering is then used in order to eliminate the impact of the changes in lighting. This is followed by a reduction of the information in order to reach a compressed representation comprising only of the essential features for the processing.

The algorithm performs segmentation of the discretisation grid by dividing it in correlation subareas. In order to ensure the successful processing of the results of the photogrammetric measurements through correlation analysis it is necessary that every correlation subarea has a unique image, provided in this case by the fine raster structure of colored dots. The achievement of high resolution requires good autocorrelation coefficients for the correlation subareas over the entire observed surface. A size for the subareas of 64x64 pixels is selected for this case as a compromise between the requirement for greater measurement accuracy and the reasonable computational time. This results in the stable detection of the correlation maximums at a good resolution for the grid of correlation subareas. Their number is 60 along the length of the specimen and ranges from 16 to 43 through their width.

The processing of the results is performed through the use of a standard correlation formula of the type:

$$C_{fg}(u, v) = \frac{1}{N} \sum \sum f(i, j) g^*(i + u, j + v) \quad (2)$$

In which the functions f and g are the values of the intensity of the halftone images of the sought pattern and the deformed area in which said pattern is sought; u and v are the respective displacements along the axes X and Y ; $*$ is a complex conjugate of the argument; N is the amount of pixels; i and j are indices of summation.

The binarization enhances the correlation maximum and increases the area in which it can be found with an increase in the deformation.

The correlation subareas in the area of deformation, as well as the abovementioned main sections, are shown in Figure 3.

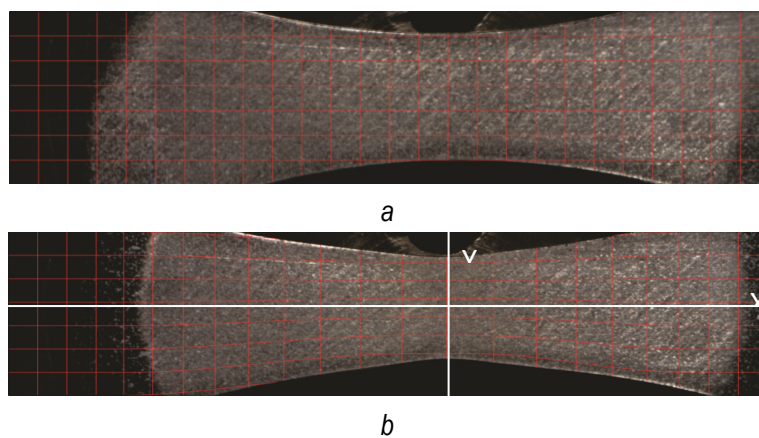


Figure 3. Correlation subareas on the specimen: a) in undeformed state; b) immediately prior to breaking following the emergence of a concentrated localization.

4. Results and discussion

The following figures contain some of the possibilities for the visualisation of the data obtained from the testing of the specimen from series 4 with longitudinal axis at an angle of 0° towards the direction of rolling – color distributions graphics, vector fields, distribution curves for sections.

Shown below are the results from two characteristic moments in time from the deformation process: following the emergence of a concentrated localisation and immediately prior to breaking. The data for the moment of breaking, uniquely determined by the distortion of the image plane, is taken from the last valid frame. Reverse

temporal synchronisation is performed between the captured series of optical images and the registered data on the temporal variation of the pressure. Based on the analysis of the results of preliminary experiments is established a threshold value to define the moment of the loss of stability. It is established that a pressure drop of 0.01 MPa after it has reached the maximum on the curve of its temporal variation eliminates its random fluctuations to the order of 0.005-0.007 MPa and provides adequate reproducibility of the results.

Figure 4 shows X-Y-cards of the displacement of the correlation subareas. The area with minimal displacements can be distinguished, in which later forms the area of focused localisation. It is evident, that its location remains unchanged over time, while in the surrounding areas the displacements increase with the distance from it and in the course of the deformation process.

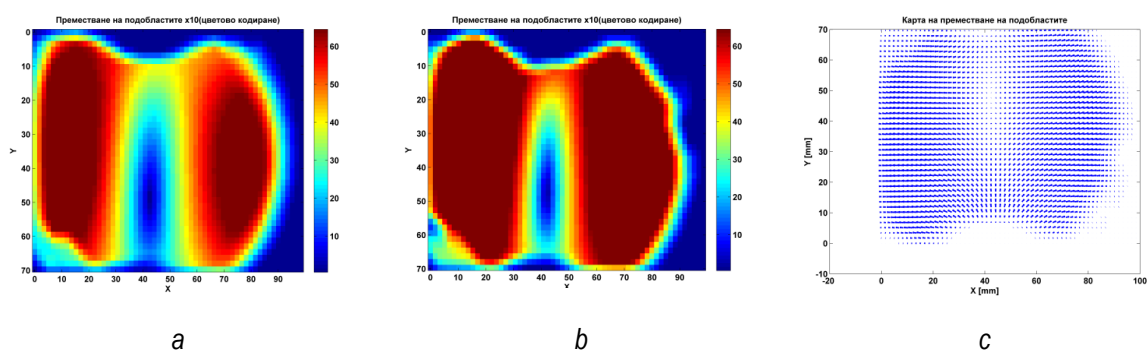


Figure 4. X-Y-maps: displacement of the correlation subareas in color code at the moment of the loss of stability (a) and immediately prior to breaking (b) and the vector fields of the displacements (c).

The distribution curves of the displacements along the axes X (a) and Y (b) are shown in Figure 5.

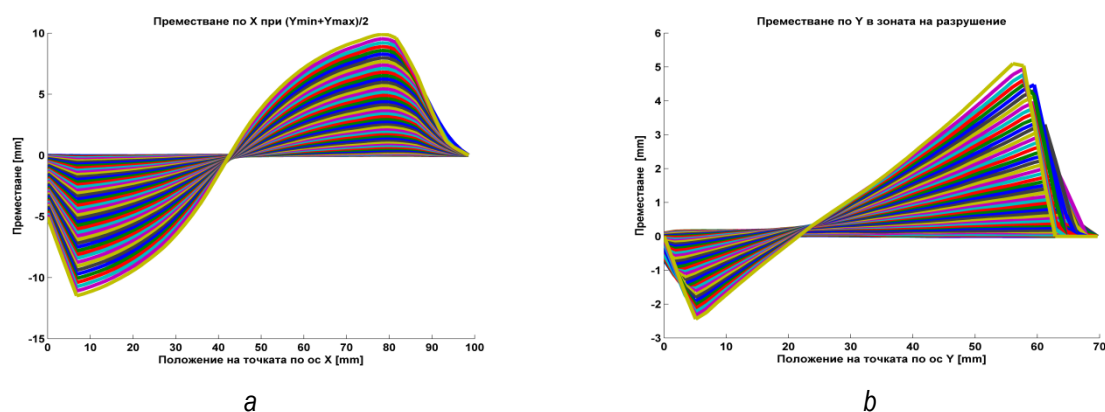


Figure 5. Displacement along the axes X (a) and Y (b) immediately prior to breaking.

The following figures show in graphical form the deformations in the correlation subareas marked in Figure 3, calculated using specialised software. The presented X-Y-cards reveal (Figure 6) that the location of the area of the localisation remains unchanged until the moment of breaking, while the deformation within it is gradually increasing with the attenuation of the process of plastic flow in the other areas. This is clearly illustrated by the

vector field of the deformations from the last valid frame before the breaking, in which the area of localisation is clearly outlined (Figure 6c).

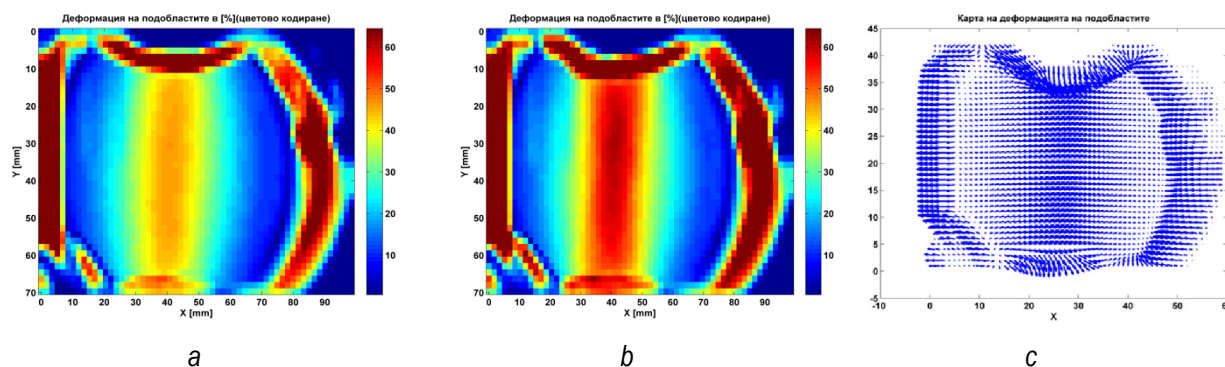


Figure 6. X-Y-maps: the deformations of the correlation subareas in color code at the moment of the loss of stability (a) and immediately prior to breaking (b) and the vector field of the deformations (c).

The distribution curves of the deformations along the axes X and Y, shown in the following Figure 7, confirm those observations.

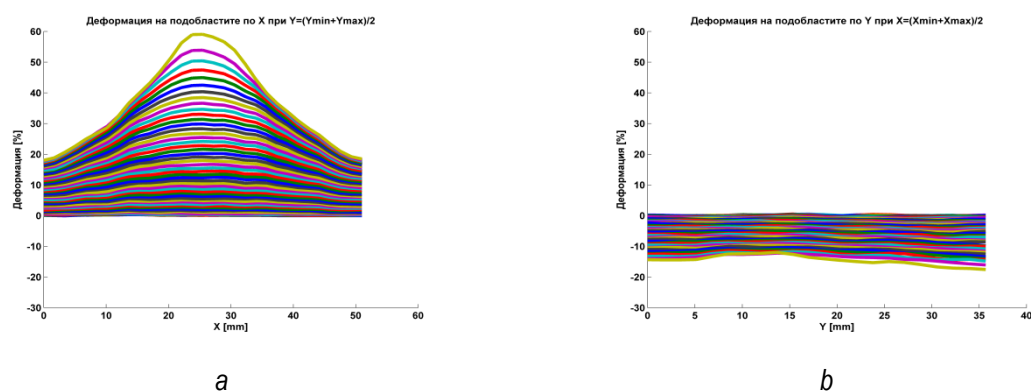


Figure 7. Distribution curves of the deformations.

Figure 8 shows the X-Y-cards of the grid with isolines and vectors of deformation, while Figure 9 shows the three-dimensional deformations in color code at the moment of the loss of stability and immediately prior to breaking throughout the entire observed surface of the specimen. These graphs reveal the most accurate presentation of the location and the shape of the area of the concentrated localization.

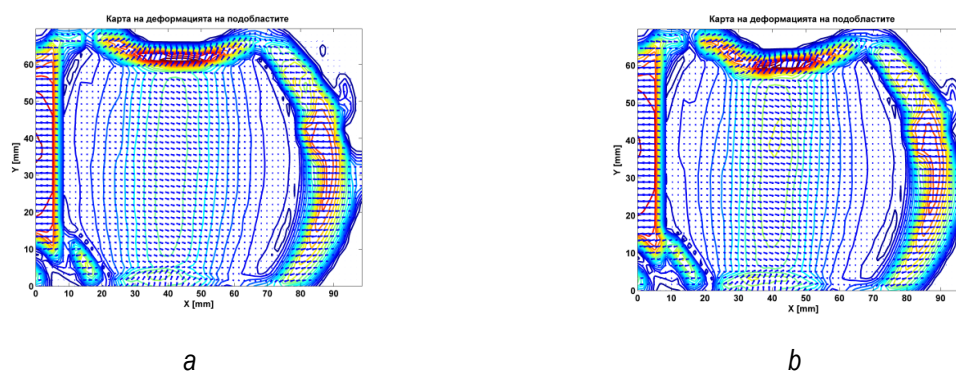


Figure 8. X-Y-cards of the grid with isolines and vectors of deformation at the moment of the loss of stability (a) and immediately prior to breaking (b).

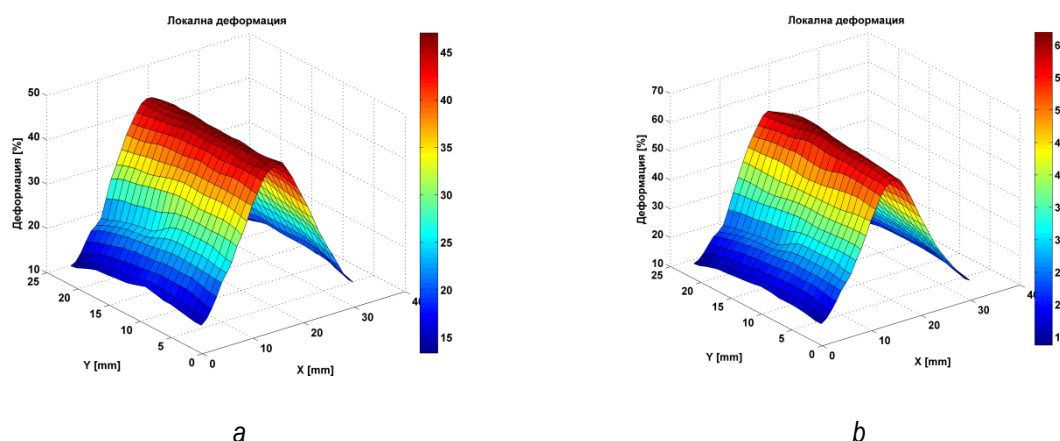


Figure 9. X-Y-maps of the three-dimensional distribution of the deformations in color code at the moment of the loss of stability (a) and immediately prior to breaking (b).

6. Acknowledgements

The studies presented in this article are a part of the scientific research project under contract № 141ПД0033-05 of the Scientific Research Sector at the Technical University – Sofia.

7. Bibliography

- [Генов, 2004] Генов Й., Р. Хесапчиева, В. Камбуров, Диаграми на пределното формоизменение на листов метал, Сборник доклади, I конференция с международно участие 4-6 ноември, София, 2004, 146-151.
- [Генов, 2007] Генов Й., В. Камбуров, А. Димитрова, Прогнозиране на пределните деформации при пластично формоизменение на листов метал, "Машини, технологии, материали", № 4-5, 2007, 13-15.
- [Ламбин, 1987] Ламбин Н.Е., П.И. Соловей, А.Ф. Даниленко, В.А. Деркач, Исследование точности определения статических деформаций фотоустройством Ф-1, Москва, Геодезические работы на подрабатываемых территориях, научные труды ВАГО, 1987, сс. 77-81.
- [Напряшкин, 2008] Напряшкин А.А., А.И. Солодушкин, В.С. Плешанов, В.В. Кибиткин, Автоматизация процесса получения и анализа оптических изображений для исследования механизмов деформации и разрушения материалов, Труды XV Всероссийской научно-методической конференции Телематика'2008, Том 1, с. 267.
- [Пригоровский, 1983] Пригоровский Н., Методы и средства определения полей деформаций и напряжений, Машиностроение, Москва, 1983.
- [Рыбин, 1986] Рыбин В., Большие пластические деформации и разрушение металлов. Москва, Metallurgiya, 1986.
- [Хубенов, 2008] Хубенов С., И. Русев, А. Шулев, Оптично определяне на момента на загуба на устойчивост при пластично формоизменение на листов метал, "Машиностроение и машинознание, бр. 4, год. III, кн. 1, Варна, 2008, 66-71.
- [Eberle, 2008] Eberle B., Entwicklung eines robusten numerischen Verfahrens zur automatischen Bestimmung experimenteller FLC-Kurven, Masterarbeit, IVP, ETH Zürich, 2008.
- [Feldmann, 2006] Feldmann, P., M. Schatz, Effective Evaluation of FLC-Tests with The Optical In-Process Strained Analysis System AutoGrid, Proceedings of the FLC Zurich, 15th – 16th March 2006, IVP, ETH Zurich, Switzerland.
- [Heymann, 1986] Heymann J., A. Lingener, Meßverfahren der experimentellen Mechanik, Springer-Verlag, Berlin, 1986.

- [Hora, 1994] Hora P., Tong L.: Prediction methods for ductile sheet metal failure using FE-simulation, Proc. of the IDDRG'94, Lisbon, Portugal, 1994.
- [Hora, 2003] Hora P., L. Tong, J. Reissner, Mathematical prediction of FLC using macroscopic instability criteria combined with micro structural crack propagation models, Int. Symposium on Plasticity, Quebec, Canada, 2003.
- [Khan, 2001] Khan A. S., X. Wang, Strain measurements and stress analysis, Prentice Hall, USA, 2001.
- [Marciniak, 1967] Marciniak Z., Kuczinski, K., Limit Strains in the process of stretch-forming sheet metal. Int. Journal of Mech. Science 9, 1967, pp. 609-620.
- [Nakazima, 1968] Nakazima, K., Kikuma, T., Asaku, K., Study on the formability of steel sheet. Yawata Technical Reports 264, 1968.
- [Rastogi, 1999] Rastogi P., Photomechanics, Topics in Applied Physics, Vol. 77, Springer. 1999.
- [Stoilov, 2009] Stoilov G., J. Genov, V. Kavardzhikov, J. Kovachev, A Software System for Determination of Forming Limit Diagrams, i.Tech 2009, Varna, Bulgaria, June-July 2009, International Book Series "Information Science and Computing", 11
- [Swift, 1952] Swift H., Plastic instability under plane stress, Journal of the Mechanics and Physics of Solids, Vol. 1, 1952, p. 187.
- [Volk, 2006] Volk W., New experimental and numerical approach in the evaluation of the FLD with the FE-method, Proc. of the FLC Zurich, Switzerland, pp. 26-30, 2006.
- [Zeng, 1992] Zeng Tan, Melin L., Magnusson C.: Application of an image processing technique in strain measurement in sheet metal forming, Journal of Processing Technology, 33 (1992) 3, p 299-310.
- [ISO/DIS 12004-2, 2008] ISO/DIS 12004-2: Metallic materials - Sheet and strip – Determination of forming limit curves - Part 2: Determination of forming limit curves in laboratory, 2008.
- [www.gom.com, 2009] Copyright © 2009 GOM mbH Rev. A (de) 21012009 www.gom.com.

Author's Information



Spilko Atanasov Hubenov – Technical University – Sofia, 8 Kliment Ohridski Blvd., Sofia-1756, Bulgaria, tel.: +359 2 965 2710, e-mail: s_hubenov@tu-sofia.bg

Major Fields of Scientific Research: Mechanical testing, nondestructive testing, formability analysis and technological deformability of sheet metal



Georgi Ivanov Stoilov – Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bontchev St., bl. 4, 1113 Sofia, Bulgaria, tel.: +359 2 979 64 62, e-mail: gstoilov@imbm.bas.bg

Major Fields of Scientific Research: Digital image correlation, digital holography, interferometry, material testing

TABLE OF CONTENT OF IJ IMA VOL.3, NO.1, 2014

Neural Network Approach to the Formation Models of Multiattribute Utility	
Stanislav Mikoni	3
Simulation of Weather in Agent Models	
Vitaliy Lytvynov, Artem Zadorozhnii	10
Information System of Forecasting Based on Combined Models with Time Series Clustering	
Alexander Berzlev	16
Semantic Net from Concepts as a Model of Student's Knowledge: How Stable are the Results of Experimental Study?	
Evgeny A. Eremin	24
Fuzzy Expected Value Model with Inspection Errors And Two Level of Trade Credit in One Replenishment Cycle	
Olha Yegorova	37
Shaping the Citation-Paper Rank Distributions: Beyond Hirsch's Model	
Vladimir Atanassov, Ekaterina Detcheva	53
Self-Citations Effect on Scientometric Indexes	
Vladimir Atanassov, Ekaterina Detcheva	68
The Creation of Strategy for Innovation Development of Socio-Economic Systems	
Vladimir Pankratov	84

TABLE OF CONTENT OF IJ IMA VOL.3, NO.2, 2014

<i>Index Matrices with Function-Type of Elements</i>	
Krassimir Atanassov	103
<i>Reduced Generalized Nets with Characteristics of the Places</i>	
Velin Andonov	113
<i>On The Algorithmic Aspect of the Modified Weighted Hausdorff Distance</i>	
Evgeniy Marinov	126
<i>Comparative Analysis of the Languages for Business Rules Specification</i>	
Krassimir Manev, Neli Maneva	136
<i>Method of Behavioral Software Models Synchronization</i>	
Elena Chebanyuk	147
<i>Informational Models of the Advanced Systems of Radioaccess to the Telecommunications Networks Comparison</i>	
Sergey Mikhailov, Sergey Shreyner	164
<i>IT-Communicators for Medicine</i>	

Ozar Mintser, Volodymyr Romanov, Igor Galelyuka, Oleksandr Voronenko.....	169
<i>Hierarchical Two Layers Control Commutator for Implementation of Fully Non-Conflict Schedule</i>	
Kiril Kolchakov, Vladimir Monov.....	175
<i>Application of Fiber-Optical Modulators as Measuring Devices in Bioinformatics Researches</i>	
Oleksandr Ryabtsov.....	182
<i>Method for Controlling State Channel Wireless Networks under a Priori Uncertainty</i>	
Sergey Zaitsev.....	186
<i>About Usage Possibility of Principles of Excitations Transmission by Nerve Fibers for Telecommunications</i>	
Galina Gayvoronska, Maxim Solomitsky.....	195

TABLE OF CONTENT OF IJ IMA VOL.3, NO.3, 2014

<i>Analysis Finitely Small Values in Computer Simulation</i>	
Oleksii Voloshyn, Vladimir Kudin.....	203
<i>A Study of Application of Neural Network TECHNIQUE on Software Repositories</i>	
Ana Maria Bautista, Tomas San Feliu.....	215
<i>Method of Domain Models Designing</i>	
Elena Chebanyuk.....	233
<i>Method for Video Shot Detection and Separation</i>	
David Asatryan, Manuk Zakaryan.....	247
<i>Market-Required Competence Topic Dynamics</i>	
Tigran Topchyan.....	252
<i>Human and Telecommunication Technologies Life Cycles Comparison</i>	
Galina Gayvoronska, Illia Ganytskyi, Petr Yatsuk.....	257
<i>Features of Planning Technologies in Computer-Aided Management Systems of an Agricultural Enterprise</i>	
Iryna Balchenko, Vitaliy Lytvynov, Maria Shulga.....	263
<i>Algorithms of Automate Model Construction for Business Game Execution Subsystem</i>	
Alexandr Deryabin, Lidia Shestakova, Olga Vikentyeva.....	271
<i>Three RD Models for Two-Hop Relay Routing With Limited Packets Lifetime in AD Hoc Networks</i>	
Irma Aslanishvili.....	280
<i>Computerized Model for Aphasia Assessment, Based on Kertezs Test</i>	
Iliya Pendzhurov, Polina Mihova, Atanas Dashovski.....	287
<i>Organizing of Knowledge Bases for Emergency Situations Using Social Networks</i>	
Oleksandr Kuzomin, Bohdan Tkachenko, Mariia Lysenko.....	293

TABLE OF CONTENT OF IJ IMA VOL.3, NO.4, 2014

<i>RDFArM - a System for Storing Large Sets of RDF Triples and Quadruples by Means of Natural Language Addressing</i>	
Krassimira Ivanova.....	303
<i>Model of the Money Circulation in the Bulgarian Lands during the Antiquity and the middle Ages</i>	
Jordan Tabov	323
<i>Комплексные инструментальные средства инженерии онтологий</i>	
Виталий Величко, Кирилл Малахов, Виталий Семенов, Александр Стрижак.....	336
<i>Научные подходы системы формирования конкурентоспособного учителя</i>	
Юрий Иванович Завалевский.....	362
<i>On the Training of Consolidated Information Analysts</i>	
Kateryna Solovyova, Andrey Danilov, Olga Pishchukhina, Panasovska Yuliia, Olga Lashyna, Maria Kozlova, Maryna Khoroshenko, Maksym Voroniy.....	372
<i>Photogrammetric Technique for the Determination of The Moment in Time and the Place of the Loss of Stability During the Plastic Deformation of Sheet Metal</i>	
Spilko Hubenov, Georgi Stoilov	387
Table of content of IJ IMA Vol.3, No.1, 2014	398
Table of content of IJ IMA Vol.3, No.2, 2014	398
Table of content of IJ IMA Vol.3, No.3, 2014	399
Table of content of IJ IMA Vol.3, No.4, 2014.....	400