

An approach to service-oriented information systems architecture development based on semantic closure measure

Viktor Mokerov

D. Serikbayev East Kazakhstan state technical university
 69 A.K. Protozanov Street, Ust-Kamenogorsk, The Republic of Kazakhstan
 Email: Viktor.O.Mokerov@ieee.org

Abstract—The goal of this research is to design the approach to automated development of service-oriented architecture of information systems based on ontology. The approach presented in the paper is based on methodology of system and ontology design. Novelty of the approach is in application of semantic closure of system functions as clustering criteria. The results of the research were used in practice for design of e-university knowledge database. This paper is a part of the author's PhD thesis.

I. INTRODUCTION

APPLICATION of ontology in information systems design tasks became widely used mostly in recent years. On one hand, ontology is based on sophisticated mathematical tool - formal logic. On the other hand, expressive graphical notations developed for it make its application not more complicated than other modeling languages. In papers [1], [2], [3], [4] approaches to generation of information system structures based on system ontology and structure samples are presented. Generation of system ontology based on descriptive models is a critical task as for existing system modification as well as one of steps for system design. Issues of system ontological description development based on project documentation and views in modeling languages are described in papers [5], [6], [7], [8].

Possible solution for tasks of systems modernization and integration using knowledge database is suggested in [9], [10], [11], [12], [13]. Papers [14], [15], [16], [17], [18] are devoted to issues of knowledge databases design for information systems based on ontological approach.

One of the main characteristics of service-oriented architecture is a loose coupling of systems components that provides for support of system modifiability and scalability. The task of the architect designing a system based on service-oriented architecture is to develop system structure that provides for minimum relations between different services. Architect decomposes functions into services in accordance with functions semantic. Functions working mostly with the same fragment of subject area will be related to the same service, while functions that are not related by semantic will be related to different services. Possessing knowledge of functions semantic, this task can be automated to a large degree. This paper suggests approach to solving automation task based on system ontology.

Services structure development automation allows, on one hand, partially reduce load on the architect in solving routine tasks. On the other hand, given large quantity of initial data the application of suggested methods allows creating system prototype that can be used by the architect.

II. ONTOLOGY OF SERVICE-ORIENTED SYSTEM STRUCTURE

To solve task of design automation, it is required to reflect system elements in ontology. The main structural elements for service-oriented system are services and their operations (functions). Services and functions cannot exist on its own, they always relate to some business process that they implement, as well as possess information on physical objects of a system - contents. Knowledge of system structure, as well as of semantic of processed information, is required during system modernization. This is why it is feasible to perform ontological engineering of subject area with the purpose of developing separate ontology containing subject area semantic. Information system structure ontology in notation IDEF5 is given in Fig. 1.

Concept *Entity* describes all contents of a subject area that are described in subject ontology. *Entity* serves as super-concept for all contents of the subject area ontology. Concept *Process* presents business process of the subject area. Samples of this concept are developed based on the results of system analysis of the subject area. In accordance with Zachman's model, business processes are reflected in concept view. To model business processes the visual modeling languages can be used: UML [11], BPMN [12], IDEF0 [13]. Description of business processes given in notations can be brought to ontology as suggested in [10], [14], [15]. Concept *ProcessFunction* describes functions that provide for completion of business processes. Concept *Service* describes information systems services. Concept *ServiceOperation* describes operations of a service. Service operation is used to perform one or more functions in business process.

The following assumptions are used as limitations:

- each business process consists of at least one function;
- each function can write or read information from at least one content;

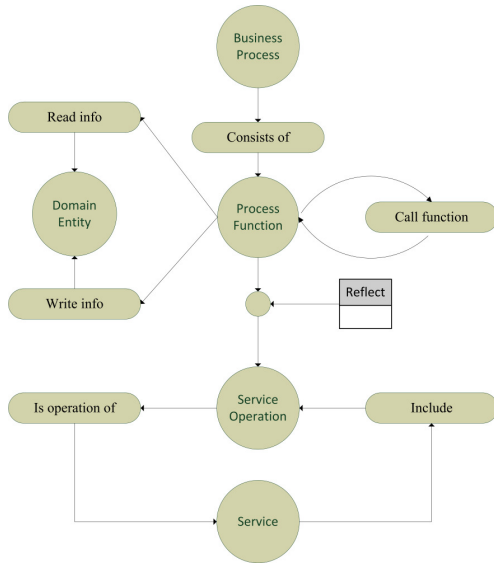


Fig. 1. Ontology of service-oriented system structure

- business process function can call for another function of this or another business process;
- service consists of at least one operation.

Role *ConsistsOf* determines membership of function in business process. Role *Call* is used to express call of one function from another. Role *ReadInfo* sets reading relations between a function and a content of subject area, role *WriteInfo* sets writing relations. Role *Include* determines inclusion of operation into service composition. Reverse role *isOperationOf* sets functional relation of function membership in service. Role *Reflect* is introduced to reflect service operations at business processes functions.

III. FORMALIZATION OF SERVICE-ORIENTED SYSTEM STRUCTURE FOR AUTOMATION OF ITS DEVELOPMENT

Task of generation of services set S comes to clustering task for system functions set $F = \{f_1, \dots, f_n\}$ with unknown quantity of clusters.

Functional maximizing semantic closure between functions of one service is used as criteria of clustering quality:

$$G = \sum_i^k \sum_{f_x \in S_i} SI(f_x, f_i) \rightarrow \max \quad (1)$$

Where k is the number of clusters, f_i is the center of cluster S_i , SI is the similarity functions between f_x and f_i business process functions.

Function of cluster elements similarity (SI) calculates measure of semantic closure for two functions f_x and f_y of subject

area in accordance with the following relations:

$$SI(f_x, f_y) = \sum_{i=1}^n w_i SI_i \quad (2)$$

$$\sum w_i = 1$$

$$SI_{BP}(f_x, f_y) =$$

$$\frac{|\exists Process.ConsistsOf.f_x \cap \exists Process.ConsistsOf.f_y|}{|\exists Process.ConsistsOf.f_x \cup \exists Process.ConsistsOf.f_y|}$$

$$SI_{ER}(f_x, f_y) =$$

$$\frac{|\exists f_x.ReadInfo.Entity \cap \exists f_y.ReadInfo.Entity|}{|\exists f_x.ReadInfo.Entity \cup \exists f_y.ReadInfo.Entity|}$$

$$SI_{ER}(f_x, f_y) =$$

$$\frac{|\exists f_x.WriteInfo.Entity \cap \exists f_y.WriteInfo.Entity|}{|\exists f_x.WriteInfo.Entity \cup \exists f_y.WriteInfo.Entity|}$$

Where $SI_{BP}(f_x, f_y)$ defines semantic closeness as a ratio of the cardinality of the subset of business processes, which connected by *ConsistsOf* role with both functions f_x, f_y , to the cardinality of the subset of business processes, which connected by *ConsistsOf* role with any of the functions f_x, f_y .

$SI_{ER}(f_x, f_y)$ - defines semantic closeness as a ratio of the cardinality of the subset of entities, which connected by *ReadInfo* role with both functions f_x, f_y , to the cardinality of the subset of entities, which connected with *ReadInfo* role with any of the functions f_x, f_y .

$SI_{ER}(f_x, f_y)$ - defines semantic closeness as a ratio of the cardinality of the subset of entities, which connected by *WriteInfo* role with both functions f_x, f_y , to the cardinality of the subset of entities, which connected with *WriteInfo* role with any of the functions f_x, f_y .

Service structure obtained as a result of algorithm performance meets the requirement of weak system relatedness.

IV. EXPERIMENTAL EVALUATION OF THE METHOD

Experimental evaluation of the suggested method was performed using processes of educational portal of East Kazakhstan State Technical University developed within the program of grants financing No. state registration 0112PK01674 (2012-2014), as application of e-university knowledge database. The fragment of electronic portal chosen for method evaluation included:

- processes: Curriculum development (P1), Education program expertise (P2), Choice of individual educational path (P3), Selection of graduates by qualification (P4);
- contents: Goal of educational program (E1), Results of educational program learning (E2); Educational program (E3), Module (E4), Course (E5), Competence (E6), Student (E7);
- functions: Search of module prerequisites (F1), Evaluation of completeness of educational program (F2), Evaluation of correctness of educational program (F3), Evaluation of goals and results of educational programs

TABLE I
SEMANTIC CLOSURE OF FUNCTIONS

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
F1	0.80	0.35	0.35	0.05	0.40	0.25	0.35	0.56	0.43	0.08	0.12	0.08
F2	0.35	0.80	0.60	0.25	0.05	0.18	0.10	0.45	0.10	0.09	0.17	0.09
F3	0.35	0.60	0.80	0.12	0.00	0.13	0.10	0.40	0.10	0.08	0.12	0.00
F4	0.05	0.25	0.12	0.30	0.00	0.06	0.12	0.24	0.05	0.10	0.13	0.04
F5	0.40	0.05	0.00	0.00	1.00	0.13	0.50	0.25	0.32	0.00	0.06	0.10
F6	0.25	0.18	0.13	0.06	0.13	0.80	0.28	0.25	0.38	0.13	0.19	0.13
F7	0.35	0.10	0.10	0.12	0.50	0.28	1.00	0.31	0.27	0.08	0.12	0.00
F8	0.56	0.45	0.40	0.24	0.25	0.25	0.31	0.80	0.19	0.12	0.09	0.05
F9	0.43	0.10	0.10	0.05	0.32	0.38	0.27	0.19	0.80	0.24	0.29	0.24
F10	0.08	0.09	0.08	0.10	0.00	0.13	0.08	0.12	0.24	0.80	0.60	0.15
F11	0.12	0.17	0.12	0.13	0.06	0.19	0.12	0.09	0.29	0.60	0.80	0.10
F12	0.08	0.09	0.00	0.04	0.10	0.13	0.00	0.05	0.24	0.15	0.10	0.80

TABLE II
MATRIX OF RELATIONS OF FUNCTIONS AND PROCESSES OF EDUCATIONAL PORTAL

Function	P1	P2	P3	P4
F1	1	1		
F2		1		
F3		1		
F4	1	1		
F5	1			
F6	1	1	1	1
F7	1			
F8	1	1		
F9	1		1	1
F10			1	
F11			1	
F12				1

TABLE III
MATRIX OF RELATIONS OF FUNCTIONS AND CONTENTS OF SUBJECT AREA BY ROLES *WriteInfo / ReadInfo*

Function	E1	E2	E3	E4	E5	E6	E7
F1				/1		/1	
F2	/1	/1	/1	/1	/1	/1	
F3			/1	/1			
F4	/1	/1	/1	/1	/1		
F5				1/	1/	1/	
F6						/1	
F7	1/	1/	1/	/1	/1		
F8	/1	/1	/1	/1			
F9				/1		/1	
F10		/1		/1			/1
F11			/1	/1	/1	/1	/1
F12		/1				1/	/1

(F4), Generation of courses module (F5), Course search (F6), Module curriculum generation (F7), Evaluation of curriculum (F8), Search of alternative modules (F9), Evaluation of influence of individual curriculum of a students on results of education (F10), Generation of individual student curriculum (F11), Search of graduates by required competencies (F12).

Relations of functions and processes by role *ConsistsOf* were determined based on process model of educational portal. They are presented in table II.

Data of table II shows that some system functions are involved in few processes at the same time, which makes it impossible to use processes as main criteria for functions uniting.

Based on the results of the analysis of information calls to database, functions relations with subject area contents by roles *ReadInfo* and *WriteInfo* were chosen. They are presented in table III.

Results of calculating measure of semantic closure of functions *SI* in accordance with (2) are given in table I.

Obtained evaluation of semantic closure was used for functions clustering. Algorithm FOREL was used to perform clustering. Experiment was conducted for neighbor elements search radius of 0.76 - average value of distance matrix composed from one complements of semantic closure matrix elements. The following clusters were obtained for search

radius:

- F1;F5;F6;F7;F9;F10;F11;F12;
- F2;F3;F4;F8;

There are no linear call sequences in obtained clusters, therefore, all obtained functions can be presented as service operations.

As the results of experimental evaluation showed, set of developed services conforms to services, which were designed empirically without using automation means.

Results showed:

- 1) functions close by semantic entered into one cluster;
- 2) functions with weak relations, not depending on used neighbor elements radius search, were chosen into separate cluster (service);
- 3) distribution of functions with high degree of relatedness depends on clustering parameters that allows identify those functions during further attempts varying neighbor elements search radius;
- 4) when radius close to average of distance matrix is used, formed clusters to a larger degree conform to services formed during empirical designing without using automatic generation of system structure.

The following services can be conditionally highlighted based on clusters formed using search radius 0.76:

- data search and input;

- analysis and expertise.

Identification of functions with high degree of relatedness allows designer to perform optimization of process model, in order to reduce relatedness of system elements.

V. CONCLUSION

Accumulated experience in information systems design can be formalized in form of knowledge database that will allow automating routine operations related to designing. Generation of architecture elements and realization of system as well as documentation and interface, allows designers to concentrate on solving application tasks. Semantic of functions relations with business-processes and subject area fragments can be used to solve task of developing information system services. Clear indication of such semantic in form of ontology allows automating process of system functions clustering into services.

REFERENCES

- [1] C. Pahl, "Semantic model-driven architecting of service-based software systems," *Information and Software Technology*, vol. 49, pp. 838-850, 2007.
- [2] J. Davies, D. Faitelson, and J. Welch, "Domain-specific Semantics and Data Refinement of Object Models," *Electronic Notes in Theoretical Computer Science*, vol. 195, pp. 151-170, 2008.
- [3] Yu. A. Orlova, "Analysis of models and methods of increasing efficiency of software design," *IZVESTIYA VolgGTU*, vol. 11, no. 9, pp. 137-141, 2010.
- [4] D. Fogli and L. P. Provenza, "A meta-design approach to the development of e-government services," *Journal of Visual Language and Computing*, vol. 23, no. 2, pp. 47-62, 2012.
- [5] V. Anaya, G. Berio, M. Harzallah, P. Heymans, A. L. Opdahl, and M. Jose, "The Unified Enterprise Modelling Language – Overview and further work," *Computers in Industry*, vol. 61, pp. 99-111, 2010.
- [6] C. Lopez, V. Codocedo, H. Astudillo, and L. Marcio, "Bridging the gap between software architecture rationale formalisms and actual architecture documents: An ontology-driven approach," *Science of Computer Programming*, vol. 77, no. 1, pp. 66-80, 2012.
- [7] K. Robles, A. Fraga, J. Morato, and J. Llorens, "Towards an ontology-based retrieval of UML Class Diagrams," *Information and Software Technology*, vol. 54, no. 1, pp. 72-86, 2012.
- [8] R. C. De Boer and H. Van Vliet, "Architectural knowledge discovery with latent semantic analysis – Constructing a reading guide for software product audits," *The Journal of Systems & Software*, vol. 81, pp. 1456-1469, 2008.
- [9] Y. Ma, "Dynamic evolutions based on ontologies," *Knowledge-Based Systems*, vol. 20, pp. 98-109, 2007.
- [10] R. Valencia-garci, R. Marti, and F. Garci, "An ontology , intelligent agent-based framework for the provision of semantic web services," *Expert Systems with Applications*, vol. 36, pp. 3167-3187, 2009.
- [11] C. Zanni-merk and D. Cavallucci, "Use of formal ontologies as a foundation for inventive design studies," *Computers in Industry*, vol. 62, pp. 323-336, 2011.
- [12] L. Rao, G. Mansingh, and K. Osei-bryson, "Building ontology based knowledge maps to assist business process re-engineering," *Decision Support Systems*, vol. 52, no. 3, pp. 577-589, 2012.
- [13] D. Strasunskas and S. E. Hakkarainen, "Domain model-driven software engineering – A method for discovery of dependency links," *Information and Software Technology*, vol. 54, no. 11, pp. 1239-1249, 2012.
- [14] T. Massoni, "A Framework for Establishing Formal Conformance between Object Models and Object-Oriented Programs," *Electronic Notes in Theoretical Computer Science*, vol. 195, pp. 189-209, 2008.
- [15] L. Thiry and B. Thirion, "Functional metamodels for systems and software," *The Journal of Systems & Software*, vol. 82, no. 7, pp. 1125-1136, 2009.
- [16] N. Bolloju, C. Schneider, and V. Sugumaran, "A knowledge-based system for improving the consistency between object models and use case narratives," *Expert Systems With Applications*, vol. 39, no. 10, pp. 9398-9410, 2012.
- [17] B. Henderson-Seller, "Bridging metamodels and ontologies in software engineering," *The Journal of Systems & Software*, vol. 84, no. 2, pp. 301-313, 2011.
- [18] A. De Nicola, M. Missikoff, and R. Navigli, "A software engineering approach to ontology building," *Information Systems*, vol. 34, pp. 258-275, 2009.