ISSN 2217-3269

e - Society Journal
Research and Applications

Volume 2 Number 1
October, 2011

E-Society Journal
Research and Applications
ISSN 2217-3269
COBISS.SR-ID 255833863

Volume 2, Number 1, October 2011
Regular Papers

Contents

E-Society Journal

Research and Applications

Publisher:

University of Novi Sad
Technical Faculty “Mihajlo Pupin”
Djure Djakovica bb, Zrenjanin, Serbia

Editor:

Zora Konjović
Faculty of Technical Sciences
University of Novi Sad

CIP – Каталогизација у публикацији
Библиотека Матице српске, Нови Сад

621.3
004.4

E-Society Journal : research and applications / editor Miodrag Ivković. - Vol. 2, No. 1 (Oct. 2011)- . - Zrenjanin : University of Novi Sad, Technical Faculty "Mihajlo Pupin", 2010-. - 25 cm

Dva puta godišnje.
ISSN 2217-3269

COBISS.SR-ID 255833863

Printed by: PC centar Magus, Zrenjanin
Printing: 50 copies

Zrenjanin 2011.

Word of the Editor

Zora Konjović

Faculty of Technical Sciences/Computing and Control Department, Novi Sad, Serbia

Information society and consequently, information technologies and information management are the "hot topics" both for the scientific community and the practitioners from various industries.

Following this trend, the First International Conference on Information Society, Technology and Management was held in mountain resort Kopaonik, Serbia on March 7-8, 2011. The aim of the Conference was to get together researchers and practitioners interested in technological and social aspects of information society thus giving them an opportunity for presentation of the latest scientific results and exchange of the good practices related to information society development. At the Conference 34 papers were presented in total.

This special issue of the Journal is the selection of the papers presented at the Conference made by the guest editor of this Journal Special Issue.

The guest editor selected the papers based on two criteria. The first criterion was capturing the profile of the Conference, both in topics and participants. The second criterion was the quality of the papers.

Based on these criterions 12 papers have been selected for this Special issue. The papers are classified into three main groups which, by editor's opinion, capture the profile of the Conference: Future Internet, Soft Computing, and E-society.

The group **Future Internet** comprises the papers belonging to two subgroups: Semantic technologies and Internet of things.

The group **Soft Computing** comprises the papers belonging to subfields genetic algorithms, fuzzy approach and data mining.

The group E-society comprises the papers belonging to subtopics E-government and E-business, and technology enhanced learning.

Semantic Categorization of Data by Determining the Similarities of Associations of the Semantic Network

Miloš Stojković, Miodrag Manić, Milan Trifunović, Dragan Mišić
University of Niš, Faculty of Mechanical Engineering

Abstract: **In the paper is presented concept of semantic network analysis which performs semantic categorization of data in the network. The concept is based on determining the similarity between semantic network associations which is done on two levels. On the first level, the determination of associations similarity is based on similarity of attributes values of associations. On the second level, the determination of association's similarity is performed according to the similarity of graph topology, or association's subgraphs. The concept allows highly efficient semantic categorization of new concepts, which does not depend on pre-planned inputs and predefined rules of deduction. Also, the concept allows different semantic interpretations of the same concept in different semantic contexts.**

1. INTRODUCTION

Semantic categorization (classification of data according to meaning) and interpretation of data (SCD, SID), represent one of the biggest challenges faced by modern information technologies. In fact, this problem is closely related to the ability of computer applications to attach certain meaning to data which is being processed. The motive for the solution of this problem lies in the ever increasing need to enable software applications to provide meaningful answers when it is not possible to predict the input, and consequently the code by which a meaningful response is programmed. This need becomes more intense with increasing amounts of information available. Computer networking in the Internet has enabled instant access to indefinite amounts of data and information. However, it is often difficult rather than easy to find quality information - one for which the user searches. Quality of information access does not equally follow the increase of available information [1], [2]. Research in cognitive psychology suggest that the process of searching information contents should include the search of meaning (or according to the meaning) of what is required in order to conduct search significantly more precise and even faster [3], [4]. Following these recommendations, the current research trend in the field of information technologies includes the two main subdirections which complement each other: a) research and the creation of semantic data models and b) research and modeling of cognitive processes. The studies of semantic data models aim to examine the way we form information about the outside world and how information is stored in our memory in order to create data models that will be able to "accept" and "incorporate" the meaning or semantics of data. Of course, the ultimate goal is not the integration of meaning alone, but its automated and intelligent interpretation and use [5]. For that reason, studies of cognitive processes are trying to determine how this information is used to solve different types of more or less complex problems [4], and to study the mechanisms of cognition and reasoning in humans, in order to model these processes and perform them automatically by artificial-intelligence systems. These directions in science are recognized as an research area of supreme importance for the near future of the European Union [6].

2. CURRENT SOLUTIONS

In solving the problem of semantic categorization and interpretation, so-called functional modeling – FM is the most widely used approach. Basically, it uses Generic Semantic Model (GSM) whose most prominent derivatives are entity-relationship model (ER), a functional data model (FDM) and semantic data model (SDM) [7]). Essentially, this is a way of presenting the knowledge coming from the concept of semantic networks [8], and the first graphical representations of the so-called semantic memory structure. The sudden development of web technologies in the last two decades also has accelerated the evolution of functional modeling and its application [9], [10], [11]. The main motive for influence of functional modeling on web technologies and/or vice versa, is the absolute priority of improving the quality of information retrieval on the Internet [5]. In order to improve the quality of search it was necessary to use not only phenomenal but also the semantic (meaningful) combinations of keywords (some sort of semantic context). In fact, in order to quickly find the right information in a huge amount of data on the Internet it's imperative to have "information on this data" - metadata, which will allow software applications for browsing the internet to find the requested information. Metadata, therefore, should be used to describe the meaning of the requested content on the Internet. Methods and models of describing semantic features are taken from the functional modeling [9], while syntax for metadata structuring is based on extensible markup language (XML) and its derivatives (RDF / RDFS, DAML + OIL and OWL) [5]. Ontologies are often referred to as vocabularies of so-called web objects made in OWL notation [5]. Like other vocabularies, their purpose is to present the meaning of terms contained in them, and in this case, the meaning of these web objects. Web objects are actually concepts used for representation or conceptualization of data and information content on the web. Vocabularies are organized according to the semantic networks model, where the nodes are concepts or web objects and functional relationships between web objects are represented by semantic network relations. It is precisely this organization that supports and enables the semantic categorization and interpretation of web objects.

2.1. Limitations

However, ontologies also have significant limitations on in terms of their ability to enable SCD and SID. These limitations originate from imperfections in the structure of the vocabulary, especially concerning semantic relations, and the lack of appropriate methods of cognitive data processing and semantic interpretation of data (web-objects, concepts). With ontologies SID is conducted on two levels.

On the first level, functional or semantic relations in the segment of semantic network around the currently observed (new) concept are analyzed and new functional relationships from or to a new concept are created according to pre-prepared set of rules. Which functional relation will be created depends upon analyzed functional relations graph. In essence, the rule set contains an encoded first-order logic which is used by ontologies inference engine to create the adequate functional relation between currently considered concept and other concepts in the semantic network for pre-planned cases of functional relations subgraphs (Fig. 1). Creation of new functional relations between data that is currently considered and data already described in the vocabulary results in semantic categorization of a new concept.

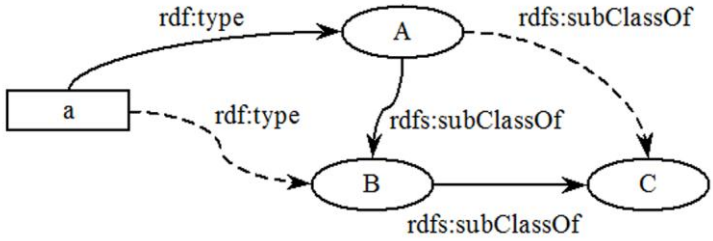

Figure 1: Propagation type inferencing schema

Although this set of rules brings partial analyticity of semantic interpretation data, autonomy and flexibility of the semantic interpretation of data is completely abolished. Each new case of functional relations subgraphs, which is not provided by the set of rules, can not be used for semantic interpretation. Another important aspect which further compromise the intelligence of ontologies "inference engine" is the fact that embedded set of rules covers only the simplest relations subgraphs. "Inference engine" is not designed to analyze large subgraphs. Thereby ontologies "inference engine" is not able to perceive the wider context of functional relations which further reduces the analyticity SID. Also, it disables ontologies "inference engine" to connect SID with semantic contexts, or to analyze the analogies between semantically distant subgraphs.

At another, deeper level, by using of SWRL (Semantic Web Rule Language) syntax an additional domain knowledge in the form of production rules can be embedded in the sole content of vocabulary. Of course, this kind of knowledge can also be used for semantic categorization of data, but only for the cases with provided input (an example of this kind of rule is presented in expressions 1 and 2):

IF

$$P_j^i \in comp(A_i), j = 1,2,\dots,m \wedge FF_{jk} \in comp(P_j^i) \quad (1)$$

where P_j^i is a part, A_i is an assembly, FF_{jk} is a form feature (surface, edge, etc.)

THEN

$$B_{jk}^{(i)} : FF_{jk} : \rightarrow P_j^i, k = 1,2,\dots,n, \quad (2)$$

where „→“ designates property: *belongsTo*

In addition to inference engine limitations regarding cognitive data processing, there is a limitation regarding vocabulary structure. In fact, the method of structuring or incorporating the meaning of data can be considered very limited. First of all, functional relations are treated as concept features, i.e. structurally relations are encapsulated into the concept definitions in the form of object-oriented programming - for appropriate so-called *domain* and *range*. This approach to building relations indicate that the meaning of the concept and knowledge of it is seen as valuable content of the box which can be reached only by opening the box (through the concept). The flexibility and analyticity of semantic categorization of data is significantly limited by this ontology feature because it is not easy to take advantage of the functional relations from one domain to another, especially if domains are semantically distant or, even worse, if domains are semantically similar, but syntactically different. In such cases, ontologies inference engine remains to rely on a limited set of rules from the first-order logic. One

problem that directly arises from these limitations is the lack of so-called ontologies interoperability. Two although, semantically very close vocabularies usually only for syntactic differences can not be recognized as such therefore can not be linked, complemented or modified mutually.

Another drawback which results from modeling approach to the meaning of the concept by encapsulation into the concept structure, is the leanness of relations attributes. Given that the focus of creating the meaning is not on relations, but on the structure of concepts, there is no obvious need for more detailed semantic differentiation of relations. This, in turn, directly limits the mechanism for the analysis of relations graphs and subgraphs, and consequently all the other deficiencies of cognitive processing.

Finally, of no lesser importance is following limitation, not being flexible ontologies inference engine is unable to self-learn and reprogram itself on basis of it's own work and feedback from users/teachers.

3. ACTIVE SEMANTIC MODEL (STRUCTURE)

An alternative approach to modeling of semantic features of data in a semantic network is based on the that the semantic designation of certain concept (network node) is contained in the semantic relations between that and other concepts. Based on this hypothesis a special model of semantic network which is called the *Active Semantic* Data *Model* - ASM has been designed and developed [13], [14]. In ASM semantic relations are modeled as separate data structures that are (structurally) independent of the concepts (network nodes). Moreover, the names of concepts that they connect are their attributes (expression 3). In addition to the concepts connected by them, semantic relations are modeled (defined) with eight more parameters.

$$\mathbf{A}(\boldsymbol{cpt_i}\leftrightarrow\boldsymbol{cpt_j})_{ij}: \{ctx_i, \boldsymbol{cpt_i}, r_i, t, h, s, c, d, r_j, \boldsymbol{cpt_j}\} \quad (3)$$

One of these eight parameters (ctx_i) is used to identify a semantic context to which given association belongs - the other seven parameters are considered correct for given context. (Fig. 2). Two parameters: *accuracy* (***h***) and *significance* (***s***) of the association represent so-called weight parameters used to define assumed accuracy and importance of semantic features stated by the relation for given context (Fig 2).

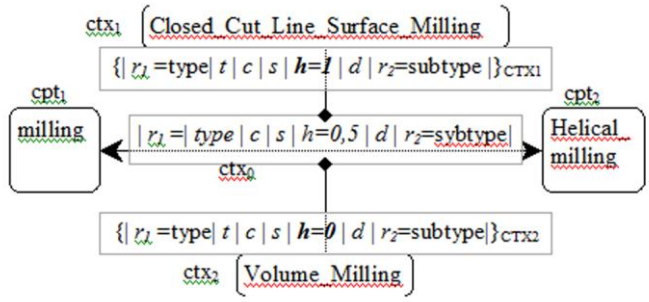

Figure 2. The value of associations' parameters depends on context

The remaining five parameters are named *topological* because in addition of stating a functional relation between two concepts (*type of semantic relation* ***t*** and their *roles* in it, $\boldsymbol{r_1}$, $\boldsymbol{r_2}$) and method of association, they define the topological features of associations graph (*character* - ***c*** and *direction* - ***d*** of associating).

4. SCD & SID BASED ON ANALOGIES

The procedure of SCD and SID in ASM is built on the claim that the use of analogy is the cognitive process core and can be considered as the primary process of cognition and communication. [15], [16], [17]. Examples of the most common heuristics as well as general strategies used in solving the problem are based on the use of analogies. [18]. In fact, well-known procedure that has proven successful in solving previous similar problems is used to solve the new problem. Condition for the success of this strategy is the *recognition* (*Einsicht* - a sudden perception of significant *relations* that leads to solutions to problems [19]) analogy between the two problems and the recall of previously implemented solutions. One of the reasons why sometimes it is difficult to recognize the analogy between the two problems is because their elements are different, although the relations are the same. Insight on the common elements of the solution is actually hiding in the similarity and/or identity of the relation between these elements [4]. This thesis was used to develop algorithms for SCD and SID in ASM.

SCD starts with *introducing* the concept whose semantic categorization is needed (the new concept) to ASM through the establishment of initial semantic relations between new concept and concepts that already exist in the network [13]. Initial introduction of a new concept with ASM is performed through so-called *communication context* - a kind of matrix that helps to define the initial associations plexus (subgraphs) between new concept and other concepts in the network.

The next step is *refinement* by allocating a subset of *significant* and *accurate* associations and for the the so-called general and the appropriate particular context (expressions 4, 5, and 6). Weighting parameters ***s*** and ***h***, and the attribute of belonging to a particular context ***ctx*** are used for precise control of the segmentation of semantic network where to look for an analog network patterns (associations plexus) on which SCD and SID will be executed.

$$\bar{S} = AVG\left(s_{A(CPT_X \leftrightarrow CPT_1)}, s_{A(CPT_N \leftrightarrow CPT_1)}\right) \quad (4)$$

where $A_{(CPT_{X,N} \leftrightarrow CPT_1)} \in \{CTX_i\}$

$$\bar{H} = AVG\left(h_{A(CPT_X \leftrightarrow CPT_1)}, h_{A(CPT_N \leftrightarrow CPT_1)}\right) \quad (5)$$

$$\bar{S} \cdot \bar{H} \geq T, \text{ where } T \text{ is a defined threshold} \quad (6)$$

All those who are found to have sufficient accuracy and significance for given context (which are subject to the condition 6) shall be taken into consideration in further proceedings in SCD and SID.

After *refinement* SCD (and SID) implements the so-called procedure of *determining the parameters of association* - *DPA* [13]. This procedure results in the creation of new associations between the new and known concepts of the network, thus actually conducting

semantic categorization of a new concept. As before launching SCD new concept has some initially created associations to known concepts in the network, SCD increases in the number of semantic relations that describe it ("deepening" of knowledge about the concept). On that occasion, with new associations, the knowledge of some known concepts is also being "deepened".

DPA procedure begins with the analysis of refined set of associations which consists of associations between new concept and existing concepts in the network (created in the communication context), and associations that continue to link these existing concepts with other concepts in the network. Of course, the procedure of DPA takes place simultaneously and separately for every identified context (Fig. 3). Determining the value of each parameter of the new association within the DPA procedure is conducted by a special subalgorithm (expression 7).

$$p\left(A_{CPT_X \leftrightarrow CPT_N}\right) = F_p \begin{Bmatrix} \boldsymbol{P}\left(A_{CPT_X \leftrightarrow CPT_1}\right), \boldsymbol{P}\left(A_{CPT_1 \leftrightarrow CPT_N}\right) \\ \dots \\ \boldsymbol{P}\left(A_{CPT_X \leftrightarrow CPT_t}\right), \boldsymbol{P}\left(A_{CPT_t \leftrightarrow CPT_N}\right) \\ \dots \\ \boldsymbol{P}\left(A_{CPT_X \leftrightarrow CPT_k}\right), \boldsymbol{P}\left(A_{CPT_k \leftrightarrow CPT_N}\right) \end{Bmatrix} \tag{7}$$

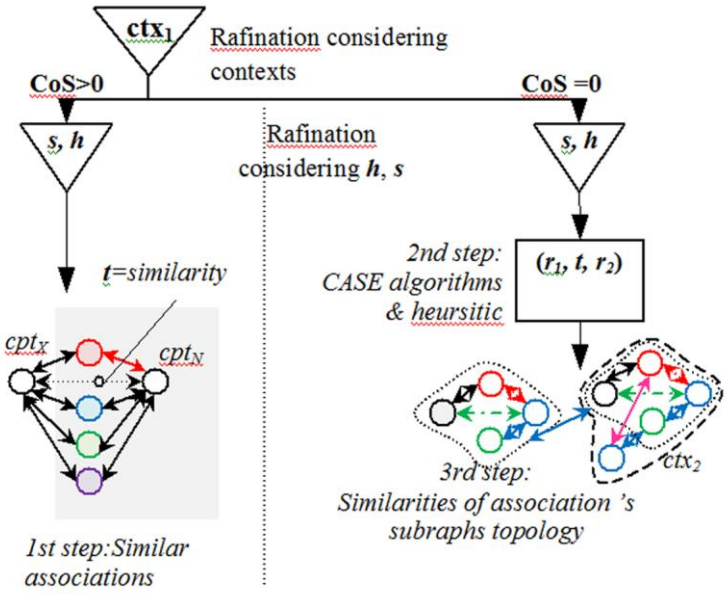

Figure 3. Branching off the *type* determination algorithm

The first in a series is a subalgorithm for determining type of association. This algorithm determines not only the type of new associations but also the role of each of the concepts that will be connected by the new association. It is implemented in three steps. However, before executing any of these three steps, additional refinement is performed. The set of association pairs whose types are the same (so-called first and upper classes of similarity CoS> 0) and a set of association pairs whose types are not the same (so-called zero class of similarity CoS = 0) are extracted.

In the first step is executing subalgorithms for determining other parameters of associations that will emerge between the new concept and those concepts whose association pairs are at least of the same type (Fig. 3). Resulting asociations from this branch of the DPA procedure are of different degree and class of *similarity*, or *differentness* (e.g., a new concept CPT_X and well known concept CPT_N are synonyms, antonyms, similar concepts with significant difference, incompletely opposite, etc.). What degree and class of semantic similarity/differentness of concepts are induced by these association depends on the obtained values of other parameters.

In the second and third step, DPA procedure processes associations pairs that are not of the same type (nor the role of concepts coincide). ASM decides the type of association that will connect the concepts CPT_X and CPT_N depending on the topological class of associations *plexus* i.e. kind of topology of association's subgraphs.

The second step is characterized by the execution of so-called CASE subalgorithms designed for pre-planned kind of asociation plexus (e.g. association plexus that describe an action, assembly, subject etc.). CASE subalgorithm is built based on the explicitly defined causal mechanism (rule) for each provided case of asociations plexus. This part of the DPA is in many ways reminiscent of the coded first-order logic which ontologies inference engine uses for pre-planned cases of functional relations subgraphs. ASM offers the possibility of constant expansion of the CASE subalgorithms base, based on monitoring of the efficiency and factors of decisions usefulness (*knowledge crystallization algorithm* can not be presented in this paper due to its volume).

Still, true flexibility, autonomy and analyticity of the DPA procedure (i.e. SCD and SID) is shown in the third step of the subalgorithm for determining of association type. This part of the procedure is intended for upfront unforeseen topology types of association subgraphs. The algorithm is based on recognizing the similarity between topologies of association plexuses (subgraphs). ASM scans the network and extracts set of association plexuses $\{CTX_N\}$ which are topologically analogous with input association plexus CTX_x in which the new concept appears CPT^x_1 ($CTX_x \asymp F(CTX_N)$) (Fig 4.)

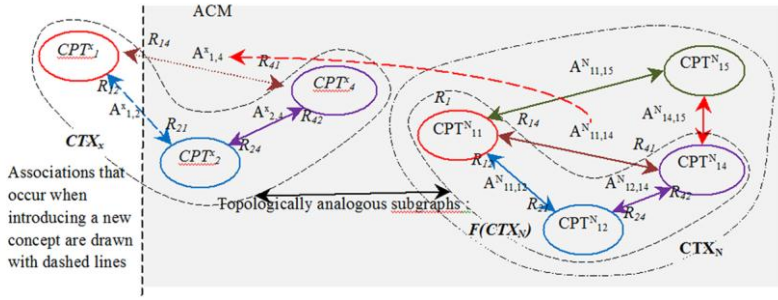

Figure 4. Type determination schema is based on similarity between the topologies of subgraphs of associations

ASM singles out this set of association plexuses by recognition of so-called *topologically corresponding associations* and *concepts*. Two associations that belong to different contexts and CTX_i и CTX_j and thereby have similar values of weight parameters and the same values of topological parameters are called *topologically corresponding associations* of two different contexts – TCAs. Concepts that belong to TCAs of two different contexts or association plexuses (CTX_i and CTX_j), and thereby have the same role in the TCAs, are called *topologically*

corresponding concepts of two different contexts or TCCs. The algorithm for recognition of TCA and TCC is located in the core of ASM's SCD and SID. Basically it is a kind of algorithm for recognition of subgraph isomorphism, which was originally designed in Laboratory of Intelligent Production Systems of University of Nis [13].

Once it recognize the topological analogy between a simple input subgraph CTX_x and previously stored subgraphs $\{CTX_N\}$, a procedure for upgrading the input subgraph is preformed in accordance to the exemplar of the subgraphs that are topologically analogous to CTX_x.

ASM, in general, may allocate a number of different elementary subgraphs (basic association subgraph is one that contains two associations and three concepts) and their sets, which "offer different exemplars" for input subgraph upgrade. The suitability of selected exemplar associations is evaluated primarily by the degree of topological similarity of basic subgraphs (or sets of basic subgraphs) they belong to, and input subgraph (or a input set of basic subgraphs). The degree of topological similarity of exemplary sets of basic subgraphs and input set is measured by the number of exemplary sets of basic subgraphs that are topologically analogous. The exemplary association (and its type), which ASM first recommends is TCA that comes from the largest set of exemplary elementary subgraphs. This emphasizes the importance of the degree of topological similarity of the subgraphs of association for the type determination.

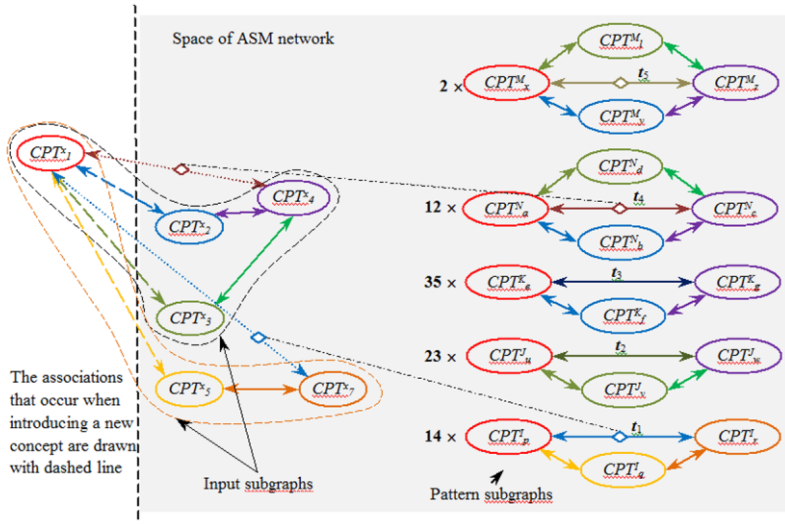

Figure 5. Type determination in the case of several candidates: basic subgraphs of associations of different type

$$\left\{\begin{matrix} CPT_1^x \leftrightarrow CPT_2^x \leftrightarrow CPT_4^x \\ CPT_1^x \leftrightarrow CPT_3^x \leftrightarrow CPT_4^x \\ CPT_1^x \left(\overset{t_4}{\leftrightarrow}\right) CPT_4^x \end{matrix}\right\} \rightleftharpoons \left\{\begin{matrix} CPT_a^N \leftrightarrow CPT_b^N \leftrightarrow CPT_c^N \\ CPT_a^N \leftrightarrow CPT_d^N \leftrightarrow CPT_c^N \\ CPT_a^N \overset{t_4}{\leftrightarrow} CPT_c^N \end{matrix}\right\} \quad (5)$$

$$\left\{\begin{matrix} CPT_1^x \leftrightarrow CPT_2^x \leftrightarrow CPT_4^x \\ CPT_1^x \leftrightarrow CPT_3^x \leftrightarrow CPT_4^x \\ CPT_1^x \left(\overset{t_3}{\leftrightarrow}\right) CPT_4^x \end{matrix}\right\} \rightleftharpoons \left\{\begin{matrix} CPT_e^K \leftrightarrow CPT_f^K \leftrightarrow CPT_g^K \\ \\ CPT_e^K \overset{t_3}{\leftrightarrow} CPT_g^K \end{matrix}\right\} \quad (6)$$

In the case that there is more topologically different sets of exemplary basic subgraphs that are equally topologically similar with the input set of basic subgraphs, ASM recommends exemplary association from that set of basic subgraphs which are more in the network. This emphasizes the importance of frequency, that is representativeness of certain topological class of subgraphs for determination of the type of association (expressions 7 and 8).

$$12 \times \begin{Bmatrix} CPT_1^x \leftrightarrow CPT_2^x \leftrightarrow CPT_4^x \\ CPT_1^x \leftrightarrow CPT_3^x \leftrightarrow CPT_4^x \\ CPT_1^x \left(\overset{t_3}{\leftrightarrow}\right) CPT_4^x \end{Bmatrix} = \begin{Bmatrix} CPT_a^N \leftrightarrow CPT_b^N \leftrightarrow CPT_c^N \\ CPT_a^N \leftrightarrow CPT_d^N \leftrightarrow CPT_c^N \\ CPT_a^N \overset{t_4}{\leftrightarrow} CPT_c^N \end{Bmatrix} \tag{7}$$

$$2 \times \begin{Bmatrix} CPT_1^x \leftrightarrow CPT_2^x \leftrightarrow CPT_4^x \\ CPT_1^x \leftrightarrow CPT_3^x \leftrightarrow CPT_4^x \\ CPT_1^x \left(\overset{t_3}{\leftrightarrow}\right) CPT_4^x \end{Bmatrix} = \begin{Bmatrix} CPT_x^M \leftrightarrow CPT_y^M \leftrightarrow CPT_z^M \\ CPT_x^N \leftrightarrow CPT_1^N \leftrightarrow CPT_z^N \\ CPT_x^N \overset{t_5}{\leftrightarrow} CPT_z^N \end{Bmatrix} \tag{8}$$

Belonging of each association to a particular context enables DPA algorithms to focus the analyses according to context. This approach in modeling the semantic content and the SCD procedure allows the same concept to be semantically interpreted differently in different contexts, which is very important feature of ASM.

5. RESULTS

The concept of semantic networks ASM and its analysis for SCD have been developed primarily in order to be used as add-in application in CAD / CAM systems, which should help the engineer in making certain decisions in situations that are characterized by significant uncertainty. Particularly interesting application of ASM could be in modules for reverse modeling and free-form design (bioforms and art design). The uniqueness of these forms intrudes the need for a system that is capable to interpret the geometry features of the model semantically and thus to help the designer to decide what features to use for further modeling and remodeling.

Preliminary tests of the application of the concept were carried out for workflow management system [20]. ASM and its algorithms for SCD and SID have shown the ability to recognize and categorize semantically unexpected exceptions that cause disruption to workflow. (e.g. the workflow of designing the bone implant). The following verification of the solution will include thorough tests that have to prove ability of ASM to automate and optimize selection and composition of the design features to create a geometric model of bio- or art-forms.

6. CONCLUSION

The concept of SCD and SID by determining the similarities of associations of the semantic network was created as a kind of modification or extension of known concepts of analogy based reasoning. SCD and SID that ASM's algorithms provide is based on recognition and

determining the similarity of concepts, and this, in turn, on recognition and determining the similarity of associations between the concepts and similarity of subgraphs of associations. Determining the similarity of associations and subgraphs of associations is performed by the recognition of topological analogy between subgraphs of associations.

Compared to inference engine of OWL, ASM has autonomous, flexible and significantly more analytical mechanism of SCD. Besides the expandable base of CASE algorithms, which are used for efficient inferring in cases of known and predicted inputs and situations, ASM makes the meaningful judgments, conclusions and decisions in new and unexpected situations using the algorithms for determination of similarity of associations and topological analogy of association's subgraphs.

7. ACKNOWLEDGEMENT

The paper presents a concept that is a result of the multidisciplinary research from the domain of IT application in bioengineering. The research project is sponsored by the Ministry of Science and Technology of the Republic of Serbia - project id 41017

REFERENCES

[1] [1] Kearney, A. T., "Study Shows $750 Billion Waste of Time,http://www.knowledgeboard.com/cgi-bin/item.cgi?id=44235, 2002.

[2] Krill P., "Overcoming Information Overload" InfoWorld, 7-Jan-2000, 2000.

[3] Peterson, L. R., Peterson, M. J., Miller, G. A. "Short-term retention and meaningfulness" Canadian Journal of Psychology, Vol 15, pp 143-147,1961.

[4] Kostić, A. Kognitivna psihologija, Zavod za udžbenike i nastavna sredstva, Beograd, 2006.

[5] Daconta, M. C., Obrst, L. J., Smith, K. T., The Semantic Web. Wiley Publishing, Indianapolis, 2003.

[6] ICT Summit. ICT-WP-Call3_Challenge2. Retrieved from http://www.ictsummit.eu/template/fs/documents/ ict-wp-Call3_Challenge2_en.pdf, 2007.

[7] Hull, R., King, R. "Semantic Database Modeling: Survey, Applications, and Research Issues" ACM Computing Surveys , Vol 19 No. 3, pp 201-260, 1987.

[8] Quillian, R. M. "Semantic Memory", in M. Minsky (Ed.), Semantic information processing. MIT Press, Cambridge, 1968.

[9] Chandrasekaran, B. "Representing function: Relating functional representation and functional modeling research streams", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Vol. 19, pp 65–74, 2005.

[10] Fensel, D., A Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, Heidelberg (Berlin), 2001.

[11] Mizoguchi, R., "Tutorial on ontological: Engineering part 3 advanced course of ontological engineering", New Generation Computing , Vol. 22 No. 2, 2004

[12] Kyoung-Yun, K., David, M. G., Hyungjeong, Y., "Ontology-based assembly design and information sharing for collaborative product development", Computer Aided Design, Vol. 38, pp 1233–1250, 2006

[13] Stojković, M. "Analysis of the Manufacturability Parameters Based on Semantic Structures of the Digital Product Model", PhD thesis, 2011.

[14] Stojkovich M., Manich M., Trajanovich M., Korunovich N., "Semantic Structures In The Product Data Model", Proceedings: Product Lifecycle Management - Assessing the industrial relevance PLM-SP3, pp 227-234, Milano, 2007.

[15] Gentner, D., Holyoak, K. J., & Kokinov, B. N., The Analogical Mind: Perspectives from cognitive science, MIT Press, Cambridge, 2001.

[16] Foxwell, H. J., & Menasce, D. A., "MARVIN: A Web-Based System for Representing, Retrieving, and Visualizing Analogies", World Wide Web: Internet and Web Information Systems, Vol. 7, pp 385-419, 2004.

[17] Bar, M., "The Proactive brain: using analogies and associations to generate predictions" Trends in Cognitive Science , Vol. 11 No. 7, pp 280-289, 2007.

[18] Mednick, S., "The associative basis of the creative thinking", Psychological Review, Vol. 69, pp 562-569, 1962.

[19] Köhler, W, The Mentality of Apes, Harcourt, Brace, New York, 1927.

[20] Mišić, D., Stojkovic, M., Domazet, D., Trajanovic, M., Manic, M., & Trifunovic, M., "Exception detection in business process management systems", Journal of Scientific & Industrial Research, Vol. 69, pp 188-193, 201.

Towards a Formal Framework for Semantic Interoperability in Supply Networks

Milan Zdravković, Miroslav Trajanović
University of Niš, Faculty of Mechanical Engineering

Abstract: **In this position paper, we summarize our achievements in the research and development of formal framework for semantic interoperability in supply networks. First, we identify common problems of traditional supply chains and argue that interoperability of enterprise information systems is crucial for their resolution. Second, we discuss on the foundational aspects of semantic interoperability and criteria for its evaluation. Third, multiple abstraction layers of this framework are described, including 1) individual enterprises' realities, namely, local ontologies, 2) common collaborative practices, namely, formalizations of relevant reference models, and 3) application ontologies. Finally, we identify gaps and provide guidelines for the future research in the topic.**

1. INTRODUCTION

Supply chain is a complex, dynamic networked environment which resembles a number of different actors, assets, goals, competencies, functions and roles. The interest in creating a new discipline of supply chain management (SCM) was developed in the early '60s with the initial motivation to investigate the increase in demand fluctuation (known as "bullwhip effect") which occurred in deeper levels of the manufacturing supply tree [1]. Despite the advances in ICT, organizational sciences and rich implementation experience, there are many evidences that the level of adoption of the paradigm of SCM and related tools, methods and practices is still low. One of the effects of the development in this domain is that the manufacturers started to view their suppliers as extensions of themselves. Consequently, supply chains are characterized by high level of integration, technical and organizational, focusing on a cost reduction as a key motivation for collaboration. We consider this type of supply chains as traditional. Traditional approach to supply chains' configuration may have negative impact to their performance. First, high-speed, low-cost supply chains are often unable to respond efficiently to unexpected structural changes in demand or supply. Second, high level of integration reduces flexibility of small and medium enterprises, main constituents of the lower levels of supply chain. Third, investments in technical framework for enterprise integration cannot be returned in a short term. Furthermore, starting collaboration in such traditional settings is reactive decision. Namely, relationship establishment is motivated by internal, rather than external factors: complexity and volume of supply relationships, potential for cost reduction [2], high frequency of transactions between parties [3], degree of asset specificity [4], etc.

In a response to the weaknesses of static architecture of the supply chain, a notion of virtual enterprise has been introduced and widely discussed. Virtual enterprise is a temporary network of independent enterprises, who come together quickly to exploit fast-changing opportunities and then dissolve [5]. It is characterized by a short-living appearance of a supply chain, capable to produce low volume of high variety of products, by drawing from the loosely-coupled, heterogeneous environment of available competences, capabilities and resources, referred to as Virtual Breeding Environment [6]. One of the main challenges of bringing these paradigms to

reality is complexity of the ICT environment and heterogeneity of corresponding enterprise information systems (EIS). Those challenges are related to realization of some of the fundamental requirements for ICT applications - enterprise integration and interoperability [7]. The dominant integration approach today, is Service-Oriented Architecture (SOA) [8]. Still, its implementation is resource and time-consuming. For the achievement of full interoperability, the federated approach is proposed [9]. It is characterized by dynamic accommodation of the systems, based on a pre-determined meta-model. In the remainder of this paper, we describe the formal framework for semantic interoperability in supply networks, which is based on this approach.

2. FORMAL FRAMEWORK FOR SEMANTIC INTEROPERABILITY IN SUPPLY NETWORKS

Paradigms of virtual enterprises and their breeding environments are based on the capability of an enterprise to configure or reconfigure quickly, according to the circumstances of the market, often not known in advance, or even in the moment of configuration. Hence, efficiency and effectiveness of this joint endeavor depends on the interoperability of enterprises, rather than their integration.

The main prerequisite for achievement of interoperability of the loosely coupled systems is to maximize the amount of semantics which can be utilized and make it increasingly explicit [10], and consequently, to make the systems semantically interoperable.

2.1. Foundational aspects of semantic interoperability

ISO/IEC 2382 defines interoperability as the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units [11]. Semantic interoperability builds upon this notion and it means ensuring that the precise meaning of exchanged information is uniquely interpreted by any system not initially developed for the purpose of its interpretation. It enables systems to combine and process received information with other information resources and thus, to improve the expressivity of the underlying ontologies. In our research, we adopt the definition of John Sowa [12], because we can use it to evaluate semantic interoperability of EISs. We represent this definition in controlled natural language, as asymmetric logical function semantically-interoperable(S,R):

$$
\begin{aligned}
&\text{data}(p) \wedge \text{system}(S) \wedge \text{system}(R) \wedge \text{semantically-interoperable}(S,R) \Rightarrow \\
&\forall p\,(\\
&(\text{transmitted-from}(p,S) \wedge \text{transmitted-to}(p,R)) \wedge \\
&\forall q(\text{statement-of}(q,S) \wedge p \Rightarrow q)\ \exists q'(\text{statement-of}(q',R) \\
&\wedge\ p \Rightarrow q' \wedge q' \Leftrightarrow q) \\
&)
\end{aligned}
$$

Figure 1 illustrates the following assumption of semantic interoperability of systems, represented by the local ontologies: when two different application ontologies of two partners in the supply chain (or two departments or contexts of the same enterprise) are mapped to the

same domain ontology, relevant EISs whose knowledge they represent will become fully or partially semantically interoperable in specific direction, depending on the mappings. Thus, if there exist two isolated EISs S_1 and S_2 and corresponding application ontologies O_{L1} and O_{L2} and if there are mappings M_{L1D1} and M_{L2D1}, established between the concepts of O_{L1}, O_{L2} and domain ontology O_{D1}, respectively, then there exist mappings M_{O1O2} which can be inferred as logical functions of M_{L1D1} and M_{L2D1}. Each of the local ontologies may represent one of the contexts of the enterprise (C_1-C_n).

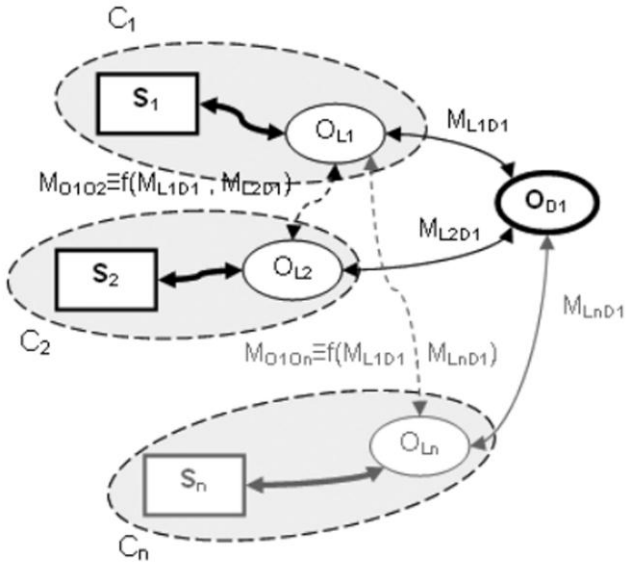

Figure 1. Semantic interoperability of systems

2.2. Description of the ontological framework

The concepts and tools presented in this paper are using the formal framework of supply chain operations [13] presented at Fig. 2. It is based on Supply Chain Operations Reference (SCOR) [14] - a standard approach for analysis, design and implementation of five integrated processes in supply chains: plan, source, make, deliver and return. SCOR aims at integrating processes, metrics, best practices and technologies with the objective to improve collaboration between supply chain partners.

The framework is developed with goals to enable the semantic interoperability between SCOR-based systems and other relevant EISs, and to improve the expressivity of SCOR-based models.

The approach to its development is based on a premise that domain knowledge evolves at highest rate at lower levels of abstraction in domain community interaction. Consensus on the specific notions is more likely to be reached than agreement on the generalizations and abstractions. However, this level of abstraction is often characterized by the implicit semantics of the standards, reference models, etc. Hence, we consider coherence between creation, evolution and use of highly contextualized knowledge and development of formal expressive models as very important factor for usability of the models.

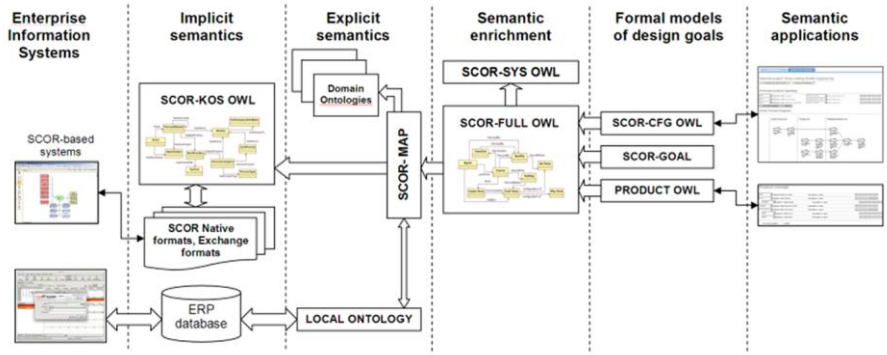

Figure 2. Formal framework of supply chain operations

In process of development of an ontological framework for supply chain operations, we start with representing the implicit semantics of SCOR model as knowledge organization system (KOS) by using OWL (SCOR-KOS OWL). Thus, we ensure the interoperability of SCOR-based enterprise applications with semantic systems. Next, semantics of the SCOR elements is made explicit by the concepts and relations of SCOR-Full ontology and rules, used for mapping those to SCOR-KOS OWL concepts (SCOR-MAP). Various design goals, such as generation of process models, acquisition of product data and goal reconciliation are formalized by the application ontologies which are used by the corresponding systems. Finally, the realities of the enterprises are represented in the framework by the local ontologies, which transform the implicit database schemas to explicit semantic models.

2.3. Domain ontologies

SCOR-Full formalizes knowledge about supply chain operations, by identifying and aggregating common enterprise notions. It is using those to define the semantics of chosen generalizations, namely, the notions of Course, Setting, Quality, Function and Resource.

SCOR-Full ontology does not aim at formalizing the supply chain, but only to resolve semantic inconsistencies of a SCOR reference model. Thus, its scope is strictly limited to using the common enterprise notions for expressing the existing elements of SCOR model. Central notion of the SCOR-Full ontology (as it is the case for SCOR model) is a generalization of process, in the sense that it acts as the main context for semantic definition of other concepts in the ontology.

In SCOR-Full ontology, "Agent" is the concept which describes an executive role and entails all entities which perform individual or set of tasks within the supply network, classified with the concepts of equipment, organization, supply chain, supply chain network, facility and system. "Course" classifies prescriptions of ordered sets of tasks: activity, process, method, procedure, strategy or plan. The notion of Course generalizes “doable” things with common properties of environment (enabling and resulting states, constraints, requirements, etc.), quality (cost, duration, capacity, performance, etc.) and organization (agent and business function). "Setting" concept provides the description of environment of a course. It aggregates semantically defined features of the context in which course take place – its motivation, drivers

and constraints. Thus, it classifies rules, metrics, requirements, constraints, objectives, goals or assumptions of a prescribed set of actions. "Quality" is the general attribute of a course, agent or function which can be perceived or measured, eg. capability, capacity, availability, performance, cost or time/location data. "Function" concept entails elements of the horizontal business organization, such as stocking, shipping, control, sales, replenishment, return, delivery, disposition, maintenance, production, etc.

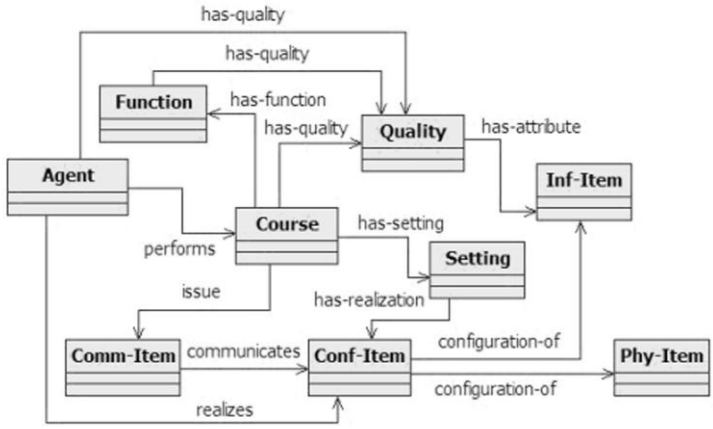

Figure 3. Main concepts of SCOR-Full ontology and relationships between them

Instead of representing process flows, SCOR-Full is used to model enabling and caused states of the relevant activities. These states are represented by the concept of configured item. A resource item is a general term which encloses communicated ("Comm-Item", e.g. notification, response, request) and configured ("Conf-Item", with defined state) information items ("Inf-Item"), such as order, forecast, report, etc., and physical items ("Phy-Item"). Where information items are the attributes of a quality, their configurations are realizations of the rules, metrics, requirements, constraints, goals or assumptions of a course. Configured items model state semantics of the resource – physical or information item, the notions which are used to aggregate the atomic, exchangeable objects in enterprise environment, and are characterized by their states. Examples of information items are Order, Forecast, Budget, Bill-Of-Material, etc. Their semantics is not addressed by SCOR-Full ontology. From this perspective, these are the atomic concepts which can be semantically defined when mapped to other enterprise ontologies.

2.4. Application ontologies

One of the layers of the formal framework presented in this paper consists of the application ontologies. These are considered as formalizations of design goals and are usually used as software application meta-models. Layering of application and domain representation models reflects the paradigm of separation of domain and task-solving knowledge [15] and assume their mutual independence [16]. Thus, arbitrary design goals can be defined, formalized to set of competency questions and used for development of a task-solving, application ontology. In our framework, we developed application ontologies for supply chain

processes generation, product data acquisition and goal reconciliation. Those are shortly described below.

SCOR-CFG ontology [13] is designed with a goal to address the problem of generation of a SCOR thread diagram - a standard tool used in implementation of a SCOR model. In this case, it is inferred as configuration of source, make and deliver processes, on basis of asserted product topology, participants and production strategies for each component. Different process patterns (and roles) are inferred as a result of SPARQL queries executed against SCOR-CFG model in each of the three possible manufacturing strategies: made-to-stock, made-to-order or engineered-to-order. Main features of the corresponding semantic application are: a) development of complex thread diagrams (including horizontal organization of individual supply chain actors); b) generation of process models (including SCOR PLAN activities) and c) workflows and generation of implementation roadmap (with best practices, systems, resource tracking and performance measurement).

Another example of the application ontology we use in our framework is product ontology for inter-organizational networks [17]. It aims at facilitating cooperative response in product information acquisition and management and is used in a process of semantic alignment of two perspectives of the product information - design and functional. The process is expected to decrease human intervention in product data exchange in networked environments, and to create added value, through possible recognition of design intent and automated referencing to related manufacturing competences. Current prototype of system comprises of interfaces for topological model submission and semantic refinement.

Third application ontology facilitates the process of goal reconciliation in supply chain networks. Where supply chain has a singular objective, its actors are individually characterized by different goals, not necessarily compatible with the overall objective. Misalignment of individual goals and objectives can have a negative impact on the capability of an enterprise to act upon its business strategy, when the enterprise is involved in more than one supply chain. SCOR-GOAL ontology formalizes the notions of cooperative goals, hard and soft goals and other relevant terms and is expected to facilitate coherent, system-wide decision making process, by providing the guidelines for individual enterprises' goals reconciliation. Also, it is expected to drive future research on using the intelligent software agents in SCM.

2.5. Local ontologies

One of the major challenges in the efficient use of computer systems is interoperability between multiple representations of reality (data, processes, etc.) stored inside the systems and reality itself – systems' users and their perception of reality. Where latter can be formalized by the domain ontologies, as shared specifications of the conceptualizations, former relies upon the local ontologies – wrappers for heterogeneous sources of information, business logic and presentation rules.

In our work, the range of semantic interoperability is set to EISs. Their conceptualizations are made on basis of the business logic, which is usually hidden in the actual code, and data model, represented by the corresponding database schema. We consider EIS's databases as legitimate starting point for building a relevant local ontology. Obviously, business logic which is encapsulated in the EIS will remain hidden – only underlying data model is exposed by ontology. The exceptions are database's triggers, which can be considered as business rules, if they are not implemented only to enforce referential integrity of the database. In order to enable the implementation of local ontologies within the formal framework, we developed a

new approach to database to ontology mapping, which overcomes the weaknesses of existing methods by using the full expressivity of OWL for making the implicit semantics of the database schema - explicit and enable translation of semantic to SQL queries.

Generation process consists of 4 phases: a) data import and classification of ER entities; b) classification (inference) of OWL types and properties; c) lexical refinement; d) generation of local ontology; and is illustrated on Figure 4, below. The process is supported by a web application, which consists of modules for data import/assertion of ER meta-model instances, lexical refinement and transformation of classified OWL types and properties to a local ontology.

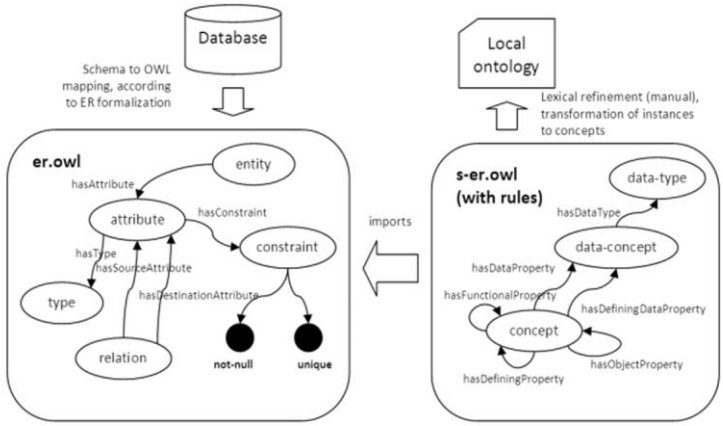

Figure 4. Approach to database-to-ontology mapping

First, database schema is investigated and OWL representation of the ER-model is constructed. This is realized by developed application, which connects to the database, uses introspection queries to discover its structure and asserts the relations between the artifacts by using proposed ER formalization (er.owl).

Second, resulting (serialized) OWL representation of the database ER-model is imported into meta-model (s-er.owl), which classifies future OWL concepts and domains and ranges of the object and data properties, according to defined rules. In this phase, approach takes into account existential constraints from the ER-model. They are associated to an explicit semantics of the resulting ontology, namely, necessary conditions for entailment of the corresponding concepts. According to these constraints, rules for intensional conceptualization (necessary conditions, or inherited anonymous classes) for particular entity are inferred. Also, the approach considers functionality of the properties. Functional property is property that can have only one (unique) value y for each instance x. They are classified when relation one-to-one is identified between two concepts.

The rules classify instances of the OWL representation of the database ER model (er.owl) into a meta-model (s-er.owl). Inferred triples can be edited in a simple web application, which also launches the process of local ontology generation. In this process, meta-model entities are transformed into corresponding OWL, RDF and RDFS constructs – a resulting local ontology.

Concepts of the generated local ontology are annotated with URI's of the corresponding ER entities from er.owl model. Thus, translation of semantic to SQL queries becomes possible.

The approach above is implemented on the case of OpenERP EIS. OpenERP is an open source suite of business applications including sales, CRM, project management, warehouse management, manufacturing, accounting and human resources. It uses PostgreSQL relational database for data storage and application server for enterprise logic. Fig. 5 shows the portions of generated local ontology in Protégé.

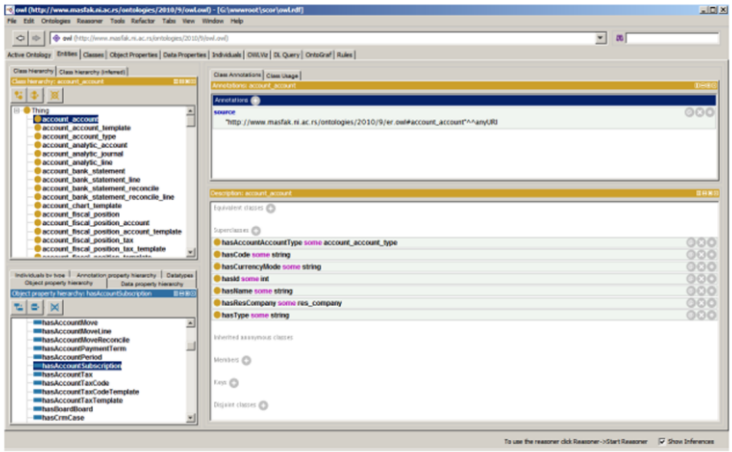

Figure 5. Generated local ontology in Protégé

One of the benefits of the semantically interoperable systems (see Fig. 1) is the possibility to use the single criterion (or criteria) to infer the statements that hold true in all these systems, despite their heterogeneous structure. Namely, specific semantic query executed against the local ontology O_{Li} would normally infer triples of information from the database of S_i. However, if mappings (or their logical function) between O_{Li} and O_{Lj} exist, inferred triples will also include information from the database of S_j. For example, in supply chain networks, a single semantic query can be used to find out the availability of specific resource or competence, of all - owned and used by the enterprises from the Virtual Breeding Environment.

2.6. Queries' translation

Semantic query can be considered as a pair (O, C), where O is a set of concepts which need to be inferred and C - a set of restrictions to be applied on their properties, namely value and cardinality constraints. This consideration corresponds to a simplified representation of a SQL query which includes tables (and fields) and comparison predicate, namely restrictions posed on the rows returned by a query. In addition, different types of property restrictions correspond to different cases (or patterns, where complex semantic query is mapped) of SQL queries.

Where relevant entailments can be reasoned only by property domain and range inferences, a set C may be considered as sufficient for representation of the semantic query. For example, in openERP ontology (see Fig. 5), a DL query "hasAccountAccountType some (hasCode value 3)" returns all instances of account_account concept whose type's code is exactly 3. This kind

of query representation (only by using properties restrictions) may produce unpredictable and misleading results where the restrictions are posed on the common lexical notions of different concepts, such as “name”, “type”, “id”, etc. Ambiguity of the corresponding properties is reflected on the relevant ontology in the sense that their domains are typically defined as union of large number of concepts. However, this may be considered as an advantage in some cases. Value restrictions on ambiguous data properties may produce relevant inferences and thus, facilitate semantic querying without a need to have extensive knowledge on the underlying ontology structure. This kind of query is mapped to a SQL UNION query which combines SELECT sub-queries made on the each element of the property domain, with the WHERE statement corresponding to the relevant rows restrictions. When corresponding element of the UNION query is assembled, a static field with appropriate label (a reference to the concept) is added to each of the elements, so as to become possible to determine which sub-query actually returned the results.

In the first step, decomposition and semantic analysis of the input query is performed. The 4-tuplets in forms of (subject predicate some|only|min n|max m|exactly o bNode) and (subject predicate value {type}) are extracted from the input query. In case of the DL query which returns all concepts which are related to a company whose primary currency is EURO (“hasResCompany some (hasResCurrency some (hasName value "EUR"))”), following 4-tuplets are identified:

X hasResCompany some bNode1

bNode1 hasResCurrency some bNode2

bNode2 hasName value "EUR"

Next, a database connection is established and SQL query is constructed and executed for each 4-tuplet, in reverse order, as a result of above analysis. Each query returns data which is used to generate OWL statements which are asserted to a temporary model. Each set of the OWL statements corresponds to a sub-graph whose focal individual is an instance of the concept, inferred on basis of the 4-tuplet’s property domain or returned result (label). Other individuals or values correspond to defining properties of this concept (inherited anonymous classes). In case of ambiguity, resulting blank nodes are represented as the sets, which are filtered as a result of range inference of the parent 4-tuplet, in a final stage of the method.

3. CONCLUSIONS AND FUTURE RESEARCH

In this position paper, a research of semantic interoperability in supply chain networks is presented. It is based on formalization of widely adopted supply chain process reference model and includes development of its OWL representation, semantically enriched model and correspondences with other models. It transforms implicit semantics of the reference model to the explicit specification which uses common enterprise notions, assumingly defined in other domain ontologies and/or conceptualizations of relevant enterprise models, architectures and frameworks. Used formalization approach is characterized by the multiple, cross-referenced levels of abstraction, represented by the OWL models of different expressivity. Modular design contributes to the usability of the ontology framework, by facilitating the maintenance, avoiding performance related problems in reasoning and by providing increased potential for

ontology matching. Thus, it is expected to facilitate the semantic interoperability in supply chain networks.

With regard to local ontologies generation, significant research efforts are needed for representation and exposition of the enterprise business logic, which is hard-coded in the systems, as well as the semantics of the instances, namely information which is stored in the database (for example, occurrence patterns). Another line of research in the future will aim at enactment of the generated ontologies, as they are considered only as intermediary models. Enactment will also facilitate the process semantic matching of the local ontologies with SCOR-Full. At the explicit side of the semantic framework, considerable efforts are needed for formalization of what is considered as domain state-of-the-art knowledge – architectures, standards, models, etc. of the enterprise, and mapping with SCOR-Full, in order to improve the expressivity and completeness of the framework and delivery of complete SCM domain ontology. Finally, in the implementation context, evaluation of the approach, by using the definition of the semantic interoperability and selected cases will bring evidences on its feasibility and validness.

We consider the research directions above as important for increasing collaboration in a supply chain network, as the fulfillment of the above objectives will facilitate logic driven, automatic and transparent decision making, thus, enabling a transition from traditional supply chains to virtual enterprise and related paradigms.

REFERENCES

[1] J.W. Forrester, “Industrial Dynamics”, The M.I.T. Press, Cambridge, MA, (1961)

[2] Lamber, D.M., Knemeyer, A.M., “We're in This Together”, Harvard Business Review on Supply Chain Management, December 2004

[3] Jespersen, B.D., Larse, T.S., “Supply Chain Management - in Theory and Practice”, Copenhagen Business School Press, 2006

[4] Williamson, O., “The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting”, The Free Press, New York, 1985

[5] Browne, J., Zhang, J., “Extended and virtual enterprises – similarities and differences”, International Journal of Agile Management Systems, Vol.1, No.1, pp.30-36, 1999

[6] Sánchez, N.G., Apolinar, D., Zubiaga, G., Atahualpa, J., González. I., Molina, A., “Virtual Breeding Environment: A First Approach to Understanding Working and Sharing Principles”, In Proceedings of InterOp-ESA’05

[7] H. Panetto, A. Molina, “Enterprise integration and interoperability in manufacturing systems: Trends and issues”, Computers in Industry, Vol.59, No. 7, pp.641-646, 2008

[8] Q. Li, J. Zhou, Q.R. Peng, C.Q. Li, C. Wang, J. Wu, B.E. Shao, “Business processes oriented heterogeneous systems integration platform for networked enterprises”, Computers in Industry, Vol.61, No. 2, pp.127-144, 2010

[9] H. Panetto, G. Berio, K. Benali, N. Boudjlida, M. Petit, “A Unified Enterprise Modeling Language for enhanced interoperability of Enterprise Models”, In: Proceedings of the 11th IFAC INCOM2004 Symposium, April 5th-7th, 2004, Bahia, Brazil

[10] Obrst, L., "Ontologies for semantically interoperable systems". In: Proceedings of the 12th International Conference on Information and Knowledge Management, New Orleans, USA, 2003

[11] ISO/IEC 2382, 01.01.47 Interoperability

[12] Mailing list of the IEEE Standard Upper Ontology working group, http://suo.ieee.org/email/msg07542.html

[13] Zdravković, M., Panetto, H., Trajanović, M., "Towards an approach for formalizing the supply chain operations", I-SEMANTICS '10 Proceedings of the 6th International Conference on Semantic Systems, 2010

[14] G. Stewart, "Supply-chain operations reference model (SCOR): the first cross-industry framework for integrated supply-chain management", Logistics Information Management, Vol.10, No. 2, pp.62–67, 1997

[15] Gangemi, A., "Ontology Design Patterns for Semantic Web Content", Lecture Notes in Computer Science, Springer Berlin, Heidelberg, Vol.3729, pp.262-276, 2005

[16] Guarino, N., "Understanding, building and using ontologies", International Journal of Human-Computer Studies, Vol.46, No.2–3, pp.293–310, 1997

[17] M. Zdravković, M. Trajanović, "Integrated Product Ontologies for Inter-Organizational Networks", Computer Science and Information Systems, Vol.6, No.2, pp.29-46, 2009

The Future of Web

Jaka Sodnik[1], Matija Jekovec[2], Grega Jakus[1] and Sašo Tomažič[1]
[1]Faculty of Electrical Engineering, University of Ljubljana, Slovenia
[2]Klaro d.o.o., Ljubljana, Slovenia

Abstract: **In this paper we summarize some important features and functionalities of the upcoming web technologies. The common idea of all the proposed standards is the generalization and formal specification of different technologies which have already been in use through different plug-ins and custom developed extensions. The majority of these features will now be implemented in browsers. HTML5 as the new version of HTML (Hyper Text Markup Language) enforces a strict separation of the page content from its style. Changes in style can only be performed by the use of CSS (Cascading Style Sheets) language with the CSS3 version proposed as the next generation. The latter has a modular structure in which different modules define different stylistic features. The development cycles of the individual modules are independent as well as their support and implementation in various browsers. HTML5 also introduces a semantic markup, which can be used for marking the document structure as well as its elements and data. In the new version, a wide support for multimedia and dynamic graphic rendering is expected. In addition, the new version will enable the possibility of storing data locally within the browser or acquiring the information on geolocation. All these and other features will be available through a set of newly defined application programming interfaces.**

1. INTRODUCTION

Hyper Text Markup Language (HTML) is the basic mechanism for defining and describing web pages. It uses markup tags for describing structural semantics of a web page by denoting its elements: sections, paragraphs, headings, tables, lists, interactive forms and others. Elements with their corresponding attributes can be nested one in another, forming a typical tree structure. External resources, such as images, videos and other objects can also be included into web documents thus becoming parts of a web page.

The important good practice in a modern web development is a separation of structure and style of web documents. The general structure of a web pages and its content is defined with HTML, while its final presentation and style are in the domain of CSS (Cascading Style Sheets). Such separation enables better flexibility and control over the final appearance of a web page. It also reduces the complexity of HTML record and enables the reuse of style definitions. This way more web pages can share the same style or a single page can use many different styles simultaneously.

In all modern dynamic web pages JavaScript is a mandatory supplement to HTML and CSS. JavaScript is a scripting language which is interpreted by a web browser and provides web pages with interactivity and dynamics. Its code can interact with the DOM (Document Object Model) through the various API (Application Programming Interface) libraries based on a mechanism of user-triggered events.

In the 1990's majority of web pages were static and intended primarily for reading and browsing while the first decade of the new century brings more dynamic web pages and applications. Users not only "browse" the Web but also contribute to it by producing and up-

loading their own content. The so-called Web 2.0 evolved and brought some major changes also in web development. The new way of interaction with Web calls also for the evolution of web languages with the main intention to formalize some of the already established good practices in web development.

In this paper, we present some important novelties in future web design and implementations. We start with the new version of HTML by presenting its new elements and extensions. We list also some most important JavaScript APIs which enable entirely new way of web development by providing a browser based database, information on geolocation, full-duplex communication between a browser and a server and other new exciting features. We point out also some new options in CSS syntax. The latter introduces many new design options and also simplifies and standardizes some features that have been available before but suffered from poor browser support. It is important to point that the majority of the technologies and features presented here are already supported or at least partially supported by the modern and most popular browsers.

2. HTML5

The majority of HTML features and functionalities have been defined through specifications, but some of them are also result of good development practices and the implementations of HTML in the popular browsers. The current HTML version - HTML4 - has been in use for almost a decade [1].

The development of HTML5 began within the WHATGW (Web Hypertext Application Technology Working Group) initiative and the organization W3C in order to provide better flexibility and interoperability of the HTML implementations and at the same time make web pages more interactive and offering better user experience. The development itself is based on the study of existing HTML4 implementations, good practices and analyses of already deployed web content.

HTML5 [2] will be backward compatible with HTML4 and XHTML1, supporting both, HTML and XML (eXtensible Markup Language) syntax. It will also introduce new interfaces to support contemporary trends, such as rich internet applications (RIA). Currently these interfaces depend strongly on the use of complex JavaScript code and proprietary plug-ins, such as Adobe Flash [3], Microsoft Silverlight [4] and Sun JavaFX [5]. The general idea suggested by the web developers is to implement the key functionalities for such interfaces in browsers themselves and remove the dependence on various proprietary plug-ins.

It is expected that HTML5 will achieve the candidate for recommendation status in 2012 and become a recommendation in 2022 [6]. Although the work on HTML5 will not be completed yet in the next few years, more and more of its functionalities are supported by web browsers [7].

2.1. Changes in the Language

In the majority of modern web pages common structures such headers, footers and sidebars are used to define the semantic structure of the page. Since no special markup is available for this purpose web developers use div and span elements, assigning them a unique id and/or arranging them into classes. HTML5 introduces a set of new elements, which allow semantic

marking of the document structure. They represent more specific replacement for the general *div* and *span* elements. The current and the new approach to structuring a web page are shown in Figure 1.

Figure 1. The current (using div elements) and the new approach (using new elements which are presented in bold text) to structuring a web page

HTML5 introduces some other new elements, with the most interesting being:

- *embed* is used for the embedded content, handled by the plug-ins;
- *video*, *audio* and *canvas*: elements that support multimedia and graphic content;
- elements for the display of quantities (*progress*, *meter*, *time* etc.);

In HTML5 content and styling of a web page are strictly separated which is manifested in the absence of the presentational attributes (e.g. align, height, border, size) and elements (e.g. font, center, strike, u). Page styling and design can only be done with CSS. Frames are also not supported in HTML5, due to their negative impact on the usability of a web page.

The elements *a* and *area* have a new attribute, called ping. It defines the URLs (Uniform Resource Locators), where a browser can send a notification when the user visits a hyperlink. It also enables the user agent to inform the user, which addresses will be notified. In case of privacy concerns, a user can turn off the notifications while he or she can not influence the redirects.

2.2. Web Forms

Web forms enable the data entered by the user to be sent to a server, which responds according to the received values (e.g. logins the user to the system). A good web development practice is the validation of more complex data on the client side by using JavaScript or any other client side scripting languages. In order to simplify the process of data validation a series of new form widgets were developed that can be used through third-party JavaScript libraries.

The developer needs encouraged the development of a new generation of Web forms, called Web Forms 2.0, which found their way into HTML5 specifications. New widgets are introduced as new values of the attribute type of the input element (tel, search, url, email,

datetime, date, month, week, time, datetime-local, number, range, color). Beside new widgets, HTML5 also introduces the enhancements of the existing ones as well as the automatic validation of the entered data.

Two more novelties should also be mentioned in this context. The form elements do not have to be nested inside the form tag anymore. They can be placed at arbitrary position within the HTML document instead and linked to a proper form by using their new form attribute. The second novelty presents the support for more HTTP (Hyper-Text Transfer Protocol) methods. In HTML5, beside GET and POST methods PUT and DELETE are also supported.

```
<input type='color'/>
```

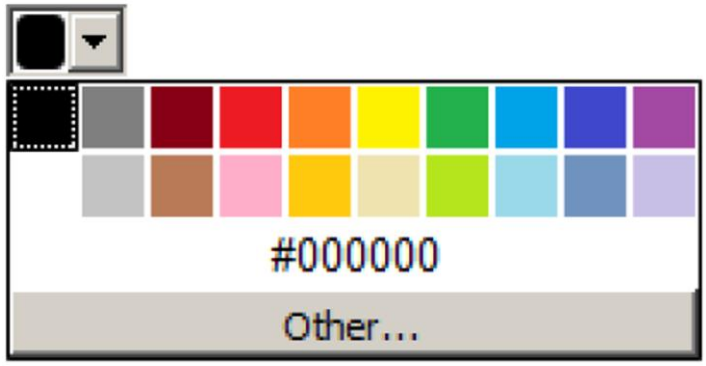

Figure 2. The new color form widget

2.3. Microdata

Microdata [8] in HTML documents enable the information, primarily intended for the end users (contact information, location information etc.), to be read and interpreted by the machines and could therefore be used for automatic processing (e.g. for indexing, search-ing, storing, cross-referencing, analyzing etc.).

The model of microdata consists of groups of properties named items. The items and their properties are presented in the context of the existing elements. An item can be created using the *itemscope* attribute in any HTML element. The *itemprop* attributes with the corresponding values are then added to different item's descendants in order to express their meaning (Figure 3). Additional attributes *itemtype* (in the form of URLs) and *itemid* are used in order to globally define some specific information and to share microdata among different authors and readers.

```
<div itemscope itemtype="http://example.org/superHero">
   <p>My name is
      <span itemprop='firstName'>Chuck</span>.
      <span itemprop='surname'>Norris</span>.
   </p>
   <p>I live in
      <span itemprop='residence'>Texas</span>.
   </p>
</div>
```

Figure 3. An example of using microdata

2.4. The accessibility

The frequently updated advanced user interfaces developed using the combination of AJAX (Asynchronous JavaScript and XML), HTML and JavaScript cause the accessibility problems to disabled users, especially to those that depend on assistive tools (e.g. screen readers). Such tools, namely, do not properly understand the roles, states and properties of components of advanced user interfaces and cannot follow the dynamically updating content on the web pages.

To overcome the accessibility problems, the ARIA (Accessible Rich Internet Applications) [9] project was initiated by the Web Accessibility Initiative (WAI). The main contribution of the project is a semantic model which allows assistive technologies to gather enough information to make advanced web applications usable also to disabled users. The model enables the authors to semantically describe the structure of a web document, the areas of the document that could potentially be updated as also the individual elements and their behavior. Authors can assign the roles the elements have on the page and define their semantic states and properties. While states and properties can change when the page updates, the role of an element always remains unaffected.

2.5. Audio and Video

Displaying of audio and video contents in a browser is presently only possible using third-party plug-ins. The leading technology in the field is the Adobe Flash. By introducing new elements for embedding audio and video contents in a web page and defining an interface for manipulation with such content, HTML 5 brings the end to the dependence on third-party plug-ins.

Nevertheless, the selection of a common format supported by all browsers still remains an open issue. The basic requirement for such format is that it should not be proprietary. It should also have good compression and picture quality as well as small processing requirements. A hardware solution for the decoding should also exist. The W3C organization chose Ogg Vorbis for audio and Ogg Theora for video records, mainly because they are both license-free, which is not true for the rival (and otherwise more popular) MP3 and H.264 formats.

However, the consensus between the W3C and browser vendors has not been reached yet. The actual draft of the HTML5 specifications does not specify the supported formats and therefore, browser vendors can individually decide on the formats built in their browsers. Apple, for example, does not support the Ogg Theora in its Safari browser, due to the lack of the hardware acceleration and some uncertainties regarding the patents. On the other hand, Mozilla Foundation and Opera oppose to the use of H.264 due to the expensive licensing. Google's Chrome supports both Theora and H.264. The latter is supported mostly because it is used by Google's YouTube which is the World's most important video portal.

```
<video src=cars.mp4' autoplay controls />
```

Figure 4. HTML 5 video support

2.6. Graphics

Similarly to the displaying of audio and video contents, the graphic rendering is also currently only feasible using the plug-ins, such as Flash or Silverlight. With the HTML5, the functionality of these plug-ins is implemented in browsers in the form of Canvas and SVG (Scalable Vector Graphics) technologies.

The canvas technology enables dynamic rendering of graphics (e.g. graphs, bitmap images, animations and games) using scripting. The display region is placed into a web page using the *canvas* element and its attributes *width* and *height*. The region is accessible through JavaScript code using the drawing functions of the Canvas API.

The basic elements of canvas graphic are pixels. Because canvas is not aware of the individual graphic objects in the graphic or their inter-relations, such objects cannot be accessed, manipulated or interacted with. The rendered graphic is therefore final and cannot be rescaled. In order to make any changes, the whole graphic must be redrawn.

The basic Canvas API currently enables 2D rendering. 3D rendering will also be possible using the WebGL standard. Although the WebGL is still in the development, it is already experimentally supported in most popular browsers.

Beside Canvas, HTML5 also supports browser built-in rendering using Scalable Vector Graphics (SVG). Nowadays, SVG is used mostly for displaying static contents (maps, plans etc.) with the aid of browser plug-ins. SVG is based on a special XML object model through which the individual graphic objects can be accessed and manipulated with using JavaScript. Unlike Canvas, SVG enables rendering in high resolution at any level of magnification as also the interactivity by using event-handlers assigned to a graphic object.

```
<canvas id="canvas" width="838" height="220"/>
<script type="text/javascript">
        var canvas = document.getElementById('canvas');
        var context = canvas.getContext('2d');
        context.fillStyle = "rgba(100,100,255,0.6)";
        context.beginPath();
        context.moveTo(30,180);
        context.lineTo(210,60);
        context.lineTo(190,220);
        context.fill();

        context.fillStyle = "rgba(255,0,0, 0.8)";
        context.beginPath();
        context.arc(120, 120, 80, Math.PI * 1/8,
        Math.PI * 4/8, true);
        context.lineWidth = 20;
        context.lineCap = "round";
        context.strokeStyle = "rgba(255,0,1,0.9)";
        context.stroke();
        context.fill();
</script>
```

Figure 5. Canvas element and the manipulation of its content using JavaScript

3. PRESENTATION AND CASCADE STYLE SHEETS (CSS)

Markup languages are used primarily to define basic structure of web documents and pages while on the other hand the final presentation and rendering is usually defined with CSS – a style sheet language which defines presentation semantics. The technique of designing web pages with CSS is almost 14 years old with the CSS 2.1 being the leading standard for the last 12 years. It is widely supported by the majority of web browsers and it evolved rapidly over the last decade. Along with the new markup languages the new CSS is being proposed as well. CSS3 is the new potential standard currently in the state of working draft or candidate recommendation. The massive specification of CSS3 standard [11] is divided into several modules which are developed individually with different progress speeds and dynamics. Various modules enable browser vendors to implement them incrementally. Several CSS3 modules are already supported by the majority of modern browsers. The new standard is completely backwards compatible with the addition of new properties and functionalities.

The new Selectors module introduces a variety of new methods for attaching elements to the corresponding style. The most exciting addition is the ability to select markup elements based on their placement in the DOM. The `:last-child`, `:nth-child(n)` and `:nth-last-child(n)` commands enable the targeting of elements based on their positions in a parent's list of child elements. For example the command `:nth-child(3n+2)` would mach a group of three elements after the second element. Other commands enable also the matching of elements which are checked, without children or elements that do not match the specified declaration. The markup elements can also be selected based on the existence of specific attribute and also

just a part of an attribute. For example, the `img[alt*="good"]` would select all images that have an alt attribute containing the word "good".

```
<style type="text/css">
   img {
         border-width: 50px 50px 50px 50px;
         -webkit-border-image:  url(ball.jpg)  100%  100%
100%
         100% repeat repeat;
           }
</style>
<body>
  <div>
          <img src="kobe.jpg">
  </div>
</body>
```

Figure 6. The use of one image as a border of the other image

A module on Backgrounds and Borders enables the use of multiple backgrounds which can be resized and positioned relatively or absolutely. It enables the reuse of images in several different contexts and more accurate filling of various areas. The module enables also the borders to use gradients, rounded corners, shadows and even border images. The *border-image* property allows an image file to be used as border of an object. An example is shown in Figure 6. The *gradient* property enables the borders or backgrounds to shift from one color to another programmatically.

The Color module defines different capabilities for setting all colors in the document. The new *rgba* command enables the specification of the color as well as the opacity of an element. Red, blue and green colors have to be defined with an integer value or with percentages while the alpha value should be between 0.0 and 1.0. For example, the alpha value of 0.8 defines the element with only 20% transparency.

Since now all web designers had to be aware of a set of so called Web-safe fonts. The new Fonts module allows the font file to be included as an external file and accessed through *font-family* property. The new @font-face rule allows the fonts to be called from an online directory, such as for example the command

```
@font-face { font-family: 'myFont'; src: url (../myFonts.ttf') format
('truetype'); }.
```

If the command is not supported by the browser it reverts to the next specified font in the *font-family* property. The main issues with new font types are font licensing and the copyright. The embedded fonts can easily be downloaded from any page. The two exiting new modules are also Transitions and Animations which enable the changes and movements of page elements in 2D or 3D space. With transitions the developer can specify the specific CSS property to animate from one state to another in a smooth transition. For example, a click on an image could trigger a change of the size of that image. The animations on the other hand

enable the iteration of multiple CSS properties simultaneously for a number of times. The animation can be divided into stages using keyframes. Individual stages of elements between keyframes are approximated automatically by the browser.

4. JAVASCRIPT APIS

HTML5 introduces a number of APIs that standardize features somewhat present in today's browsers. Support for these features can be tracked on sites like [7] or [12].

4.1. Offline features

The Web has evolved from a Web of simple hyper-linked documents to an ecosystem of web applications running on a range of devices. As applications are becoming similar to desktop applications they tend to require similar facilities.

Detecting if a user agent is connected to the Internet is described in W3C's Offline Web Applications [13] proposal. By using “online” and “offline” events dispatched by the Window object or querying status of the *onLine* attribute of the Navigator object, a developer can determine status of the connection.

In recent past, using cookies was the only way to store data locally [14]. Several drawbacks such as limited size (in the range of few kB) and performance issues (they are sent over the network for each request) limits the use of cookies for storage in real-world applications.

HTML5 introduces the idea of simple storage with the Web Storage draft [15] that recognizes local storage (data persists even if browser is closed) and session storage (data is destroyed on along with the session). The biggest critique of this storage mechanism is a security of such storage which depends solely on the user agent. Web storage is implemented in most major browsers with a varying degree of adherence to standards [12].

SQL compliant database was also proposed to W3C for standardization, but the draft was frozen as it was too similar to SQLite[17] and no other implementations existed. Another W3C draft for advanced storage is called Indexed Database API [18] (renamed from Web Simple DB [19] in late 2009). This draft proposes storage based on key/value pairs but has advanced functionalities for querying and storing data that developers working with modern RDBMSs (Relational Database Management System) are used to. Currently Firefox 4 beta and Chromium dev (an open source version of Chrome browser) have partial support for this standard [20] while most browsers support SQLite [21].

Beside the object level access to storage, some drafts have also been developed to allow access to the file system of the application. A basic File API [22] and extensions File API: Directories and System and File API: Writer define how a web application is allowed to access the file system. While basic “File API” is used to define how to list files and retrieve file data from files selected through the File Upload functionality of the user agent, “Directories and System” draft also defines ways to provide complete read-only access to the underlying system. “Writer” draft also defines methods for writing data to the underlying file system. In contrast with other storage methods, access to file system is regulated through the “Permissions API”.

4.2. Mobile device oriented specifications

HTML has become an increasingly popular technology on mobile devices and an answer to cross platform problems on mobile devices. As operating system market on mobile is fragmenting (iOS, Android, MeeGo, Symbian, Windows Phone 7), HTML technology ecosystem is the only real alternative for cross-platform applications.

Some limited progress towards integration with device calendars, camera, contacts, file system and location has also been made with the Device API Requirements draft [22] but more specific drafts are being developed. Messaging API is a draft that will try to standardize the way web applications communicate with the device to send and receive SMS, MMS and e-mail messages. Additionally, Contacts API is designed to access and manage device's contacts and Media Capture API defines how to do sound recording, photographing or filming using a web application on a device.

When determining the location of a device, software does not rely on a single, but a range of technologies including GPS (Global Positioning System), Wi-Fi, RFID (Radio Frequency Identification) and mobile radio technologies. The Geolocation API [23] candidate recommendation document was created to standardize the way web application retrieve or track location of the host device.

```
//output last cached location
navigator.geolocation.getCurrentPosition(
    function(pos){
        alert("lat:" + pos.coords.latitude
            + ", long:" + pos.coords.longitude)
    }
);
```

Figure 7. Example of using geolocation API [23] through Navigator object by retrieving last cached location

4.3. User agent multithreading

While JavaScript within browser supports asynchronous communication with servers, the user interface is always interacted with in a single thread. This means that while computationally intensive tasks can be done on the server without slowing down the user interface, there is no way to do them within a browser.

Web Workers draft is designed to cope with this issue [24]. The idea behind web workers is to allow background execution of time consuming or computationally demanding tasks without interrupting or slowing down user's interaction with the browser. Two types of web workers are defined: *normals* that are accessible only from the current window (or tab) of a user agent and *shared* web workers, that can be accessed by any instance of web application running in separate tabs.

4.4. Two-way communication between a web browser and a server-side process

In the original model of the Web, a browser always requests a web page or parts of a web page. HTTP 1.1 specifications limit browsers to maximum of two simultaneous connections with a web server which enables faster loading and rendering of pages with multiple pictures or multimedia content. In the past the "Comet" model has changed this classical model by keeping the one of the two connections alive permanently for real-time data exchange with the

server. A web server can use this connection to push data to a browser without the browser to explicitly requesting it. In some cases just one long lasting connection can be used instead of two simultaneous connections. There have been several different implementations of this technique but all of them rely on AJAX and its XMLHttpRequest object.

The HTML5 WebSocket protocol [25] is a new uniform technique for pushing data to browsers. It is a part of HTML5 specification and it is intended to be used within scripts in web pages. It provides a full duplex connection without the need for multiple HTTP connections. A WebSocket request is initiated as a standard HTTP request and then upgraded to the WebSocket protocol with the special initial handshake between the client and the server. The latter establishes a persistent HTTP connection which can be used for sending data in full duplex mode. The communication is handled through JavaScript which requires the data to be in a text-based format.

WebSocket protocol places much less burden on web servers and enables the existing machines to handle twice the number of simultaneous connections. Another great benefit of the protocol is its ability to traverse firewalls and proxies. It detects a proxy server and automatically establishes a tunnel to pass through the proxy by issuing an HTTP CONNECT statement. The latter requests the proxy to open a TCP/IP connection to a specific host. The similar solution works also for establishing secure connections (Secure Sockets Layer, SSL).

5. CONCLUSION

The general idea of HTML5 and accompanying standards presented in this paper is the formal specification of functionalities needed to further extend and enriching the Web. Although some functionalities are already implemented, its' implementations vary in completeness between different user agents.

By defining native video and audio support, user agents do not rely on proprietary plug-ins like Flash to deliver multimedia. Javascript toolkits that extend HTML functionality to support desktop-like applications will not be needed when standards get implemented in major browsers. Fear of growing fragmentation in the mobile OS market is probably one of the reasons behind the latest HTML related drafts that address functionalities commonly found on mobile devices (contacts, messaging, geolocation …).

As more and more "users" of the Web are machines, and the amount of data continues to rapidly grow, HTML needed to adapt by introducing special markup to convey semantics to computers. The so-called Semantic Web [26] is expected to be the next step in the evolution of the Web where data will not be only human, but also computer readable.

REFERENCES

[1] HTML5 differences from HTML4, W3C Working Draft, 13. 1. 2011, http://www.w3.org/TR/html5-diff/, acc. 25. 1. 2011.

[2] HTML5, A vocabulary and associated APIs for HTML and XHTML, W3C Working Draft, 13. 1. 2011, http://www.w3.org/TR/html5/, acc. 25. 1. 2011.

[3] Adobe Flash, http://www.adobe.com/flashplatform/, acc. 25. 1. 2011.

[4] Microsoft Silverlight, http://www.silverlight.net/, acc. 25. 1. 2011.

[5] JavaFX, http://javafx.com/, acc. 25. 1. 2011.

[6] WHATWG Wiki FAQ, http://wiki.whatwg.org/wiki/FAQ, acc. 25. 1. 2011.

[7] WHATWG Wiki, Implementations in Web browsers, http://wiki.whatwg.org/wiki/Implementations_in_Web_browsers, acc. 25. 1. 2011.

[8] HTML Microdata, W3C Working Draft, 24. 6. 2010, http://www.w3.org/TR/microdata/, acc. 23.1. 2011.

[9] Accessible Rich Internet Applications (WAI-ARIA) 1.0, W3C Working Draft, 15. 12 2009, http://www.w3.org/TR/wai-aria/, acc. 25. 1. 2011.

[10] HTML5, Web Development to the next level, html5rocks.com, http://slides.html5rocks.com/, acc. 25. 1. 2011.

[11] Cascading Style Sheets, Current Work http://www.w3.org/Style/CSS/current-work, acc. 25. 1. 2011.

[12] Offline Web Applications, W3C Working Group Note, 30. 5 2008, http://www.w3.org/TR/offline-webapps/, acc. 25. 1. 2011.

[13] Peng, W., Cisna, J., "HTTP cookies – a promising technology", Online Information Review, Vol. 24, No. 2, 2000.

[14] Web Storage, Editor's Draft, 20. 1. 2011, http://dev.w3.org/html5/webstorage/, acc. 25. 1. 2011.

[15] Trivero, A. "Abusing HTML 5 Structured Client-side Storage",http://trivero.secdiscover.com/html5whitepaper.pdf, acc. 23.1.2011.

[16] Web Designer's Checklist, http://www.findmebyip.com /litmus/, acc. 25. 1. 2011.

[17] SQLite Home Page, http://www.sqlite.org/, acc. 25. 1. 2011.

[18] Indexed Database API, W3C Working Draft, 5. 1. 2010, http://www.w3.org/TR/IndexedDB/, acc. 25. 1. 2011.

[19] WebSimpleDB API, W3C Working Draft, 29.9.2009, http://www.w3.org/TR/WebSimpleDB/, access. 25. 1. 2011.

[20] https://developer.mozilla.org/en/IndexedDB, acc. 25. 1. 2011.

[21] Ranganathan, A. "Beyond HTML5: Database APIs and the Road to IndexedDB, Mozilla.org", http://hacks.mozilla.org/2010/06/beyond-html5-database-apis-and-the-road-to-indexeddb/, acc. 25. 1. 2011.

[22] http://www.w3.org/TR/2010/, acc. 25. 1. 2011.

[23] Geolocation API Specification, W3C Working Draft, 7. 9. 2010, http://www.w3.org/TR/geolocation-API/, acc. 25. 1. 2011.

[24] Web Workers, Editor's Draft, 20. 1. 2011 http://dev.w3.org/html5/workers/, acc. 25. 1. 2011.

[25] The Web Sockets API, W3C Work. Draft, 22. 12. 2009, http://www.w3.org/TR/websockets/, acc. 25. 1. 2011.

[26] Berners-Lee, T., Hendler, J., Lassila, O. "The Semantic Web", Scientific American Magazine, 2001.

The Networking Protocol Interactions for the Internet of Things

Carlo Fischione, Piergiuseppe Di Marco, Pangun Park
ACCESS Linnaeus Center, Electrical Engineering,
Royal Institute of Technology, Stockholm, Sweden

Abstract: **Wireless medium access control (MAC) and routing protocols are the basic building blocks making it possible the Internet of Things (IoT). These technologies have been the focus of substantial research in the last decade. Nevertheless, as new networking standards are being proposed and different existing solutions patched together, evaluating the performance of the network becomes challenging. Specific solutions that were supposed to be individually efficient, when stacked together may have detrimental effects on the overall network. In this paper, an overview of the fundamental MAC and routing IoT protocols, the IEEE 802.15.4 MAC and IETF RPL, and their cross-layer interactions are discussed. It is argued that a mathematical description is essential to truly understand the protocols mutual effects and their dynamics. Analytical and experimental results show that the MAC protocol may hurt the routing protocol, unless these two are carefully optimized together.**

Index Terms**: Internet of Things, Medium Access Control, Routing.**

1. INTRODUCTION

Internet of Things is a new technology for which any physical object equipped with wireless communication, and thus often referred to as "smart object", can be reached via the Internet. Potentially, any Internet terminal such as a laptop, a mobile phone, or a server can close the loop with such smart objects if the Internet Protocol is used. A variety of new applications is expected to flourish by the IoT vision, including building and industrial automation, healthcare, smart grids, and security. To support such an increasing number of applications, the wireless medium access control (MAC) and routing protocols must be inherently efficient, scalable, interoperable, and have a solid standardization base.

There is a consensus that the protocols for the physical layer and MAC for low data rate and low power applications will be based on the flexible IEEE 802.15.4 standard with its useful and numerous variants [1]. The standard has been adopted with some modification also by a number of other solutions, including ZigBee, WirelessHART, ISA-100, and according to recent surveys, it already represents more than 50% of building and industrial automation market [2].

By contrast, routing protocols are still being under standardization. The Internet Engineering Task Force (IETF) is currently working for creating a reference standard for IoT, the IETF Routing Protocol over Low power and lossy networks (RPL) [3]. This new standard proposal is supposed to be compatible with many existing MAC protocols, particularly the IEEE 802.15.4 MAC. However, we believe that it may be inefficient to design MAC and routing separately. Their complex interactions and dynamics must be considered altogether for choosing adequately the networking resources and optimize typical performance indicators. For example, for IoT control applications, the process state information must be transmitted to

the controller over a network guaranteeing some delay and reliability so the stability of the closed loop is ensured while making an optimal use of the network resources.

In this tutorial paper, we summarize the two essential protocols for IoT networking, IEEE 802.15.4 and IETF RPL, we then give an overview on an analytical method to characterize MAC and routing interactions. We conclude that it is challenging to satisfy desired levels of performance with the current state of the protocols, unless their dynamics and interactions are properly jointly understood and regulated.

2. THE MAC AND ROUTING INTERACTIONS

The interaction between MAC and routing is illustrated through the feedback loop in Fig. 1. The *application requirements* are what the IoT application asks to the networking protocols. These requirements include end-to-end delay or reliability with which information must be received. The network topology (e.g., number of nodes and connectivity graph) and the traffic generated by the nodes of the network impact the MAC and routing performance. Routing decisions, i.e., the path choices over which route information, determine the distribution of the traffic over the network. These decisions consider the application requirements and locally available network information called *routing metrics*. Typical metrics are the end-to-end reliability, the end-to-end delay, the total energy consumption, and the router congestions that are expected to be experienced while routing information to destination over a given path. Notice that the application requirements are what the IoT application demands, where the metrics are routing rules based on what is actually experienced. Routing metrics are given by the sum of per-link performance indicators, which are determined by the MAC. It follows that the routing metrics are directly influenced by the performance of MAC layer. For example, consider node V_s generating a packet (we retain the node's notation of the IETF RPL documents [3]), V_r relaying such a packet and V_d receiving it as a destination. The probability that the packet reaches V_d is $R_{s,r} \times R_{r,d}$, where $R_{s,r}$ and $R_{r,d}$ is the reliability over the link V_s to V_r and V_r to V_d, respectively. Same considerations hold for the end-to-end delay, energy consumption, and congestions.

Since different MACs provide us with different per-link performance, if the routing is designed or selected regardless the MAC, the closed loop system of Fig. 1 may result highly unstable in the sense that end-to-end performance indicators could be far from what asked by the IoT application. Same problem shows up when MACs protocols are selected regardless of the routing. The wireless channel is shared, thus a choice of a MAC for a link between two nodes may reflect in unexpected packet collisions or interferences over other links. Therefore, highly efficient networks for IoT pose the question of how to put together MAC and routing protocols.

We advocate a design method for which the protocol layer interactions are mathematically described as function of the protocol MAC and routing parameters, namely the free parameters that can be regulated to influence the protocol performance. It is only by the availability of such descriptions that the dynamics between the networking layers and the IoT applications can be properly understood and predicted, the closed loop system of Fig. 1 is well behaving, and the network performance can be maximized. The idea of analytical description and optimization of the cross layer interactions is not new, but as a matter of fact, the standardization processes does not appear to consider it.

In this talk, we show the choices available for the MAC and routing selections, and we illustrate the analysis by IEEE 802.15.4 and IETF RPL.

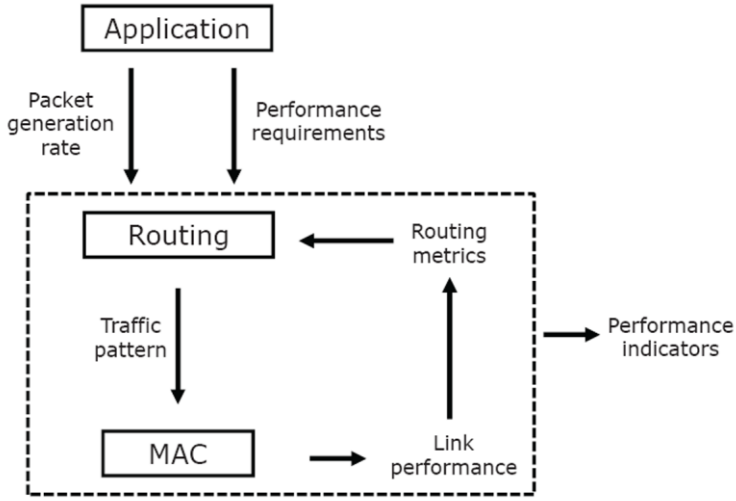

Figure 1. The loop of MAC and routing interactions. The closed loop can lead to instabilities if these interactions are not considered when selecting the protocols

3. MEDIUM ACCESS CONTROL PROTOCOLS

We classify MAC for IoT into three categories based on the access mechanism, in contrast to previous surveys from the literature [4]–[6] that categorize MAC protocol according to some specific energy-saving technique: *schedule*, *contention*, and *hybrid*-based access. The motivation is that in IoT, not all the nodes of the network need to be necessarily energy efficient. Observe that every MAC falling in one of the classes will offer its own performance metrics and its own parameters, which could be greatly different from protocol to protocol.

- Schedule-based MAC: Time is divided into time slots. Nodes wake up and listen to the channel during their assigned slots and then go back to sleep in other slots to save energy. The network topology must be centrally known to establish a schedule without having interference with other transmissions. There is a high complexity and high cost for the scheduling, a low adaptability for dynamic topologies, and a overhead due to time synchronization.
- Contention-based MAC: These MAC protocols differ from schedule-based approaches because there is no centralized scheduler. Nodes listen if the channel is free, in which case transmit, otherwise randomly defer the transmission. The MAC is simple and flexible and does not require either global synchronization or topology knowledge. An effective enhancement to save energy is the duty-cycling between sleep/active periods. The throughput is low when traffic load increases due to high contentions. A mathematical understanding of the performance of these MAC protocols is challenging.
- Hybrid-based MAC: They combine the advantages of both a random access with contention-based MAC and a deterministic access with schedule-based MAC to offer a flexible quality of service for several classes of applications. The pros of this

MAC are a medium complexity, a high throughput, and high flexibility. The optimal size of active period is a function of traffic load, network topology, and hardware specifications which appears difficult to characterize mathematically due to the coexistence of the deterministic behavior of scheduled-based and the stochastic behavior of contention-based MACs.

4. ROUTING PROTOCOLS

Many routing protocols have been proposed in the literature [7]. Routing protocols of relevant interest for the IoT can be classified as follows:

- Topology aware. Global network information is needed to route packets. Routing tables are computed off-line and/or updated runtime. They are scalable and widely available, but have a high routing overhead.
- Content aware. A data centric approach is used, where routing is based on IoT application data information. All nodes can receive and send information (e.g., SPIN, Directed diffusion). The routing is a simple and efficient solution, with a low scalability and adaptation to dynamic networks. However, it is sensitive to the traffic pattern.
- Location aware. Nodes know their own and their neighbors position, for example by GPS location (e.g., GAF), and the routing is based on the position. Packets me be forwarded also by using position-informed randomization (e.g., GeRaF). The advantage is a low computation and overhead. However, this routing solution may be energy consuming. Localization is critical, and might be inaccurate indoor.

5. PUTTING MAC AND ROUTING TOGETHER

This section puts together the previous two sections to give a characterization of the MAC and routing interactions. We consider the unslotted IEEE 802.15.4 MAC and the IETF RPL specifications, as described above. Given the availability of many MAC and routing solutions, with their complex performance and adaptation parameters, a natural question is how to choose the best protocols and how to adapt their parameters to meet desired IoT application requirements. To answer this question, we advocate a new approach to the protocol implementation based on the mathematical modeling of the protocols, as we see next.

5.1. Joint Modeling

We can represent the dynamical interaction of MAC and routing using a statistical model. Consider the simple network in Fig. 2. Let $_j$ be the neighboring set of node V_j . Let Q_j be the traffic generated plus the traffic to forward in the multihop routing. Let $\_$ be a vector of node traffic generation rates, where each component is associated to a node. In addition to $\_j$, node V_j has to forward traffic generated by nodes in its children set $\_j$. Recall that $\_{i,j}$ is the metric associated to link between node V_i and V_j to build the DODAG.

The RPL path selection is modeled by a real valued matrix **M**, in which element $M_{i,j}$ corresponds to the probability that the metric $\_{i,j}$ in link (i, j) is the highest among the set of candidate receivers $\square i$. The distribution of the traffic flows along the network is modeled by the matrix **M**, and by a scaling due to that only successfully received packets are forwarded. Therefore, we define a matrix **T** such that $T_{i,j} = M_{i,j}R_{i,j}$ where $R_{i,j}$ is the reliability in the link (i,

j), which clearly depend on the traffic rate of the node. It follows that the vector of node traffic generation probabilities Q is the solution of a system of flow balance equations Q = Q**T** + _. In steady state, we have

$$Q = \_ [\mathbf{I} - \mathbf{T}]^{-1}, \quad \text{(V.1)}$$

where **I** is the identity matrix and $^{-1}$ denotes the matrix inversion, which always exists because **T** is stochastic. Eq. (V.1) gives the fundamental relation between the packet generation rate, the effect of routing (via **M**) and the performance at MAC layer (via the link reliability $R_{i,j}$). Eq. (V.1) together with the expressions for the per-link reliability $R_{i,j}$, as obtained by the mathematical modeling we proposed in [2], [8], gives the distribution of the traffic in the network. In practice, Eq. (V.1) is the fundamental equation to model mathematically the joint effects of MAC and routing. The i-th component of the vector Q is the amount of traffic that a node has to forward to its parent per unit time. Such a traffic is handled by the MAC and thus determines the per-link performance such as delay, reliability, and energy consumption. inally, the end-to-end reliability of a node is the product of each link reliability in the path to the root. Similarly, the end-to-end delay is the sum of the delays in the path from the transmitter to the root node. The validity of Eq. (V.1) is shown in Section V-C by experiments.

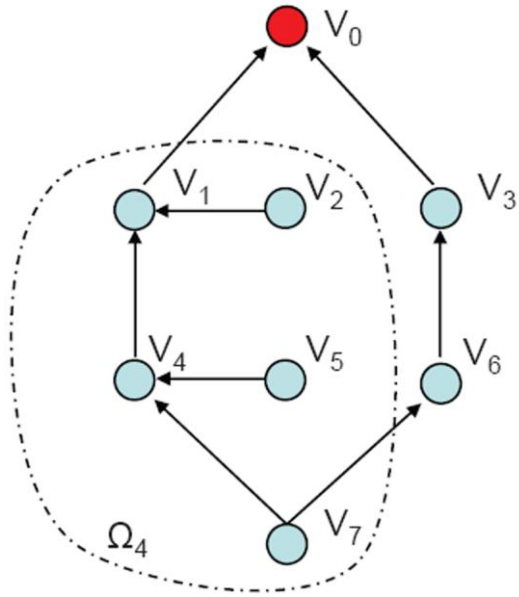

Figure 2. The shadowed area in this network represents the sensing range of node 4. Every node in the shadowed area is heard by node 4, which means that the MAC performance (the per-link performance over the link 4–1) is affected by these nodes and those who are hidden

5.2. Protocol Selection

A protocol selection should be based on a library of compatible MAC and routing protocols. The selection should pick the best protocol among the library, with optimal parameters for the specific class of IoT applications. The selection takes as input the network topology and packet generation rate, the application requirements for reliability, delay, and

energy consumption, and the resource constraints from the underlying physical layer. Every choice of the optimized MAC solution and routing metric will give a different value in (V.1), whereby the end-to-end delay, reliability, and energy can be computed before hand. Then, the protocol engine chooses the optimal protocol among its library with optimal parameters for the specific topology and packet generation rate that gives the best performance as predicted by (V.1). Thus, the MAC and routing performance can be evaluated altogether.

5.3. Experimental Results

In this section, we report experimental results for the performance of the IETF RPL and the contention-based MAC of IEEE 802.15.4. We chose this MAC solution since it is one of the recommended in the IETF RPL standard. However, the framework we have proposed above for the protocol modeling and selection can be applied to any MAC compatible with IETF RPL.

We consider a simple routing example to illustrate the analysis, see Fig. 2. However, the results we present in this section give general insights and can be derived for any other topology. We assume that each node generates same own traffic. This is without loss of generality because the forwarded traffic varies from node to node as a consequence of the routing decisions, which results in a heterogeneous traffic generation per node. We focus on two cases, which we denote by *Path 1* and *Path 2*. In Path 1, we analyze the end-to-end reliability from node V_7, which routes its packets along the path V_4–V_1 to the root node V_0. In Path 2, V_7 forwards its packets along the path V_6–V_3. Coupling is obtained by letting the carrier sensing range of nodes in a path include the nodes in the other path, up to two hops away.

Fig. 3 shows the end-to-end reliability as obtained by the mathematical model and the simulations, where Path 1 and Path 2 are coupled. The best performance is for Path 1. Nodes in Path 1 are dominant in terms of traffic load and affect negatively the performance of nodes in Path 2. The reliability of a contention-based MAC scheme increases as the average number of contenders in each time unit reduces. If a routing metric based only on the maximization of the endto- end reliability is considered, then with a strong coupling in the network the routing decision leads to an unbalanced distribution of the traffic load. This is because RPL forces the forwarded traffic to more dominant nodes, so that the average number of contenders in each time unit is lower. If not taken into account, this MAC phenomenon may be catastrophic and cause instability when considering limited node buffer sizes and energy constraints.

In Fig. 4, the end-to-end delay from node V_7 to the root node V_0 is presented. Path 1 has a lower delay than Path 2. Similarly to the reliability analysis of Fig. 3, if the routing metric is based only on the minimization of the end-to-end delay, and there is coupling in the network, the result is an unbalanced traffic distribution to dominant nodes. Once again, this could lead to a catastrophic effect.

6. CONCLUSIONS

We gave an overview of networking technologies for the IoT, with particular focus on the MAC protocol IEEE 802.15.4 and routing protocol IETF RPL. We argued that a mathematical description is essential to characterize the complex interdependence among these protocols, and that an independent selection of their parameters is not in general the best design choice.

We illustrated a simple mathematical framework for a joint optimization of the protocol parameters of MAC and routing. We believe that the inclusion of such a framework in the current standardization process could be very useful to improve the performance of IoT networks. Future innovations brought by the growth of the IoT application domain can thus prosper on top of a highly efficient and performing network.

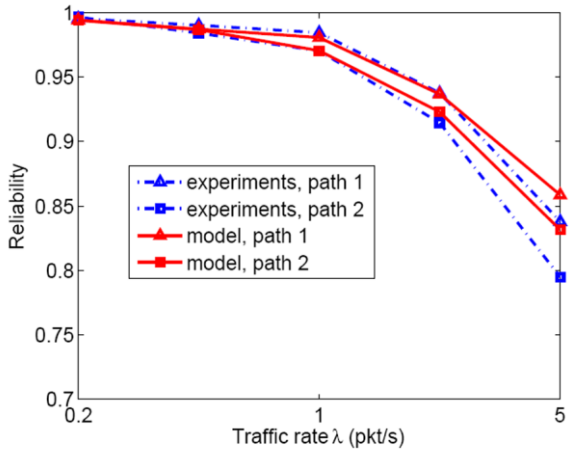

Figure 3. End-to-end reliability for coupled paths for the multi-hop topology of Fig. 2. Path 1 is V_7-V_4-V_1-V_0, when the link V_7-V_6 is disabled. Path 2 is V_7-V_6-V_3-V_0, when the link V_7-V_4 is disabled

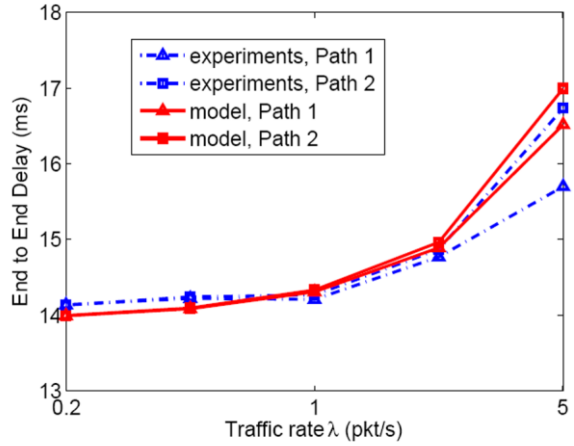

Fig. 4. End-to-end delay for coupled paths for the multi-hop topology of Fig. 2. Path 1 is V_7-V_4-V_1-V_0, when the link V_7-V_6 is disabled. Path 2 is V_7-V_6-V_3-V_0, when the link V_7-V_4 is disabled.

REFERENCES

[1] IEEE 802.15.4 standard: Wireless Medium Access Control and Physical Layer Specifications for Low-Rate Wireless Personal Area Networks, IEEE, 2006, http://www.ieee802.org/15/pub/TG4.html.

[2] P. Park, C. Fischione, and K. H. Johansson, "Adaptive IEEE 802.15.4 protocol for energy efficient, reliable and timely communications," in ACM/IEEE IPSN, 2010.

[3] Routing Over Low power and Lossy networks, Internet Engineering Task Force (IETF), http://www.ietf.org/dyn/wg/charter/roll-charter.html.

[4] I. Demirkol, C. Ersoy, and F. Alagoz, "MAC protocols for wireless sensor networks: a survey," IEEE Communications Magazine, vol. 44, no. 4, pp. 115 – 121, 2006.

[5] A. Bachir, M. Dohler, T. Watteyne, and K. K. Leung, "MAC essentials for wireless sensor networks," IEEE Communications Surveys and Tutorials, vol. 12, no. 2, pp. 222–248, 2010.

[6] K. Langendoen and A. Meier, "Analyzing MAC protocols for low datarate applications," ACM Transactions on Sensor Networks, vol. 7, no. 2, 2011, to appear.

[7] N. Vassileva and F. Barcelo-Arroyo, "A survey of routing protocols for energy constrained ad hoc wireless networks," in Future Generation Communication and Networking (FGCN 2007), vol. 1, 2007, pp. 522 – 527.

[8] P. Di Marco, P. Park, C. Fischione, and K. H. Johansson, "Analytical modelling of IEEE 802.15.4 for multi-hop networks with heterogeneous traffic and hidden terminals," in IEEE Globecom, 2010.

Internet of Things: Opportunities, Challenges, Current and Future Services

Shiddhartha Raj Bhandari, Gyu Myoung Lee, and Noel Crespi
Institut Telecom, Telecom SudParis

Abstract: **The Internet of Things (IoT) is a new paradigm coined to extend physical objects into the Internet where everything communicates and collaborates with anything, anytime, anywhere. Such capabilities allow offering a new range of services closely related with the real world called "real world services". In this paper, we explain IoT, its service evolution, use cases and challenges. Considering the requirements, we propose integrating IoT with cloud which solves resource requirement of objects and opens up flexible resource utilization and seamless service provisioning from cloud.**

***Index Terms*—Internet of Things, real world services, ubiquitous network, connected objects.**

1. INTRODUCTION

Internet of Things (IoT) is a novel paradigm getting popular in research and industries. The basic idea is IoT is to connect things or objects around us (electronic, electrical, non electrical) to provide seamless communication and contextual interactions [1]. Development of RFID tags, sensors, actuators, mobile phones make it possible to materialize it. The main part of the Internet of Things is made up by nodes associated to objects with sensorial and actuation capabilities; those nodes are tiny and wirelessly networked which facilitate pervasive deployments. Such capabilities allow offering a new range of services closely related with the real world so they are called "real world services". The real world service paradigm requires new interaction forms usually classified in three groups: object-object, object-human and object-Internet. This activity could generate massive amounts of data taking place to the knowledge of the system.

There are so many applications that are possible because of IoT. Individuals, professionals, industries, all sectors will get benefitted with the plethora of contextual and real time services. For individual users, IoT brings useful applications like home automation, security, automated devices monitoring, and management of daily tasks. For professionals, automated applications provide useful contextual information all the time to help on their works and decision making. Industries, with sensors and actuators operations can be rapid, efficient and more economic.

Managers who need to keep eye on many things can automate tasks connecting digital and physical objects together. Every sectors energy, computing, management, security, transportation are going to be benefitted with this new paradigm. Domestic living, e-assisted living [2], e-health, elearning, Automation, industry management, logistics, business processing are also the application scenarios where IoT will have effect on. As we move forward, there will be a big network of things; this will create a large number of connections and enormous amounts of data [3]. In this paper we explain IoT, its service development and current usages scenarios. We further explain future challenges and service requirements and present one solution of integrating IoT with cloud in in order to support resource requirement

of large number of objects. Integrating IoT with cloud enables flexible resource utilization and seamless service provisioning.

Organization of this paper is as follows. In section II, we explain IoT and its service evolution; in section III we describe technologies behind IoT and current use cases. In section IV, we explain some challenges and describe our proposed IoT cloud integration. Section V concludes this paper.

2. IOT AND ITS SERVICES

2.1. Development of IoT and services

The very first concept of IoT came from things oriented perspective. It is attributed to the Auto-ID Labs of MIT. It is dedicated to creating the IoT using RFID and wireless sensor networks. Working together with EPC Global, Auto-ID Labs started to create a global system for tracking goods using single numbering system called the electronic product code. Their focus was to develop industry-driven standards for EPCglobal network to improve object visibility, traceability of an object and its status and current location [4]. Similarly on the vision of IoT, Unique universal ubiquitous Identifier called UiD center was established with in the T-engine forum to establish core technology for automatically identifying physical objects and locations and to work toward the objective of realizing a ubiquitous computing environment [5]. Still RFID remains the key technology behind IoT however wide portfolio of devices, technologies, services will eventually build IoT. We explain various service development stages since IoT was first introduced.

Identity related services

At the beginning IoT started with identity related services. Small objects were given unique identification with which they were able to track them and monitor. Identity related service uses RFID, two-dimensional code and barcode as technology.

Information aggregation services

More advanced services started compare to identity related services. With such services, terminal used to collect and process data and report via communication network to the platform. Platform further used to process and implement it in different purposes. This category of services includes smart meter reading, elevator management, logistics, and traffic management.

Collaborative services

With development in object, communication and information management system, IoT started providing collaborative services. This is current stage where things are bringing very sophisticated services. They need terminal to terminal and terminal to people collaboration. These communications require higher reliability, delay and objects need to be smart.

Ubiquitous services

Ubiquitous services aim to provide smooth communication anytime, anywhere for anybody and everything. This is the highest communication that is possible to humankind. This concept integrates all the physical objects (real world things) in to information world, which were not there before. This will have a lot of challenges on which we need to be focused on.

2.2. Technology behind IoT

Development of several technologies made it possible to achieve the vision of Internet of things. Normal physical objects were not able to perform any computing and communication. Embedding processing, sensing and communicating capabilities enable normal object to become smart which enables range of potential applications and services. Interaction can be device-to-device or person-to device depending on the current context and requirements. Identification technology such as RFID allows each object to represent uniquely by having unique identifier. It provides possibility of tracking objects but has limited sensing capability and deployment flexibility [6]. Wireless sensor technology allows objects to provide real time environmental condition and their context. Establishing connections among objects are possible with various short range communication technologies like WiFi, Bluetooth, ZigBee, and NFC. Smart technologies allow objects to become more intelligent; not only think and communicate but also analyze and react. Nanotechnologies are helping to reduce the size of the chip incorporating more processing power and communication capabilities in a very small chip. We can attribute development of different technologies behind evolution of IoT.

3. CURRENT USE CASES AND CHALLENGES

3.1. Current usages of IoT

In this section, we list out some current use case scenarios of IoT.

Production and management

Nowadays objects produced in factory are uniquely tagged, so each item can be tracked from the production phase to distribution. Automatically production order can be received, items will be checked in and out and orders passed on to the suppliers. Goods are transferred from producer to consumer without much human interaction. Producers have real time market view based on the information they got about demand and supply. Based on that information, production can be optimized reducing energy consumption and becoming environment friendly.

E-health

The use of tagging and sensing technologies is allowing real time monitoring of patient's health information. Heart beat, blood pressure, breathing rate can be measured with the help of light weight sensors. Such sensors collect process and transmit medical information to different health unit making health care more efficient and reliable compare to before. Automatic cure and emergency alert can help medical personnel in providing better health care.

Smart grid

Smart grid is a field where Internet of Thing concept is being used. More flexible energy demand and supply features are available in current electric grid system based on two way communication. Now, households can sell surplus energy to the electric grid. It is possible because of the development of communication technology. More efficient energy utilization is possible in electric grid. Smart meters which are mounted in home communicate with all electric appliances in home exchanging energy consumption information. Meters can collect data and send it to the main grid that can plan the energy consumption in a home.

Energy efficiency

With the help of lighting and temperature sensors, houses are being able to reduce energy consumption without compromising the comfort. With sensors dynamic room temperature adjustment and lighting system are maintained. With IoT, remote home monitoring is possible which help to turn off the devices at peak time when energy price is high. Such sensing and collaborating reduces energy utilization and help in economizing.

3.2. Challenges of IoT

In this section we explain some of the challenges that IoT need to face in future and our proposal of integrating objects into cloud.

Standardization

IoT supports interaction among heterogeneous objects and sources of data. Technological standardization in the area of IoT is still in infancy [7]. Uses of standard interfaces, protocols and data models are required to ensure high degree of interoperability among diverse system. Several technologies are available, however they are proprietary and developers and users face difficulties during information exchange and interworking. One entity with in one organization or management periphery should be able to interact with another efficiently without any issue of interoperability. This allows individual and organization gain benefits from competitive marketplace performing their tasks efficiently involving heterogeneous system efficiently.

Security

Internet of things should be built ensuring easy and safe functionalities. Users should feel safety in order to enjoy the benefits avoiding risk of safety and privacy. As everything is going to be connected and interact with each other it will raise new security and safety related problems. Unauthorized access, tracking of the objects may breach the basic security. Compare to traditional distributed system; billions of objects need to update their information which may opens up for many security challenges and security techniques across multiple policies.

Scalability

Future network will consist of all the things around us .There is no limit on the number of objects present in any scenario. When they grow large, there will be the issue of managing them and provisioning resources according to their need. Object can be overwhelmed because of large number of objects interacting at the same time, for example a temperature sensor in a park. Many people want to know current weather condition either to visit there or not. In this case, providing temperature information from sensor itself will be difficult as a sensor may not be able to handle multiple requests at the same time.

4. INTEGRATING IOT WITH CLOUD

Here we explain our proposal of integrating physical objects in to cloud so that resource requirement and real world service provisioning can be fulfilled from the cloud.

To provide all the computational, memory, security and networking facility, users need to invest huge budget in physical infrastructure. Cloud, a network based computing, is becoming popular and is being implemented by large Information Technology (IT) companies and

research centers. Cloud has ability to handle massive data as per on demand service [8]. Cloud computing can be a possible solution to fulfill resource requirements of IoT. Services, platform, and infrastructures can be provided inside cloud according to need of the users with negotiated Quality of Services (QoS) via Service Legal Agreement (SLA). The cloud on the IT infrastructure can work on behalf of the object, increasing availability and performance. Scalability is possible to maintain via clouds.

4.1. Why IoT requires Cloud Computing

IoT is coming in reality and being implemented in various sectors. Deploying IoT means connecting millions of objects together. When general objects connect to the network, they have to involve in some computational activities which is difficult for them. In such case, computational, storage and memory intensive tasks can be performed in cloud infrastructure. Integrating both will fulfill each other's goal. We have listed several points in support of integrating IoT with cloud.

Flexibility of resource allocation

It is difficult to predict resource requirement for IoT in advance. Sometime there will be high resource requirements and some time less. Depending up on context, situation, and environment resource requirement might vary. In such case there is need of flexible resource allocation. Cloud can provide flexible resources to objects on their needs.

More intelligent applications

Common and more intelligent applications can be deployed in cloud. Developing same application for every new installation and operation will cost lots of investments in software development and management. Small enterprise who wants to use such utility but does not have enough budget can be benefitted using them from cloud. They subscribe it at once and pay for the use.

Energy saving

Small devices are power constraints. It is not possible always to plug it with continuous power source. One possible solution will be to delegate processing power to somewhere else in Internet. Cloud can be a good option. Devices can delegate its works to perform in cloud which conserves energy.

No onsite infrastructure

IoT applications require different infrastructures. It is costly to install expensive equipments on site. Cloud service providers maintain small to large scale infrastructure that can be leased to users. IoT applications can use infrastructures through the cloud as similar as it is present on site.

Heterogeneity of smart environment

In ubiquitous computing environment, we will see heterogeneity. Applications and services vary in intelligence, mobility and function. . System needs more intelligent, more interactive and more accurate.

Scalability and agility

Scalability is important when we talk about connecting normal objects into the network. They will be in large number. Providing and maintaining scalable solution is a big challenge.

The important feature of cloud computing is elasticity of resources. Uncertain resource requirements of IoT can be fulfilled by elastic resource availability of cloud.

Simple interfaces

One of the goals of ubiquitous networking is to provide a minimum of complexity. Complex configurations and user interfaces hinder the acceptation of a new technology. Therefore, the simplest possible user interfaces should be maintained. When most of the services will be deployed in a cloud and they become a normal commodity, more and more focus will be given to making user interfaces simple and accessible.

Virtualization

Virtualization is a technique through which resources: processors, storage, I/O and network on one or more machines can be transformed through hardware/software partitioning, time sharing and simulation/emulation into multiple execution environments, each of which can act as a complete system. Using virtualization, objects can use resource as they are single users of it.

4.2. Implementation

Physical objects are becoming smarter with the continual augmentation of communication and computing capabilities. Service Oriented Architecture (SOA) based programming, which was initially used for complex, and rather static business data sharing can now be used for small objects [9]. Objects can offer their functionalities using the Simple Object Access Protocol (SOAP) or the Representational State Transfer (REST) Application Programming Interface (API) based approaches [10]. This allows objects to interact dynamically. Devices that provide their functionality as a web service can be used by other entities such as business applications or other devices.

Figure 1 shows use case scenario how each object inside and outside home is interacting with cloud to use resources. Each objects equipped with sensor senses current environment. Web services technology will be used to interact among objects and cloud. Hypertext Transfer Protocol (HTTP) methods for example GET, PUT and DELETE can be used to perform operations. Various use cases are possible; Cloud application can access objects and its services, objects can use cloud services, user can access services and information regarding physical objects.

Connected objects using resources from cloud

In this case, objects request services in cloud. Appliances such as refrigerator, washing machine, air conditioning systems inform their electricity consumption on demand or in every specific period of time. Sensor measures current electricity usage level and uploads to cloud. Objects have web service interface to publish and let others to use its functionality.

User/cloud accessing objects

In this use case, user requests object information from cloud. Cloud provides information either from storage or immediately accessing objects. Cloud can provide information in graphs, usage history and other interactive forms. User can request different information regarding his/her interested objects. They even can operate the devices using real world services deployed in cloud.

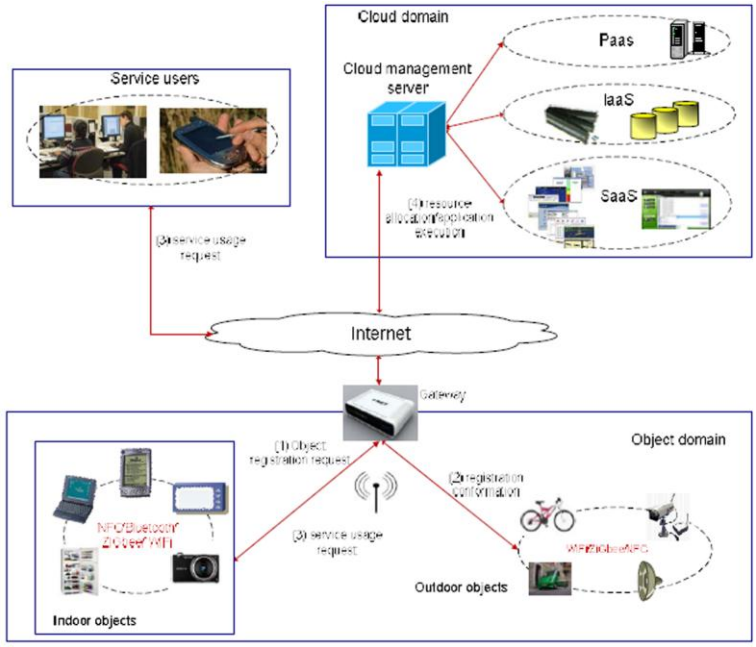

Figure 1.Configuraiton of IoT cloud integration

5. CONCLUSION

In this paper, we have explained a new paradigm, IoT, its service evolution, challenges and use cases. IoT has the potential of creating new ecosystem of communication and collaboration through a world-wide network of interconnected objects. Along with opportunities, it has posed considerable challenges among us. In this paper, we have figured out some challenges and proposed resource requirement problem to solve by integrating objects to cloud. We believe this integration opens flexible resource utilization and real world services service provisioning via cloud. Our future works will be focusing on solving scalability and service provisioning issue of proposed solution.

REFERENCES

[1] Luigi Atzori, Antonio Iera, Giacomo Morabito, “The Internet of Things: A survey,” Computer Networks, Volume 54, Issue 15, 28 October 2010,Pages 2787-2805.

[2] Zouganeli, E., Svinnset, I.E., , “Connected objects and the Internet of things — A paradigm shift,” PS '09. International Conference on Photonics in Switching, 2009, pp.1-4.

[3] Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., Schreier, G., “The Internet of Things for Ambient Assisted Living,” New Generations (ITNG), 2010 Seventh International Conference on Information Technology, April 2010, pp.804-809.

[4] http://autoid.mit.edu/cs/

[5] http://www.uidcenter.org/english/introduction.html

[6] Kortuem, G.; Kawsar, F.; Fitton, D.; Sundramoorthy, V.; "Smart objects as building blocks for the Internet of things," Internet Computing, IEEE , vol.14, no.1, Jan.-Feb. 2010,pp.44-51,

[7] ITU-T Internet Reports, "Internet of Things," November 2005.

[8] Rimal, B.P., Eunmi Choi, Lumb, I., "A Taxonomy and Survey of Cloud Computing Systems," NCM '09. Fifth International Joint Conference on INC, IMS and IDC, 2009, pp.44-51

[9] Guinard, D.; Trifa, V.; Karnouskos, S.; Spiess, P.; Savio, D.; "Interacting with the SOA-Based Internet of Things: Discovery, Query, Selection, and On-Demand Provisioning of Web Services," IEEE Services Computing,IEEE Transactions, vol.3, no.3, July-Sept. 2010. pp.223-235

[10] Malatras, A.; Asgari, A.; Bauge, T.; "Web Enabled Wireless Sensor Networks for Facilities Management," IEEE Systems Journal, vol.2, no.4, Dec. 2008, pp.500-512

Locating Weigh-In-Motion Checkpoints in Traffic Networks Using Genetic Algorithm

Milica Šelmić, Nikola Bešinović, Dušan Teodorović
University of Belgrade, Faculty of Transport and Traffic Engineering

Abstract: **The problem of optimizing locations for weigh-in-motion (WIM) checkpoints facilities is studied in this paper. The WIM checkpoints equipped with a remote system can be used for various practical applications like statistics, interval measuring and online traffic control, selecting of overloaded vehicles, estimation of the current loading of road or bridge constructions and so on. In order to check truck weight limits and obtain all mentioned information, we need to find answer to the question: where should WIM checkpoints be located? This paper develops a model to determine the locations of WIM checkpoints in traffic network. The problem is solved using the Genetic Algorithm. The purpose of this research was to develop a decision support methodology to identify the optimal locations of a finite set of WIM checkpoints on a transport network in order to minimize possible risk and to maximize total flow captured.**

1. INTRODUCTION

Several billion tons of hazardous materials are shipped yearly in the world. Authorities in some countries have established a system of checkpoints to detect leaks from trucks transporting hazardous materials such as waste oil, diesel fuel and calcium chloride. At these checkpoints, vehicles in violation are taken out of service. At truck weighing stations traffic authorities check weight limits, driver hours and service regulations, vehicle equipment safety, and collect road use taxes. For example, in some states and countries, simple portable scales enable a weighing station to be setup practically at any node along a highway. Weigh-in-motion (WIM) systems reduce the probability of occurrence of overloaded trucks on highways. Economic benefits of the WIM checkpoints far exceed the costs of implementation and usage.

WIM systems are finding increasingly widespread use as a valuable extension to conventional traffic counters and classifiers. They provide a whole spectrum of information on traffic flow, with detailed data for each individual vehicle, including:

- Dynamic weights of all axles (or if selected, even left/ right half axles),
- Gross vehicle weights,
- Axle spacing,
- Distance between vehicles,
- Speed,
- Vehicle classification according to various schemes,
- Versatile statistic representations for all types of traffic parameters.

In order to check truck weight limits and obtain all necessary information we need to find an answer to the question: where should WIM checkpoints be located?

The WIM checkpoints belong to the class of flow-intercepting facilities (including billboards, gas stations, fast food outlets, convenience stores, ATMs, retail facilities, etc.). In the case of flow-intercepting facilities, clients (drivers) are serviced if they pass through a

facility. Every client (driver) that passes through such a facility is treated as a 'captured' or 'intercepted' client.

The aim of this paper is to develop the model capable to determine optimal locations of WIM checkpoints in traffic network. The objective function to be maximized represents the combination of the largest possible risk reduction and the largest possible intercepted flow on the traffic network. The number of facilities is given in advance and treated as a constraint in the problem formulation.

Location of WIM checkpoints represents *NP* hard combinatorial problem. This fact motivated the authors of the paper to develop heuristic tehnique based on Genetic Algorithm. The proposed model is supported by numerical examples.

The paper is organized as follows. The problem statement is in the second section. The third section is devoted to the Genetic Algorithm approach to the WIM checkpoints location problem. The numerical test results follow in the fourth section and finally conclusions are provided in the last, fifth, section.

2. PROBLEM STATEMENT

The problem of determining the locations of flow-capturing facilities in a transportation network have been treated by numerous authors [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11].

One group of flow-capturing problems comprises problems such as: location of billboards, gasoline stations, fast food outlets, convenience stores, ATMs, or retail facilities. When considering the location of these facilities, the goal should be to capture the largest possible total flow. On the other hand, driving under influence checkpoints, weighing stations, routine police check-ups, or hazardous material checkpoints are related to specific risks in traffic network [12]. In each case, any client intercepted more than once counts only one time.

In the case of WIM checkpoints that are risk-related, drivers should be intercepted as soon as possible after starting their trips through traffic network. This idea reflects preventive policy. Within this policy, WIM checkpoints should be located in the network in such a way so as to maximize the largest possible risk reduction in a traffic network. On the other hand, costs of locating WIM checkpoints are very high, so it is very important to optimize location of WIM checkpoints. According to this idea, it is important to locate WIM checkpoints at the points (tunnels, viaducts, bridges) with the most intensive traffic flows.

In this paper we study location problems in the case of risk-related flow-intercepting facilities (the preventive flow-intercepting problem). We consider the WIM checkpoints location problem in the case of a non-oriented network $G=(V, A)$. The total number of nodes in the network in which it is possible to locate WIM checkpoints equals n. We denote respectively by R ($R \subseteq V$), and S ($S \subseteq V$) the set of origin and the set of destination nodes respectively. We also denote by P a set of origin-destination path flows. The total number of WIM checkpoints that we want to deploy equals m.

Let us consider a path $p \in P$ and a set of nodes V_p on the path p. Proposing the idea of preventive policy Gendreau [8] introduced the quantity $a_{ip} \geq 0$ that denotes "the reduction of risk achieved on path p if the first facility encountered along that path is located at vertex $v_i \in V_p$ ". Obviously, $a_{jp} < a_{ip}$ when node v_j follows node v_i .

Let us introduce binary variables y_i and x_{ip} defined in the following ways:

$$y_i = \begin{cases} 1, \text{ if facility located at node } v_i \\ 0, \text{ otherwise} \end{cases}$$

$$x_{ip} = \begin{cases} 1, \text{ if the flow on path } p \text{ is first intercepted at } v_i \in V_p \\ 0, \text{ otherwise} \end{cases}$$

The mathematical formulation of the WIM checkpoints location problem reads:

Maximize

$$\sum_{p \in P} \sum_{v_i \in V_p} a_{ij} f_i x_{ip} \qquad (1)$$

subject to:

$$\sum_{v_i \in V} y_i = m \qquad (2)$$

$$x_{ip} \leq y_i \quad (p \in P, v_i \in V_p) \qquad (3)$$

$$\sum_{v_i \in V_p} x_{ip} \leq 1 \quad (p \in P) \qquad (4)$$

$$y_i \in \{0,1\} \quad (v_i \in V) \qquad (5)$$

$$x_{ip} \in \{0,1\} \quad (p \in P, v_i \in V_p) \qquad (6)$$

The objective function (1) reflects our desire to maximize reduction of risk and to maximize captured traffic flow. The total number of checkpoints that should be deployed equals m (constraint (2)). The flow on the path p cannot be intercepted in node V_i if there is no WIM checkpoints deployed (constraint (3)). By constraint (4) not all paths necessarily contain a facility and each path counts at most once towards the objective function. Binary nature of variables y_i and x_{ip} is represented by constraints (5) and (6). The number of WIM checkpoints is given in advance and treated as a constraint in the problem formulations.

Large number of papers deals with WIM problems, but most of them treats different types of sensors that can be used as WIM facility [13, 14, 15, 16]. The most of WIM checkpoints work on the principle of the piezoelectric WIM Lineas ® Quartz Sensor, Slow Speed WIM sensors, PAT Bending Plate WIM sensors or Single Load Cell WIM sensors. Also, there is the paper that focus on improvement of the precision of weigh-in-motion due to vibration from

dynamic load [17]. An excellent overview of research on weigh-in-motion system is given in [18].

To the best of authors knowledge there is no paper in relevant literature that finds optimal location to the WIM checkpoints in a transport network.

3. USING GA TO OPTIMIZE WIM LOCATIONS

Genetic Algorithm (GA) is heuristic search techniqe based on the evolutionary ideas of the natural selection and genetic. The basic concept of GA is designed to simulate processes in natural system necessary for evolution. GA was first developed by Holland [19]. GA represents an intelligent utilization of a random search within a defined search space to solve a problem.

GA belongs to the class of algorithms that have the ability to find solutions close to optimal for complex combinatorial optimization problems. This method is probabilistic and perform a multidirectional search by maintaining a population of potential solutions. The new generation of solutions (individuals) is expected to be better than the parent population because only the good quality solutions (individuals) from the parent population are allowed to participate in future mating [20].

Authors suggested GA as a tool for solving proposed model of WIM checkpoints locations. GA was tailored for solving particular problem by adding more complex operations in processes of generation initial population and selection.

An instinctive way of representing solutions for the WIM checkpoints locations problem is using of binary coding.

One solution is presented as a string of cells as shown in Figure 1. The number of nodes (n) in the network is adopted as the length of each string (individual, chromosome) in the population. The value in the cell indicates the existence of a WIM facility. A value of 1 means that a WIM facility is deployed at that location and a value of 0 means there is no deployment. The sum of all cell values, or the length of the string, is equal to the number of WIM facilities to be deployed (m).

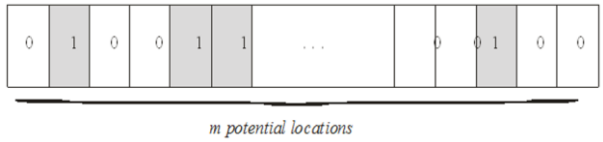

Figure 1. Solution Representation

Initial generation is selected in such way to provide a large number of feasible individuals. This means that each individual who is repeated several times in the generation is removed.

During the selection process the elitist strategy is combined with the steady-state replacement. Using elitist strategy, the best N_{elit} individuals are chosen and directly transferred to the next generation. The individuals for the other places are obtained by applying the genetic operators. In order to have more equal treatment of all individuals in a popuation, GA does correction of elitist individuals' objective function by using of correction function used in [21].

After correction, weaker individuals, that consist of good genes, have higher probability to take part in making the new generation.

The selection of individuals is done by a tournament. The number of groups and the number of individuals in each group are predefined. The individual with the highest objective function value becomes the winner of the group. Each group winner represents a parent that will form an offspring for the next generation. Offsprings are made by combining the genes extracted from the parents. The method of crossover used in the algorithm is the one-point crossover (Figure 2). Typically, the probability for crossover ranges from 0.6 to 0.95.

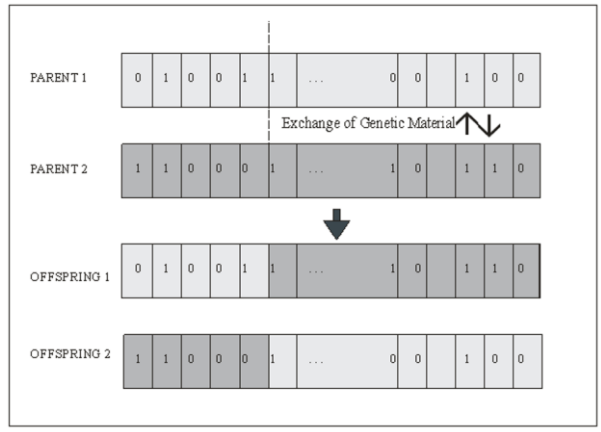

Figure 2. Random one-point crossover operator

To introduce and preserve diversity in search population, a mutation operator is applied to the generated offspring. It operates independently on each individual by probabilistic perturbing a single random gene. This type of mutation operator is called one-point mutation (Figure 3).

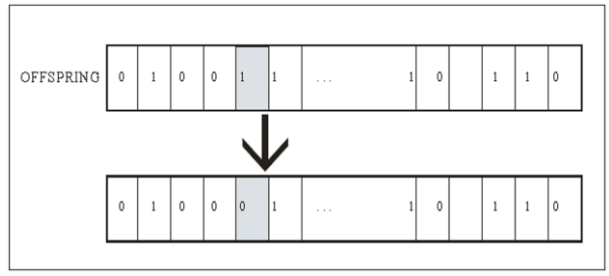

Figure 3. Mutation operator

The algorithm uses predefined number of generations as stopping criterion. Before running the algorithm, user defines the dimension of the population, the number of elitist individuals and number of generations.

4. NUMERICAL EXAMPLES

The proposed model is tested on instances that are randomly formed, due to the lack of real data. All these matrices are available upon request. We tested our model and GA on networks with 10, 20 and 30 nodes. Number of WIM checkpoints that should be located is varied from 3 to 13. Quantities a_{ip} are generated in a random way. During this procedure, we generated quantities a_{jp} and a_{ip} in such a way that $a_{jp} < a_{ip}$ when node v_j follows node v_i.

In the problem studied in this paper input parameters are: number of nodes, number of WIM checkpoints, the origin-destination matrix and the shortest paths between nodes in a transport network.

Tests were performed on the Intel Atom 1.6 MHz processor and 1 GB of RAM. Results obtained by GA were compared with optimal solutions. To find the optimal solution IBM ILOG OPL optimization programming language was used. The number of individuals in the initial population is 50. The number of generations is 50, 100 or 300, depending on problem dimensions. During the selection process we combine the elitist strategy and the steady-state replacement. Using elitist strategy, we select the best 30 individuals and directly transferred them to the next generation. The other 20 individuals are obtained by applying genetic operators. The selection of individuals is done by a tournament. We selected the number of groups (N_{gr} = 20) and the number of individuals in groups (N_i = 3). The winners of the groups represent parents that will form individuals in the next generation. In the algorithm we used one-point crossover and one-point mutation. We define crossover probability p_{cros} = 0.85 and the coefficient of mutations p_{mut} = 0.5.

Table 1. GA results

No. of nodes	No. of flows	No. of WIM checkpoints	GA BEST	OPL OPT	Cap. flow (%)	No. of gener.
10	90	3	37417	37417	81.2	<10
10	90	4	42466	42466	91.7	<10
10	90	5	45293	45293	97.2	<10
10	90	6	46739	46739	99.4	<15
20	380	3	138484	138484	79	<40
20	380	5	159212	159212	89.9	<70
20	380	8	177008	177008	98.1	<90
20	380	10	181634	181634	100	<70
30	870	3	240258	240258	58.8	<30
30	870	6	327098	327098	78.2	60
30	870	10	389199	389199	91.7	<110
30	870	13	414965	414965	96.7	<260

Table 1 shows results of applying GA on the proposed model and their comparison with optimal results. The first column of Table 1 contains the total number of nodes, while the next two columns present the total number of flows on the network and the number of WIM checkpoints to be located. The fourth column contains the best known results obtained by GA. The next column presents optimal results reached by the Optimization programming language

(objective function value, OPL OPT). In the sixth column we report the percentage of captured flow. The last column contains the number of generations that GA needs to obtain the best result.

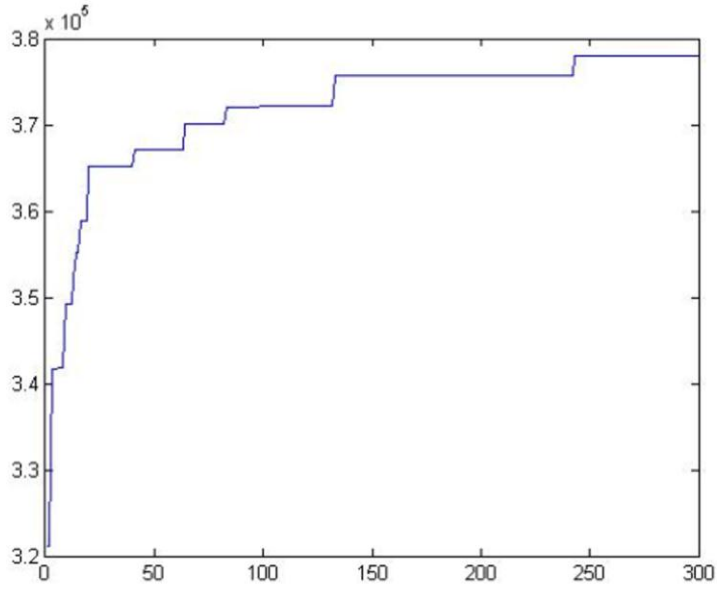

Figure 4. Maximization of objective function

From Table 1 it can be seen that GA was able to get optimal solution in all tested networks. Table 1 and Figure 4 show that all solutions are obtained using a relatively small number of generations. Also, from Table 1 it can be noticed that proposed model enable large precentage of intercepted flow with maximization of possible risk reduction. Only few WIM checkpoints are enough to capture almost total flow in a transport network.

Table 2 presents real value and percentage of captured flow considering the number of located WIM checkpoints (the particular example consists of 30 nodes and number of WIM checkpoints is varied from 1 to 13). From Figure 5 and Table 2 it can be seen that even 3 optimally placed WIM checkpoints are enough to capture almost 60% of total flow on transport network. With 8 properly located WIM checkpoints more than 85% of total flow can be intercepted. Every pair (*Number of WIM checkpoints*, *Percentage of captured flow*) corresponds to a specific decision. In this manner, a large number of different potential location choices are generated for the decision-maker.

Table 2. Captured flow

No. of nodes	No. of WIM checkpoints	GA BEST	Captured flow	Captured flow (%)
30	1	110640	120757	27.2
30	2	199880	217762	49.1
30	3	240258	259616	58.5
30	4	277193	297353	67.0
30	5	306909	327380	73.7
30	6	327098	347414	78.2
30	7	345045	364028	82.0
30	8	361709	380575	85.7
30	9	377898	396635	89.3
30	10	389199	407227	91.7
30	11	399320	416148	93.7
30	12	407998	423104	95.3
30	13	414965	429593	96.7

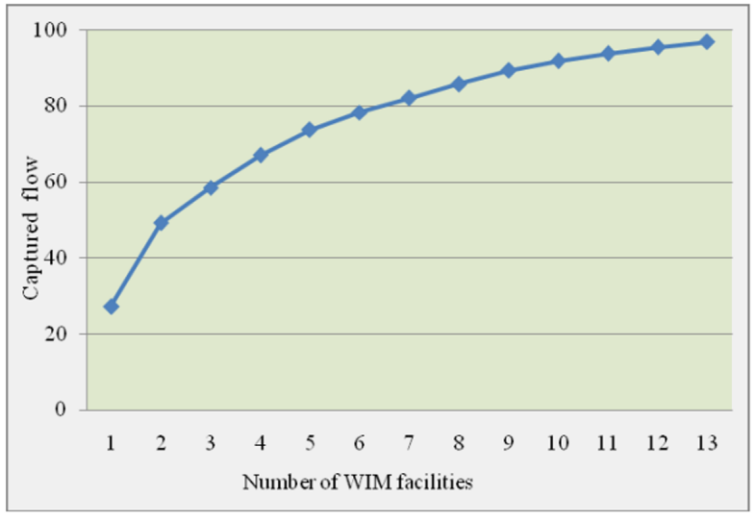

Figure 5. Percentage of captured flow

Figure 6 and Table 3 show the number of times a potential WIM checkpoint location is present in the optimal solutions.

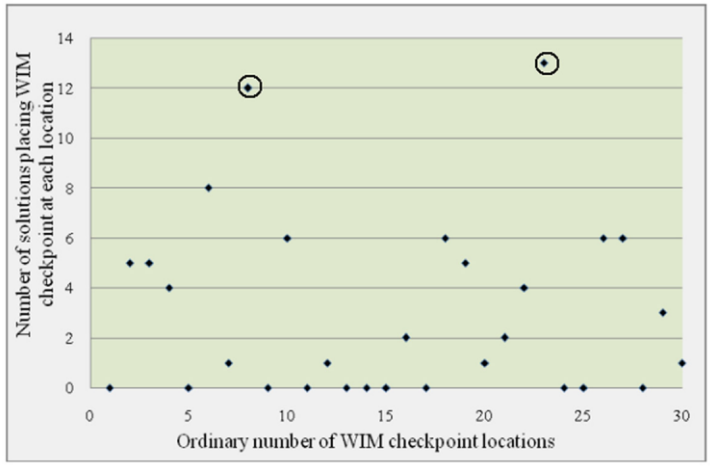

Figure 6. Frequency plot showing the number of times a WIM checkpoint is placed at each location

For the network of 30 nodes, for example, from 13 sets of WIM checkpoint deployments, locations 23 and 8 are present in 13 and 12 sets, respectively. This is an indication that these locations are very critical for the flow capturing and WIM checkpoints deployed in these locations need to be regularly maintained.

Table 3. The number of times a WIM checkpoint is placed at each location

No. of nodes	No. of WIM facilities	Locations
30	1	23
30	2	8,23
30	3	3,8,23
30	4	6,8,23,26
30	5	6,8,21,23,26
30	6	6,8,21,23,26,27
30	7	6,8,19,21,23,26,27
30	8	6,8,10,18,19,23,26,27
30	9	2,3,8,10,16,18,19,20,23
30	10	2,4,8,10,16,18,22,23,29,30
30	11	2,3,4,6,8,10,12,18,22,23,27
30	12	2,3,4,6,8,10,18,19,22,23,27,29
30	13	2,3,4,6,7,8,10,18,19,22,23,27,29

5. CONCLUSION

The trucks, loaded over their allowed weight, cause great damage to the road and traffic security. To solve these problems we must take efforts in developing more advanced WIM system and in optimizing location of WIM checkpoints.

This paper studied the WIM checkpoints location problem. We analyzed preventive policy in which risk-related drivers should be intercepted as soon as possible after starting their trips through the traffic network. Within the analyzed policy, WIM checkpoints should be located in the network in such a way to maximize largest possible *risk* decrease on traffic network and total flow captured.

Since presented problem is *NP* hard, we use Genetic algorithms as a tool for its solving. In the absence of real data, we tested our model on randomly generated instances. The results show that with optimal placement of WIM checkpoints high percentage of traffic flow can be captured. With carefully placed WIM checkpoints that are well maintained, all necesarry information can be derived.

REFERENCES

[1] Hodgson, M.J. "The location of public facilities intermediate to the journey to work", Eur J Oper Res, Vol. 6, pp 199-204, 1981.

[2] Hodgson, M.J. "A flow-capturing location-allocation model", Geogr Anal Vol. 22, pp 270-279, 1990.

[3] Berman, O. Larson, R.C. and N. Fouska, "Optimal location of discretionary service facilities", Transportation Science Vol. 26, pp 201-211, 1992.

[4] Berman, O. Hodgson, M.J. and D. Krass, Flow-interception problems, Facility location. A survey of applications and methods, Springer-Verlag, New York, pp 389-426, 1995.

[5] Mirchandani, P.B., Rebello, R., and A. Agnetis, "The inspection station location problem in hazardous materials transportation: Some heuristics and bounds", INFOR Vol. 33, pp 100-113, 1995.

[6] Hodgson, M.J., Rosing, K.E. and J. Zhang, "Locating vehicle inspection stations to protect a transportation network", Geogr Anal Vol. 28, pp 299-314, 1996.

[7] Berman, O. and D. Krass, "Flow intercepting spatial interaction model: a new approach to optimal location of competitive facilities", Location Science Vol. 6, pp 41-65, 1998.

[8] Gendreau, M., Laporte, G. and I. Parent, "Heuristics for the Location of Inspection Stations on a Network", Naval Research Logistics Vol. 47, pp 287-304, 2000.

[9] Wu, TS. And J.N. Lin, "Solving the competitive discretionary service facility location problem", European Journal of Operational Research, Vol. 144, pp 366-378, 2003.

[10] Kuby, M. and S. Lim, "The flow-refueling location problem for alternative-fuel vehicles", Socio-Economic Planning Sciences, Vol. 39, pp 125-145, 2005.

[11] Jun, Y., Min, Z., He, B. and C. Yang, "Bi-level programming model and hybrid genetic algorithm for flow interception problem with customer choice", Computers & Mathematics with Applications, Vol. 57, pp 1985-1994 , 2008.

[12] Šelmić, M., Teodorović, D. and K. Vukadinović K., "Locating inspection facilities in traffic networks: an artificial intelligence approach", Transportation Planning and Technology Vol. 33, pp 481-493, 2010.

[13] Suopajarvi, P., Pennala, R., Heikkinen, M., Karioja, P., Lyori, V., Myllayla, R., Nissila, S., Kopola, H. and H. Suni, "Fiber optic sensors for traffic monitoring applications", Proc. SPIE Proceedings of SPIE 3325, pp 222-229, 1998.

[14] Wierzba P. and B.B. Kosmowski, "Polarimetric sensors for weight-in-motion of roads vehicles", Opto-Electronics Review Vol. 8, pp 181-187, 2000.

[15] Udd, E., Kunzler, M., Layor, M., Schulz, W., Kreger, S., Corones, J., McMahon, R., Soltesz, S. and R. Edgar, "Fiber Grating Systems for Traffic Monitoring", Proceedings of SPIE 4337, pp 510-516, 2001.

[16] Kunzel, M., Edgar, R., Udd, E., Taylor, T., Schulz, W., Kunzler, W. and S. Soltesz, "Fiber Grating Traffic Monitoring Systems", Proceedings of SPIE 4696, pp 238-243, 2002.

Medical Fuzzy Models for E-Health Applications

Dragan Z. Šaletić, Milica Savić
School of Computing, University "Union", Belgrade, Serbia

Abstract: **In the paper, a possible e-health application of fuzzy models of determining the severity of respiratory distress in a patient in an intensive care unit is considered. Models of fuzzy multi-criteria decision making are applied in two situations: the first, with all features of equal importance and without interaction between criteria, and the second, with interacting criteria. The practical usage of both approaches is considered. In the first case the Bellman-Zadeh's approach is used, and in the second case, the Choquet integral is used. Directions for possible further work on the telemedicine application considered are pointed out.**

1. INTRODUCTION

E-health technologies are part of Internet society technology and are dealing wiht healthcare practice supported by electronic processes and communication. Ehealth encompasses a range of services or systems that are at the edge of medicine/healthcare and information technology, including, among other technologies, a telemedicine. Telemedicine is a rapidly developing application of clinical medicine, where medical information is transferred through interactive audiovisual media for the purpose of consulting, and sometimes-remote medical procedures or examinations. In the paper, we focus on evaluating the severity of respiratory distress in a patient in an intensive care on the base of fuzzy models, having in mind a telemedicine application. Disease severity is the focal point of medical interest. It reflects all symptoms and impairments caused by analysed disease. The goal of severity evaluation is to generate standardized indicators that are suitable for unified assessment of patients' condition. Diseasy severity is evaluated by medical examiners based on objective measurements and quantification of symptoms.

The severity rating scales used in clinical practice are regularly base on simplistic additive scoring. These scales enjoy wide acceptance because they are easy to administer. The additive scoring approach yields indicators of insufficient precision.

Current medical rating scales do not use graded logic functions and other useful features of soft computing decision models. The main goal of of this paper is to demonstrate how soft computing methods and corresponding software tools can be use as tools in medical evaluation.

Fuzzy sets theory offers lots of aggregation operators for combining membership values representing uncertain information. That is of importance not only for models expressed by fuzzy expert systems, but also for fuzzy decision-making systems, and wider.

In many applications of fuzzy sets theory, in order to model a complex and maybe not fully defined systems, fuzzy decision making systems are used. Thus, the aggregation process of fuzzy information is an important element of fuzzy decision making system.

A situation that requests fuzzy modelling exists in the process of determining the severity of respiratory distress. The idea of using fuzzy sets theory in modelling the decision-making process in determining the severity of respiratory distress was proposed in [1]. In [2], software

implementation of fuzzy decision-making process in determining the severity of respiratory distress is described.

With intelligent medical systems, the status of a patient can be described by numerical values, text and image data, i.e. multimedia data. A patient with respiratory distress is described by a set of symptoms with numerical values assigned at approximate intervals, by a set of symptoms expressed verbally (*Breathing*), and by a set of symptoms presented by an image (chest radiograph -*Rö*). The appropriate model that can deal with all these kinds of data is a fuzzy model. Some variants of simplified fuzzy models for determining the severity of respiratory distress and the developed application have been reported in [2]. In this paper possibilities of further implementation and development of the application are discussed.

Considering nonadditive integrals in the discrete finite case defines a class of aggregation functions allowing criteria interaction modelling. For this the concept of fuzzy measure (or the Choquet capacity) is used. The Choquet integral has been proposed as an aggregation operator for information fusion and pattern classification [3]. An algorithm for discriminative training of Choquet integral based fusion operators is described in [4]. This new algorithm is applied to a landmine detection problem, and compared to other techniques.

Fuzzy sets theory generated among other things, a widely accepted application, fuzzy-logic controllers , but also gives powerful motivation for aggregation operators research with results of importance not only for fuzzy modelling. Some results of that research have also been applied for evaluation of disease severity, for example in peripheral neuropathy [5].

Section 2 of this paper considers a practical problem, the problem of determining the severity of respiratory distress in a patient in an intensive care unit. The theoretical basis of the model in which the Choquet integral is used, is given in Section 3. The model is based on fuzzy measures theory and interaction representation. Section 4 treats the results. In Section 5, the problem of Internet implementation of considered problem is discussed. Finally, the conclusions are given, and possible further work is pointed out.

2. PROBLEM CONSIDERED

A medical application of a fuzzy multicriteria decisionmaking model is considered. Criteria used for an early diagnosis of ARDS (acute respiratory distress syndrome), [1], [2] include:

- clinical aspects of breathing (*Breathing*);
- chest radiograph (*Rö*);
- the arterial partial tension of oxygen (PaO_2, mmHg);
- the arterial partial tension of carbon dioxide ($PaCO_2$, mmHg);
- alveolar-arterial oxygen tension difference ($A\text{-}aDO_2$, mmHg).

Those criteria are used as an example. It is obvious that an implementation of a real medical system requires an extension of a number of considered criteria.

The progression of changes through various phases of ARDS is given by Table I.

In Table I., in column "*Phase*", N is the normal condition of a patient, I – is the first (the least severe) phase of respiratory distress (injury and resuscitation), II – the second phase of respiratory distress (subclinical), III – the third phase (established respiratory distress), and IV – the fourth

phase of distress (severe respiratory failure), which, due to medical reasons is not considered in the paper.

The features *Breathing* and *Rö* are expressed verbally, and are given a subjective membership degree (fuzzy sets theory). Other features are characterized by approximate intervals of numerical values. For these features to be interpreted as fuzzy sets, such as 'x is approximately in the interval [b, c]', they can be characterized by an ordered quadruple $A = (a, b, c, d)$, a fuzzy trapezoidal number, [1].

Table I. Decision-making parameters for determining the severity of ARDS

Phase	*Breathing*	*Rö*	PaO_2	$PaCO_2$	$A\text{-}aDO_2$
N	-	-	80 – 100	35 – 45	5 – 10
I	normal	no changes	70 – 90	30 – 40	20 – 40
II	mild to moderate tachypnea	minimal infiltrates	60 – 80	25 – 35	30 – 50
III	increasing tachypnea	confluence of infiltrates	50 – 60	20 – 35	40 – 60
IV	obvious respiratory failure	generalized infiltrates	35 – 55	40 – 55	50 – 80

Note that features *Breathing* and *Rö*, having in mind an implementation of usable medical system and in order to get an adequate score for a patient for those criteria, require the development of two pattern recognition systems. The first of these systems would be used for describing *Breathing*, and the other one (a kind of computer vision system) for providing data to a fuzzy model about the feature *Rö*.

Characteristic values of the criteria for determining the severity of respiratory distress, given by Table I, represented by fuzzy intervals formed on the basis of experience, are given in Table II.

Table II. Fuzzy decision parameters for determining the severity of ARDS

Phase	PaO_2	$PaCO_2$	$A\text{-}aO_2$
N	(70,80,100,110)	(30,35,45,50)	(0,5,10,15)
I	(50,70,90,110)	(25,30,40,45)	(10,20,40,50)
II	(40,60,80,100)	(20,25,35,40)	(20,30,50,60)
III	(40,50,60,70)	(10,20,35,45)	(30,40,60,70)
IV	(30,35,55,60)	(30,40,55,65)	(40,50,80,90)

When considering ARDS syndrome in [1], all the criteria (symptoms) are of equal importance. Then Bellman- Zadeh`s decision-making principle, [6], can be applied,: for m criteria C_j, $j = 1, \ldots, m$, and object $\mathbf{x}$, described by the sequence of feature values (performances), the overall performance of an object is determined by its weakest manifestation. From the set of these alternatives D_i, $i = 1, \ldots, N$, the alternative (decision) D^* is chosen, where an object $\mathbf{x}$ has maximal overall performance taking into account all of the criteria. In a fuzzy case, the membership degree value $\mu_j(\mathbf{x}) \in [0,1]$, $j = 1,2, \ldots m$, indicates the

point at which a criterion C_j is fulfilled by an object **x**, i.e., by the object's *j*-th performance. The value $\mu_j(\mathbf{x})$ represents the score for *j*-th criteria.

The degree by which an object **x** maximally satisfies all the criteria, is the wanted decision (the phase of syndrome), for $i = 1, 2, \ldots, N$:

$$D^*(\mathbf{x}) = \max_i \{\min[\mu_1(\mathbf{x}), \ldots, \mu_m(\mathbf{x})]\}. \qquad (1)$$

In the considered problem, the criteria are, Table I: C_1 –*Breathing*, C_2 – *Rö*, C_3 – PaO_2, C_4– $PaCO_2$, and C_5 –*A-a*DO_2, $j = 1, 2, \ldots, 5$. According to considered features (symptoms) with respect to the established criteria, and using the described procedure, a patient's condition is classified into one of the following phases (alternatives): D_1 – *N* (normal conditions), D_2 - *I*, D_3 - *II*, D_4 - *III*, (and D_5 - *IV*), i.e. $i = 1, \ldots, 5$.

Membership degrees $\mu_j(\mathbf{x})$, $j = 1, 2, \ldots, 5$, for a phase, for the given numerical values of symptoms, indicated by Table I, are determined by using trapezoidal membership function, [1], and given by Table II.

3. THE FOUNDATION OF THE ALTERNATIVE APPROACH

As an alternative to the usage of the Bellman-Zadeh's decision making principle, a nonadditive set function on a finite set of criteria can be used. This allows weight defining, not only on each criterion, but also on each subset of criteria, allowing criteria interaction modelling. For this the concept of fuzzy measure (or the Choquet capacity) is used.

A fuzzy measure (or the Choquet capacity) on C = {C1, ...,Cm} is a monotonic set function $\mu : P(C) \rightarrow [0,1]$, where P(C) is the power set of the set C, with $\mu(\emptyset)=0$ and $\mu(C)=1$. Monotonicity means that $\mu(S) \leq \mu(T)$, whenever $S \subseteq T \subseteq C$. An interpretation of $\mu(S)$ can be that it is the weight related to the subset S of criteria.

For this situation, a suitable aggregation operator is the discrete Choquet integral [3], [7].

Given μ, the Choquet integral of $x \in (R^+)^m$ with respect to μ is defined by

$$Ch_\mu(\mathbf{x}) := \sum_{i=1}^{m} (x_{(i)} - x_{(i-1)})\mu(\{(i),\ldots,(m)\}). \qquad (2)$$

In (2) $(\cdot)$ in indices means a permutation of the elements of *C* such that $x_{(1)} \leq \ldots \leq x_{(m)}$ and $x_{(0)} = 0$. Given patient's symptoms, using the Choquet integral, the rank of phases can be obtained, and the phase with the maximum global evaluation is the ARDS phase of the considered patient. The Choquet integral allows expressing physician's preferences.

4. THE EXAMPLE

For a patient whose condition is described by data given in Table III, decision-making table is given by Table IV. Using the model from [1], (1), respiratory distress is determined as being in phase II.

If the physician's preferences are: "features *Breathing* and *Rö* are less important than others, and features PaO_2 and $PaCO_2$ must be favored", a synergy between the criteria exists. Symptoms PaO_2 and $PaCO_2$, when considered together, have greater importance than when considered separately. Expressed by a fuzzy measure these preferences could be:

For the first preference:

μ (*Breathing*)= μ (*Rö*)=0.1

$\mu(PaO2)= \mu(PaCO2)= \mu(A\text{-}aDO2)=0.2$

Table III. Symptoms (Features) for a patient

Breathing	Ro	PaO2	PaCO2	A-aDO2
increasing tachypnea	confluence of infiltrates	50	32	31

Table IV. Decision-Making table for a patient

Phase	Breathing	Ro	PaO2	PaCO2	A-aDO2
N	0	0	0	0.4	0
I	0	0	0	1	1
II	0.2	0.3	0.5	1	1
III	1	1	1	1	0.1

For the second preference:

$\mu(PaO_2,PaCO_2)=0.5> \mu(PaO_2) + \mu(PaCO_2)=0.4$

The idea is that superadditivity of the fuzzy measure implies synergy between criteria, and subadditivity implies redundancy.

Note that it is up to the physician to scale these values to the extent that he feels they express the importance and interaction. For instance, for the first preference, he also could have increased $\mu(PaO_2)$, $\mu(PaCO_2)$, $\mu(A\text{-}aDO_2)$ above average instead of decreasing μ(*Breathing*) and μ(*Rö*), or he could have done both.

Back to the example: for the Choquet integral for each phase to be calculated, it is also necessary to define *v* for the required subsets of criteria: μ(*Breathing, Rö,* PaO_2, $PaCO_2$), μ(*Rö,* PaO_2, $PaCO_2$, *A-*aDO_2), μ(PaO_2, $PaCO_2$, *A-*aDO_2), $\mu PaCO_2$, *A-*aDO_2). Here, the superadditivity of the fuzzy measure is also used to imply synergy between criteria, and subadditivity to imply redundancy.

Preferences in the example imply that there is only synergy between criteria PaO_2 and $PaCO_2$. Additivity implies no synergy or redundancy between criteria.

Respiratory distress is determined as being in phase III (Ch_μ(III)=0.73) what is different from the case when symptoms maximally satisfy all the criteria.

A physician, a specialist for ARDS would say which model is more adequate.

5. FUZZY MEASURE IDENTIFICATION

The *ARDS Advisor* application based on the previous discussion was developed, [2]. The first step in using the application (Fig. 2) is entering the values for the 5 criteria widely used for determining the phase of respiratory distress syndrome: clinical aspects of breathing (*Breathing*); chest radiograph (*Rö*); the arterial partial tension of oxygen (PaO_2); the arterial partial tension of carbon dioxide ($PaCO_2$); alveolar-arterial oxygen tension difference ($A\text{-}aDO_2$, mmHg). The first two symptoms are described verbally, and the physician is required to enter the degree of truth for the given statements on a scale 0 to 1 (zero being absolutely false, one being absolutely true).

For the three remaining symptoms, the measured values for a patient are to be entered.

Phases of respiratory failure are modeled as fuzzy trapezoidal numbers. For input values, membership degree to each phase for each symptom is calculated. By clicking the *Submit* button (Fig. 2), user gets the first result computed using the Bellman-Zadeh`s decision-making

principle. By clicking the *Next* button, the user is taken to the next step, where the input of additional parameters, which are needed for the Choquet integral, starts.

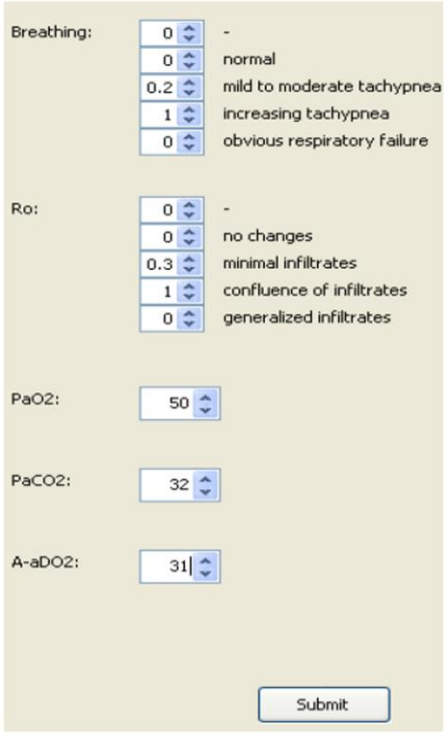

Figure 1. The first step: (Table 3 entrees)

Here, the user is first asked to enter the importance of each individual symptom. For this, a scale of 0 to 5 is proposed as intuitively most suitable, and all the values are initially set to 1 (the average value of equally important symptoms). An equally valid result will be obtained by using some other scale, for instance, from 0 to 1. By clicking on the *Next* button, indices of

importance of individual symptoms which the user has entered are shown, as well as indices of importance of all the possible combinations of symptoms, calculated by using the ordinary addition operation. (Case when there is no interaction between symptoms: importance of criteria *A* and *B* combined = importance *A* + importance *B*) (Fig. 3).

User scales these values in order to express interaction between criteria using the following principle: if the symptoms are redundant to some degree, then the index of their combined importance is to be decreased; if the symptoms have some degree of synergy, then the index of their combined importance is to be increased.

It is up to the user of the application to scale these values to the extent he feels they express the interaction. If the user submits parameters incorrectly (for instance: submitted importance of parameters *PaO*2 and *PaO*2 combined is smaller than individual importance of one of them), a prompt will be issued to correct these parameters.

Although the user is given the calculated values for the non-interacting criteria, and is asked to scale them in order to express interaction, this is still a tedious task. Also, if the number of symptoms for *ARDS Advisor* application was to grow, rather than improving the application, this could lead to an unusable application instead – that is, the task of scaling alone would be too much to grasp.

The first improvement that deals with this problem could come from the fact that representation through interaction indices is closer to the decision maker's mind than the usual measure representation.

Interaction index between any number of criteria was defined by Grabisch , [10], [11], based on [9]:

$$I^{\mu}(K) := \sum_{L \subset [m] \setminus K} \frac{(m-|L|-|K|)!|L|!}{(m-|K|+1)!} \sum_{L' \subset K} (-1)^{|K|-|L'|} \mu(L' \cup L), \quad \forall K \subset [m], \qquad (3)$$

where $[m]$ is the index set $[m] = [1,2, ..., m]$, $m \setminus K$ is a difference between sets $[m]$ and K , a coalition of elements from $[m]$ that does not include elements from K, $|L|$ is a cardinal number of a set L , and other notations are obvious.

The second, perhaps more significant improvement could come from use of algorithm for identifying fuzzy measures. Attempts in this direction already exist: algorithm of Mori and Murofushi [12], and algorithm developed by Grabisch [8]. In [8], algorithm which takes advantage of the lattice structure of the coefficients (Figure 4) is introduced.

For the description of the algorithm, the following terminology is used:

The lattice of a fuzzy measure is made from nodes related by links. The lattice has $n + 1$ horizontal layers, numbered from 0 (for the layer containing only $\mu_{\emptyset}$) to n (for the layer containing only μ_X). A path is a set of chained links, starting from the node $\mu_{\emptyset}$ and arriving to the node u_X (on figure 4, the path passing through μ_3, μ_{23} , and μ_{234} is emphasized). For a given node in layer l, its lower neighbors (resp. upper neighbors) are the set of nodes in the layer l-1 (resp. $l + 1$) linked to it. There are l lower neighbors and $n - l$ upper neighbors. The author of the algorithm describes it in 2 steps as follows:

step 1: for a given datum x, we modify only the coefficients on the path involved by x in order to decrease the error, as in a gradient descent algorithm. The modification is done in

order to preserve the monotonicity property *on the path*. Also, monotonicity is checked for neighboring nodes. This is done for all learning data, several times.

step 2: if there are too few learning data, then some nodes may have been left unmodified. These nodes are modified here in order to have the most equilibrated lattice, i.e. distance from neighbors should be as equal as possible.

The idea behind the step 2 is that in the absence of any information for some nodes, they should be arranged into the lattice in order to get a lattice as homogeneous as possible.

In case of *ARDS Adviser* application [2], the use of this (or a similar) algorithm for identifying fuzzy measures could lead to more 'user-friendly' application, especially when considering the possibility of improving *ARDS Adviser* application by adding even more symptoms to be considered when determining ARDS phase for the patient.

6. CONCLUSIONS

The application of fuzzy systems gives formal basis for modelling and developing decision-making systems based on soft computing, that can use verbally expressed experience in the considered area. In this paper, the practical implementation of fuzzy systems, that deals with independent, equally important information (the Bellman- Zadeh approach), and, also, with interacting, not equally important information (the Choquet integral), is described. Possible use of the considered approaches in telemedicine is shown. This may be an application of interest for considering some other aggregation functions.

Including interaction among criteria gives more flexible model of ARDS, and allows expressing physician's preferences in connection to the complex phenomena of criteria and subsets of criteria interaction.

ARDS Advisor can be further improved, the first by enlarging the number of considered parameters, and then by using it on larger number of cases. Other improvements are also possible, for instance, improvements can be obtained by including interaction indices, as well as identification of fuzzy measures. Intelligent medical computer aid in case of ARDS is welcome in clinical practice. It can alarm the physician, remind him of the possibility of patient developing ARDS, and signal the degree of respiratory distress. The final decision remains with the physician. To make considered application closer to clinical practice, pattern recognition systems for *Breathing* and chest radiograph images should be implemented.

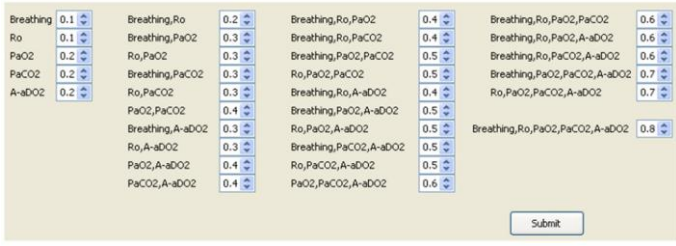

Figure 2. Indices of importance

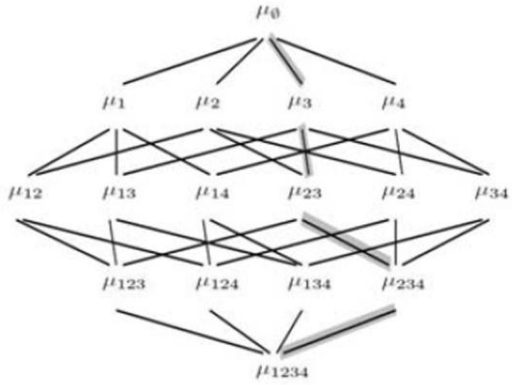

Figure 3. Lattice of the coefficients of a fuzzy measure (n=4)

The application of fuzzy systems offers a mathematical basis for developing medical systems more attuned to human cognitive processes – i.e., systems of soft computing which can make use of verbally expressed medical experience in the domain.

Reconciling and delivering relevant medical knowledge to practitioners using Internet technology are issues of universal importance. Within this context, the Java programming language is a candidate for developing distributed intelligent application, e-health application available on a variety of computing platforms, and enabling the users to make use of the (multimedia) information they have access to. The presented fuzzy models enables a distribution of not-so-well-structured information.

REFERENCE

[1] D. M. Velašević, D. Z. Šaletić, S.Z. Šaletić, "A fuzzy sets theory application in determining the severity of respiratory failure", *International Journal of Medical Informatics*, 63, 2001, pp. 101-107.

[2] D.Z.Šaletić, M.Savić, Software Implementation of Fuzzy Decision-Making in Determining the Severity of Respiratory Distress, Ed. Aniko Szakal *Proceedings of the 6th Symposium on Intelligent Systems and Informatics*, SISY 2008, September 26-27, 2008, Subotica, Serbia, 27- sisy2008-submission.pdf

[3] M. Grabisch: Fuzzy integral in multicriteria decisionmaking, *Fuzzy Sets and Systems*, 69, 1995. pp. 279-298.

[4] Mendez-Vazquez A.., P. Gader, J. M. Keller, K. Chamberlin Minimum Classification Error Training for Choquet Integrals With Applications to Landmine Detection, *IEEE Trans. Fuzzy Systems*, 16(1), February 2008, 225 – 238

[5] Dujmović, J.J. Evaluation of Disease Severity and Patient Disability Using LSP Method, *L. Magdalena, L. Ojeda- Aciego, J.L. Verdegay (eds.) Proceedings of IPMU 2008*, pp. 1398-1405

[6] Bellman, R. E., L.A. Zadeh: Decision-making in a fuzzy environment, *Management Sciences*, Ser.B 17(1970), 141- 164

[7] M.Grabish, J.L. Marishal, R. Mesiar, E. Pap, *Aggregation Functions*, Cambridge University Press, 2009

[8] Grabisch, M.,. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In: *Proc. Joint Conf. 4th IEEE Internat. Conf. on Fuzzy Systems and 2nd Internat. Fuzzy Engineering Symposium,* pp. 145-150.

[9] Murofushi, T. and S.Soneda. Techniques for reading fuzzy measures (iii): interaction index. In *Proc. 9th Fuzzy System Symposium*, pages 693-696, Sapporo, Japan, May 1993. In Japanese.

[10] Grabisch, M.. *k*-order additive discrete fuzzy measures and their representation, *Fuzzy Sets and Systems,*92, 1997., 165-

[11] Grabisch, M.. Alternative representation of discrete fuzzy measures for decision making. *Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems*, pages 587-607, 1997.

[12] Mori, T. , T. Murofushi, An analysis of evaluation model using fuzzy measure and the Choquet integral, *5th Fuzzy System Symposium*, Kobe, Japan, June 2-3 1989 (in Japanise).

Multi-Label Classification Experiments with Co-Training Based-Algorithm

Slivka Jelena, Kovačević Aleksandar, Konjović Zora
Faculty of Technical Sciences/Computing and Control Department, Novi Sad, Serbia

Abstract: **In many classification problems we cope with the lack of sufficiently large training set indispensable for correct classification. In order to eleviate the problem of laborious hand-labeling of a large training set, a set of thechniques called semi-supervised learning was proposed. Co-Training is the most widely used semi-supervised learning technique. However, it implies a problem setting in which the description (feature set) of each example can be naturally partitioned into two distinct feature sets (views), which is a property rarely found in real-life data, and moreover, it is designed to work with binary datasets. In this paper we conduct an empirical study of applicability of Co-Training on single-view multi-label datasets. We modified Co-Training to work with multi-label data and tested it on several standard single-view multi-label UCI datasets, using several Co-Training settings: random split of features, artificial maxInd split of features and Random Split Statistic Algorithm (RSSalg) we have developed earlier to address the problem of Co-Training application on single-view datasets. In the experiment conducted in this paper we confirm that, provided sufficient redundancy in the data, Co-Training applied with random split of features may be beneficial even in the case of multi-label datasets. We show that RSSalg significantly exceeds the performance of Co-Training settings it was compared by.**

1. INTRODUCTION

In order to automate a categorization task, we must firstly define a list of categories and prepare the sample of instances labeled (classified) with respect to their membership in one of the predetermined categories (classes), which we refer to as the training set. Then, a learning algorithm induces, from these examples, a general description of predefined classes. Our goal is to obtain a description with highest generalization ability possible, capable to predict (with high degree of success) the class of previously unseen examples. To achieve this, a training set which contains a large amount of various data is necessary.

However, in many classification problems we cope with the lack of sufficiently large training set indispensable for correct classification. This is due to the fact that hand-labeling requires human intellectual effort and is a time consuming and tedious process. In order to overcome the problem of laborious hand-labeling of a large training set, a set of techniques called semi-supervised learning, which induce classifiers from only a small number of labeled data and a sufficiently large number of unlabeled data, was designed. These techniques result in performances in the rank or even better then the performances of supervised techniques which use training sets labeled by experts only.

Co-training is one of the major semi-supervised learning methods. It implies a problem setting in which the description (feature set) of each example can be naturally partitioned into two distinct feature sets (views). For example, a web page can be described by the words occurring on that web page, and also by the words occurring in hyperlinks that point to that page.

Basically, Co-Training employs two base classifiers, trained on different views, which use unlabeled data to boost each others performance. It was first introduced by Blum and Mitchell in [1] where they both theoretically and experimentally show that the Co-Training algorithm works well on semi-supervised learning, if the feature set division of dataset satisfies two assumptions: (1) each set of features is sufficiently strong (redundant, i.e. the learner can learn the high prediction accuracy classifier using each view individually), and (2) the two views are conditionally independent given the class (i.e. for each example the features from the first view are not related to the features from a second view, except through the class of the example).

However, in the great majority of practical settings, the natural split of features that describe the dataset does not exist, limiting the applicability of co-training and leading researchers to look for the ideal artificial split which can be used by Co-Training.

The original Co-Training algorithm is designed for two-class problems. Although Co-Training algorithm can be easily modified to work with multicategory data, most research on Co-Training has focused on small, often binary, classification problems and it is not clear whether their conclusions would generalize to real-world classification tasks with a large number of categories [2]. A reason for this could be the need for appropriate feature split, which requires consistency of two feature sets (i. e. the target functions over each feature set predict the same label for most examples, for example, the prediction of category of a web page should be identifiable using either the page text or hyper link text), making it difficult to apply Co-Training to multiclass classification problem [3].

In practice, multi-class problems are often tackled by their decomposition into multiple binary problems and then Co-Training is used for learning individual binary classification problems, or even by converting multicategory datasets to binary by grouping some classes as positive and the rest of the classes as negative, which leads to neglecting some of the classes important in practice.

In this paper we have conducted an empirical study which extends our preliminary research in [4], with the goal to provide some empirical results for Co-Training application on multi-label datasets. We handled the multi-class cases directly with Co-Training, without reducing them to multiple two-class problems. In order to cope with the lack of natural feature split, we have tested several Co-Training settings that do not require the knowledge of the natural feature split: Co-Training applied with random split, artificially created *maxInd* feature split used in [5], and also Random Split Statistic Algorithm we have developed earlier in [4].

We have tested these settings on several single-view multicategory UCI (University of California, Irvine) datasets[1], using a 10-fold-cross validation procedure against the performance of Naive Bayes classificator trained on the labeled version of whole dataset.

Our experiment showed that Co-Training, modified to work with multi-label datasets and used with random and maxInd split, shows improvement in accuracy for two out of three datasets we tested it on, while it degrades the performance for the third dataset. Our result confirmed that the conclusion of [6] (even when there is no natural feature split, provided that there is sufficient redundancy among the features, and a fairly reasonable division of the features can be identified, then co-training algorithms can be beneficial) also stands in the case of multilabel datasets.

[1] http://archive.ics.uci.edu/ml/

MaxInd split did not proove to be significantly better then the random split, but this was expected as Feger et al. reported in [5] that maxInd does not perform it's best in combination with NaiveBayes classifier.

Our experiment confirmed the good performance of RSSalg applied on multi-label categorization task. RSSalg outperformed the settings it was compared by on all three datasets, having acchieved improvement in accuracy even on the dataset where other setting failed.

This paper is organized as follows: section 2 presents the related work. Section 3 presents the Co-Training algorithm modified to work with multi-label datasets. Sections 4 and 5 show the experiment conducted in this paper and achieved results. Finally, chapter 6 concludes this paper and gives possible direction of future work.

2. BACKGROUND

We may say that this paper connects two lines of previous research in the area of Co-Training algorithm – a search for ideal artificial feature split that can be used by Co-Training, and the problem of handling multi-label datasets with Co-Training.

Nigam and Ghani [6] have reported an empirical study on splitting single views into two views. They have found that even when there is no natural feature split, if there is sufficient redundancy in the data, and a fairly reasonable division of the features can be identified, then Co-Training algorithms may show similar advantages to other algorithms. These results have inspired many researchers to use random feature split for Co-Training in practice, but, in order for Co-Training to perform its best, a natural split of features is still needed.

Du et al. [7] stressed out that Nigam and Ghani's [6] experiments were conducted on two-class text categorization problems, while it is well-known that text data possesses large redundancy among features, so there is no guarantee that their conclusions also apply to common data (such as UCI data). They have extended their empirical study by designing several methods for splitting a single view dataset into two views. Their results show that, given the large training set, their view verification and splitting methods perform well, but in the case of small labeled sets, the two Co-training assumptions are difficult to verify, and view splitting is unreliable. Their experiments were conducted on standard UCI datasets [8]. However, in their experiments, datasets with multiple classes are converted to binary by using the majority class value as one class, and the rest of the other values as the other class. This leads to neglecting some of the classes, especially in fairly balanced datasets, and there is no guarantee that their conclusions extend on to multi-label datasets.

In order to enable successful Co-Training application on single-view datasets we have developed earlier in [4] the Random Split Statistic Algorithm (RSSalg). RSSalg runs the Co-Training algorithm for a predefined number of times, each time using a different random feature split. As the result, the training set produced by Co-Training is different in each run and it consists of initially labeled examples and examples labeled in Co-Training process. All of these Co-Training results are processed by selection of the examples that appear in most resulting training sets (the ones that exceed *example threshold*) and for which most of the resulting Co-Training classifiers agree on the label (the ones that that exceed *label threshold*). The final training set is formed from these selected results and it is used for learning a model

with much higher classification performance then the initial model learned only on labeled examples.

Feger et al. [5] applied the method called *maxInd* to create the artificial feature split to use with Co-Training algorithm. *MaxInd* algorithm is designed to create two views maximaly independent given the class, based on the second requirement given by Blum and Mitchell in [1] for successful Co-Training application. It is based on Conditional Mutual Information (CondMI) [9] as the measure for determination of views maximally independent given the class. Their experiments showed that the performance of Co-Training does not always become better by simply choosing truly independent views. In this paper we use maxInd algorithm as the benchmark for the performance of our *RSSalg*, since they have the similar goal of applying Co-Training on single-view datasets and provide some more empirical results for this algorithm, as to our knowledge, it was only tested on two-class problems so far.

Ghani in [2] presented an approach for multi-class classification that combines labeled and unlabeled data. Multi-label data was handled by decomposition of multi-class problem into multiple binary problems using error-correcting output codes, and then Co-Training was used to learn the individual binary classification problems. Ghani has shown that his methodology is especially useful for text classification problems involving a large number of categories.

3. THE MULTILABEL CO-TRAINING SETTING

The standard Co-Training algorithm [1] implies a problem setting in which the description (feature set) of each example can be naturally partitioned into two distinct feature sets (views). Co-Training uses each of these views to train a different classifier. These two classifiers are then applied to the set of unlabeled examples and each of them labels those examples from the set for which it predicts the label with most confidence. Examples labeled in such a way are then added to the initial training set and both classifiers then re-learn on the enlarged training set so that they take into account the newly added (and previously unlabelled) data. The described process is repeated for a predefined number of iterations and it can be regarded as employing two base classifiers that use unlabeled data to boost each others performance.

Co-training begins at using a weak initial hypothesis over each feature set. Though the classification noise from the weak hypothesis would be brought to the other classifier, the algorithm can learn from these labeled examples by an iterative procedure between the two classifiers, resulting in classifiers trained on both labeled and unlabeled data with significantly improved performance.

After Co-Training process, a previously unseen (test) example can be classified by multiplying the class probabilities output by resulting classifiers, and choosing the class predicted with most confidence.

The original Co-Training algorithm labels examples with respect to their membership in one of the two classes – positive and negative. In the experiment conducted in this paper we modified Co-training algorithm to work with multiple label data. This was achieved by choosing inner classifiers which can handle multi-label data and allowing each inner classifier in each iteration to label a predefined number of examples for each class for which it predicts that class with most confidence.

4. DATASETS

In order to study behavior of RSSagl, as well as different Co-Training settings, we chose 3 multi-label UCI datasets coming with the WEKA package [11] – OptDigits (Optical Recognition of Handwritten Digits), Splice (Splice-junction Gene Sequences) and Segment (Image Segmentation Data Set).

These datasets were chosen as they were already used for Co-Training evaluation [7, 12, 13] and they are balanced datasets so converting them to binary datasets by grouping classes together would lead to neglecting important classes. The basic properties of these datasets are given in Table 1.

Table 1. The basic properties of datasets used in our experiments

Dataset	Attributes	Instances	Classes
OptDigits	64	5620	0 (554), 1 (571), 2 (557), 3 (572), 4 (568), 5 (558), 6 (558), 7 (566), 8 (554), 9 (562)
Splice	60	3190	EI (767), IE (768), N (1655)
Segment	19	2310	brickface (330), sky (330), foliage (330), cement (330), window (330), path (330), grass (330)

5. EXPERIMENTAL RESULTS

In order to compare the performance of RSSalg to the performance of other Co-Training settings, we adopted the stratified 10-fold-cross validation described in [5].

The standard procedure of 10-fold-cross validation divides the experimental data in 10 folds, and in each of the 10 rounds one different fold (10% of the data) is selected to be used for testing, while the remaining 9 folds (90% of the data) are used for training.

Co-Training setting uses only a small amount of both labeled and unlabeled data, thus applying the standard 10-fold-cross validation procedure on Co-Training would result in many examples to be omitted from both testing and training data. A better utilization of the available data is to increase the size of the test set which will improve the evaluation of the classifier without significantly reducing its quality.

Thus, in the experiment, data is divided in 10 stratified folds. In each round of 10-fold-cross validation process, a different fold is selected for random selection of required number of labeled training examples. The remaining data from that fold, as well as 5 adjacent folds are used as unlabeled training examples, and finally, the remaining 4 folds are used as testing examples. In this way, in each round, 60% of the data is used for training and remaining 40% of the data is used for testing. Each fold is used exactly once for the selection of labeled data, five times it is included as unlabeled data and four times it is used as a part of the testing set.

The measure of performance used in this experiment is the accuracy measure:

$$accuracy = (tp + tn)/(tp + fp + tn + fn),$$

where *tp* and *tn* represent the number of examples correctly predicted as positive and negative, respectively, and *fp* and *fn* represent the number of examples falsely predicted as positive and negative, respectively. This measure is widely used in research considering Co-Training. In this experiment we do not consider other measures, such as precision and recall, because the datasets used are highly balanced and none of the classes is dominant.

The base classifier used in Co-Training algorithm is Naive Bayes. This classifier was chosen both for its speed (which is an important factor due to the complexity of RSSalg) and high accuracy it achieves on benchmark datasets.

The experiment was performed on several UCI datasets described in Section 6. The baseline accuracies of these datasets, achieved when classifying all test examples to the majority class of the training set are given in Table 2.

Table 2. Baseline accuracies for gold-standard datasets

Dataset	Baseline accuracy
OptDigits	10.2%
Splice	51.9%
Segment	14.3%

For comparison, three different settings are used: a traditional supervised Naive Bayes classifier trained only on labeled examples, Co-Training setting using a random split created by random selection of half of the features in the first view and using the remaining features as the second view, Co-Training setting using a *maxInd* split which was created as described in chapter 4, and finally, the *RSSalg*. Implementation of these settings relies on WEKA[2] framework as in [5].

In order to ensure that Co-Training can possibly show improvement, a small initial training sets for each dataset are chosen in a manner that "optimal gain"[3] is large enough (~20%), similar as in [7]. When forming a training set, number of examples chosen for each class are proportional to the class distribution in the dataset. The number of examples per each class labeled by Co-Training inner classifiers, h_1 and h_2 in each iteration are also chosen proportional to the class distribution in the dataset. Table 3 shows, for each dataset, the size of initial training set $|L|$, accuracy achieved on classifying the test set when only learning on initial training set *Lacc*, accuracy achieved on classifying the test set when learning on the whole dataset *Allacc*, and number of examples added to the training set in each iteration of Co-Training *noPerClass*.

[2] http://www.cs.waikato.ac.nz/ml/weka/

[3] The "optimal gain" represents the difference between the accuracy acchieved on the whole dataset (initial labeled examples and unlabeled examples with correct label) and the accuracy on initial labeled examples alone.

Table 3. Size of the initial training set, accuracy achieved when learning on initial training set, accuracy achieved when learning on the whole dataset and number of examples added to the training set in each iteration of Co-Training, for each dataset

Dataset	$\|L\|$	*noPerClass*	*Lacc*	*Allacc*
OptDigits	60	30	72.6	91.7
Splice	20	40	63.1	95.3
Segment	21	7	67.3	80.8

The size of the unlabeled pool $|U'|$ is 50 ($u=50$) for each dataset. Number of iterations used in Co-Training algorithm (in each of the settings) is 20 ($k=20$). The number of different random splits used in RSSalg is set to be 30. Label threshold and example threshold, used in RSSalg, for each dataset are listed in Table 4. All of these settings were empirically determined to be the best solution.

Table 4. Label threshold and Example threshold used in RSSalg for each dataset given in percentages

Dataset	Label threshold	Example threshold
OptDigits	68	0
Splice	64	31
Segment	70	40

Table 5 presents the average accuracy achieved on the test set in 10-fold-cros validation runs for different UCI datasets. The column "Final accuracy" in Table 5 shows the final accuracy achieved by each algorithm. For Co-Training setting with random split, for the better comparison with RSSalg, we have displayed the accuracy achieved at the end of Co-Training algorithm, i. e. after its 20_{th} iteration, averaged on 30 random splits (the same 30 splits used in RSSalg). For Co-Training with maxInd feature split this is the accuracy achieved at the end of Co-Training algorithm, i.e. after its 20_{th} iteration. For RSSalg this is the accuracy achieved on the training set which resulted from choosing best examples from Co-Training resulting training sets (Co-Training was run on 30 different random feature splits (same splits as in Co-Training with random split setting), each time for 20 iterations). Column "Increase" in Table 5 shows the difference between the final accuracy achieved by algorithm and accuracy achieved by Naive Bayes classifier trained only on small initial labeled set *L*. Thus, the positive results show the improvement, while negative results show the degradation in accuracy. The column "Gap" shows the difference between the goal accuracy and final algorithm accuracy. We refer to the goal accuracy as the average accuracy achieved by Naive Bayes classifier trained on the labeled version of all training data (both views) in 10-fold-cross validation. The "goal accuracy" may reflect the upper bound that Co-Training can achieve in accuracy.

Our experiment showed that Co-Training, modified to work with multi-label datasets, shows improvement in accuracy for OptDigits and Splice dataset, while for the Segment dataset the performance is degraded for both random and maxInd feature split.

Table 5. Accuracy [%] of different Co-Training settings achieved on different UCI datasets

Dataset	Split	Final accuracy	Increase	Gap
OptDigits	Random	77.0	4.4	14.7
	maxInd	77.6	5.0	14.1
	RSSalg	82.4	9.8	9.3
Splice	Random	79.5	16.4	15.8
	maxInd	75.4	12.3	19.9
	RSSalg	85.9	22.8	9.4
Segment	Random	59.6	-7.7	21.2
	maxInd	58.8	-8.5	22.0
	RSSalg	69.3	2.0	11.5

In order to gain some insight why doesn't Co-Training work on Segment dataset, we have checked the conditions that might affect Co-Training performance: strength of the views and the dependence between the features of each dataset. Table 6 presents strengths of the views for each dataset for different Co-Training settings. For Co-Training with maxInd setting this is the strength of each view obtained with maxInd splitting algorithm, while for Co-Training with random split this is the averaged strength of the views obtained by 30 different random splits of features. Note that average strength of the views for RSSalg is not listed for it uses the same 30 splits as Co-Training setting with random split. As we can see from Table 6, Segment dataset has significantly lower strength of the views for each setting then the two other dataset, which may be the reason for the poor performance of Co-Training algorithm.

Table 6. Strength of the views [%] of different Co-Training settings achieved on different UCI datasets

Dataset	Split	Average Strength of the views (whole dataset)	
		View 1	View 2
OptDigits	Random	85.2	84.9
	maxInd	85.3	85.3
Splice	Random	85.8	85.1
	maxInd	86.3	80.8
Segment	Random	74.4	74.6
	maxInd	74.0	74.1

In order to check the dependence between the features, a correlation matrix was calculated for each dataset. Table 7 presents the 5 most correlated pairs of attributes for each dataset. The measure for correlation used is the Pearson product-moment Correlation Coefficient (PMCC). The higher the absolute value of the PMCC, the more correlated the variables are. As we can see form Table 7, Segment dataset has the highest correlation between the attributes. Feger et

al. [5] stated the hypothesis that the reason for the poor performance of Naive Bayes classifier is its sensitivity to the independence of the features within the views, given the class. It might be that due to the high intradependence in the Segment dataset, the initial classifiers are not able to sustain improvement.

Table 7. Top 5 most correlated pairs of attributes for each dataset

Dataset	Attribute pair		PMCC
OptDigits	input3	input59	0.927
	input2	input58	0.833
	input7	input15	0.756
	input8	input16	0.747
	input4	input60	0.739
Splice	attribute_29	attribute_30	0.348
	attribute_31	attribute_32	0.284
	attribute_32	attribute_35	0.277
	attribute_28	attribute_30	0.223
	attribute_31	attribute_35	0.203
Segment	rawblue-mean	value-mean	0.999
	intensity-mean	rawred-mean	0.998
	intensity-mean	value-mean	0.997
	intensity-mean	rawgreen-mean	0.996
	intensity-mean	rawblue-mean	0.996

Our results show the strong correlation between the features and lowest avarage strength of individual views for the Segment dataset, compared to other datasets. These results prove that the conclusion of [6] (even when there is no natural feature split, provided that there is sufficient redundancy among the features, and a fairly reasonable division of the features can be identified, then co-training algorithms can be beneficial) also stands in the case of multilabel datasets.

Co-Training with maxInd feature split has shown improvement in accuracy compared to Co-Training setting with random split only for OptDigits dataset. This could be due to the usage of Naive Bayes as Feger et al. in [5] achieved better results with *maxInd* split by using other base classifiers such as SVM and RBF net. This is not surprising as paper [14] indicates that the accuracy of Naive Bayes is not directly correlated to the degree of feature dependences measured as the class-conditional mutual information between the features.

Our experiments show that the best performance was achieved by RSSalg. Even in the case of Segment datasets, where Co-Training setting with random split shows degradation in accuracy (averaged on same 30 splits used in RSSalg), the statistic drawn from those Co-Training results in RSSalg ensures the increase in accuracy compared to accuracy achieved by Naive Bayes classifier trained only on small initial labeled set *L*.

6. CONCLUSION

This paper is an empirical study which extends our preliminary research conducted in [4]. Our goal was to provide some empirical results for Co-Training application on multi-label datasets and also to test the performance of RSSalg on several other datasets.

We handled the multi-class cases directly with Co-Training, without reducing them to multiple two-class problems. In order to cope with the lack of natural feature split, we have tested several Co-Training settings that do not require the knowledge of the natural feature split: Co-Training applied with random split, artificially created *maxInd* feature split used in [5], and also RSSalg. These settings were tested on several multicategory UCI datasets, using a 10-fold-cross validation procedure against the performance of Naive Bayes classificator trained on the labeled version of whole dataset.

Our results for Co-Training application on multi-view datasets confirmed that the conclusion of [6] (even when there is no natural feature split, provided that there is sufficient redundancy among the features, and a fairly reasonable division of the features can be identified, then co-training algorithms can be beneficial) also stands in the case of multi-label datasets.

MaxInd split did not prove to be significantly better then the random split, but this was expected as Feger et al. reported in [5] that maxInd does not perform its best in combination with Naive Bayes classifier.

We demonstrated the good performance of *RSSalg* applied on multi-label categorization task. RSSalg outperformed the settings it was compared by on all benchmark datasets, having achieved improvement in accuracy even on the dataset where other setting failed. Our task in the future remains improvement of *RSSalg* in terms of finding a more general way of determination of its parameters (i.e. example threshold and label threshold).

ACKNOWLEDGMENTS

Results presented in this paper are part of the research conducted within the Project "An lternative to eGovernment based on Open Standards and Open Code," Grant No IT13009, Ministry of Science and Technological Development of the Republic of Serbia.

The authors are thankful to Dr. Felix Feger and Dr. Irena Koprinska for sharing the details of their work.

REFERENCES

[1] Blum A. and Mitchell T., "*Combining labeled and unlabeled data with co-training*," In Proceedings of the Eleventh Annual Conference on Computational Learning theory (Madison, Wisconsin, United States, July 24 - 26, 1998) COLT' 98, ACM, New York, NY, pp 92-100, 1998.

[2] Ghani R., "Combining Labeled and Unlabeled Data for Multiclass Text Categorization*,"* Proc. 19th Int'l Conf. Machine Learning, 2002.

[3] Shinnou H. and Sasaki M., "Semi-supervised learning by Fuzzy clustering and Ensemble learning," 4th international conference on Language Resources and Evaluation (LREC2004), Academic Journal 2004., pp 399-402, 2004.

[4] Slivka J., Kovačević A. and Konjović Z., "Co-Training Based Algorithm for Datasets without the Natural Feature Split," 2010 IEEE 8th International Symposium on Intelligent Systems and Informatics (September 10-11, 2010, Subotica, Serbia) SISY 2010., 2010.

[5] Feger F. and Koprinska I., "Co-training Using RBF Nets and Different Feature Splits," In Proceedings of 2006 International Joint Conference on Neural Network, pp 1878-1885, 2006.

[6] Nigam K. and Ghani R., "Analyzing the Effectiveness and Applicability of Co-Training," 9th International Conference on Information and Knowledge Management, pp. 86-93, 2000.

[7] Du J., Ling C. X. and Zhou Z., "When Does Co-Training Work in Real Data." Knowledge and Data Engineering, IEEE Transactions on, vol PP, Issue:99, pp 1-1, 2010.

[8] Asuncion, A., and Newman, D. J. UCI machine learning repository, 2007.

[9] MacKay D., Dependent Random Variables, Information Theory, Inference and Learning Algorithms: Cambridge Univ. Press., 2003.

[10] Goldman, S. and Zhou, Y., "Enhancing supervised learning with unlabeled data," In: Proceedings of the 17th International Conference on Machine Learning (San Francisco, CA), pp. 327–334 , 2000.

[11] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The weka data mining software: an update. SIGKDD Explorations, 11, 10–18, 2009.

[12] Huang J., Shirabad J. S., Matwin S., and Su J., "Improving Co-training with Agreement-Based Sampling," In Proceeding RSCTC'10 Proceedings of the 7th international conference on Rough sets and current trends in computing, pp 197-206, 2010.

[13] Yaslan Y. and Cataltepe Z., "Random Relevant and Non-redundant Feature Subspaces for Co-training", Intelligent data engineering and automated learning - IDEAL 2009, Lecture Notes in Computer Science, 2009, vol 5788/2009, pp 679–686, 2009.

[14] Rish I., "An empirical study of the Naive Bayes classifier," IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 2001

The Strategic Importance of E-Government in the Construction of E-Society (based on the example of e-Croatia)

Andrea Tušetić, Ljerka Luić
Karlovac University of Applied Sciences

Abstract: **e-Government concept originated at the beginning of 21st century, mostly as a copy of e-commerce into public sector. All intentions were directed towards the presence of the public services on the Internet. In the early years of its development, e-government follows the evolutionary e-business evolving model, which in particular means that in the early days of e-government evolvement, primary focus of the e-services was simple appearance of graphic user interfaces with no possibilities of interactions. Early enthusiasm during the mean time weakened but such experiences brought crucial acknowledgments. Today, because of those acknowledgments, the focus is on coordination and effective assessment of the needs, efficiency and public benefits for such services. The development of electronic public services enters in the new phase, which is mostly determined by reengineering of existing processes of public government. Public sector by its nature (based on information and communications) is ideal for international increase of efficiency and quality.**

1. INTRODUCTION

Public government disappointment is triggered by bureaucracy, information abuse for internal purposes, increasing cost of transactions and mostly because of lack of responsibility for the client. Especially in European countries the problem of ever-growing public sector is present, making the concept the efficient e-government even more important. Regarding the participants engaging in e-government activities, four models can be recognized:

- G2C (Government to Citizens),
- G2B (Government to Business),
- G2E (Government to Employees), and
- G2G (Government to Government). [1]

2. E-GOVERNMENT FRAMEWORK

The new doctrine of e-government requires organisations to integrate and synchronise the strategic vision and tactical delivery of services to clients with the information technology and service infrastructure needed to meet that vision and process execution. In the next few years, successful countries will restructure their public sector, process and technology infrastructure to ensure the successful realisation of e-government. [2]

ICTs were recognized to have tremendous ‘administrative’ potential. For example, ICTs may help create a networked structure for interconnectivity, service delivery, efficiency and effectiveness, interactivity, decentralization, transparency, and accountability. Electronic government (e-government) cover all these functions and, generally speaking, refers to the intensive use of ICT in providing the citizens an improved access to information related to

public administrations as well as in providing them superb service quality. E-government transformation is one of the biggest challenges within the IT-related sector from the perspective of scale and complexity. The main objective is to adapt existing e-government to new computing requirements based on the citizens' new service concept. [3]

Acording to Ebrahim and Irani the understanding of egovernment architecture framework by public sector organisations is significance strategic phase toward reliable and effective e-government adoption. In their paper they describe how to use and manage information technologies to revitalise business processes, improve business decisionmaking, and gain competitive advantage from the adoption of e-government.

The architecture framework defines standards, identifies the infrastructure components, applications and technologies that are the guidelines for e-government adoption have highlighted the importance of integrating the existing information systems and applications in public sector organisations in order to establish an efficient framework for e-government architecture. They suggest that the architecture of e-government can be divided into four layers: access layer, e-government layer, e-business layer; and infrastructure layer. The authors also analyse significant barriers to the adoption of e-government. They classify these barriers into dimensions with practical examples that include: IT infrastructure, security and privacy, IT skills, organisational issues and cost. [4]

Acording to Lam here is a set of 17 barriers which were organised into one of four categories: strategy, technology, policy and organization. Strategy barriers include common e-government goals and objectives, delivery timeframes, and ownership and governance. Technology barriers include architecture interoperability, data standards and legacy systems. Policy barriers include citizen privacy, data ownership and policy implications. Organization barriers include pace of government reform, legacy government processes and management and technical skills. E-government is not simply a technical matter of getting IT systems to talk to each other, but something that requires strategic planning and considerable change management. [5]

The Economist [6] predicted the next Internet revolution (after e-commerce and e-business) to be e-government revolution. Contrary to such optimistic expectations regarding e-government, a disappointment on how the situation evolved in following years followed. The Gartner Group in 2004 brought analysis of e-government hype cycle in world. After "peak of inflated expectations" in 2002 "trough of disillusionment" follows, and finally "slope of enlightenment" is coming. [7]

In order to achieve cost savings in a citizen-centered government through egovernment services, governments need to know expectations of e-Government services. In order to acquire the information the following should be implemented:

- Information and service needs assessment. There is a need to systematically ask citizens about their actual information and service needs,
- Technology needs assessment. Governments cannot assume that a computer and Internet access are sufficient to engage in e-government,
- Information and technology literacy. A key aspect of egovernment service delivery and availability is the ability of citizens to successfully use e-government services,
- Government literacy. How users interact with egovernment services will depend greatly on how well they understand the structure of the government,
- Usability and functionality. There is a need to engage in an iterative design process that encompasses user assessment throughout the service's development process,

- Accessibility. Increasingly aging populations and others in populations have a range of challenges tied to their abilities to use e-government services,
- Meeting user expectations. Consistency between sites and services has not been a high priority, nor has consistency of the materials provided,
- Understanding how citizens actually use e-government services.

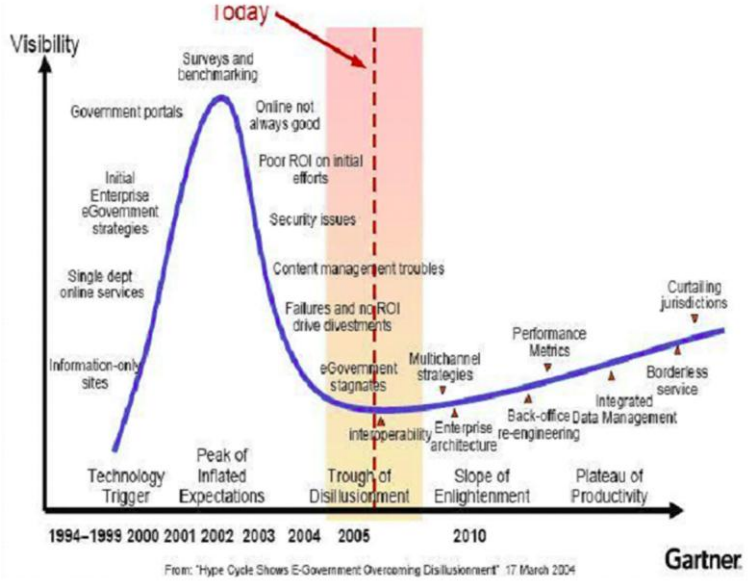

Figure 1. Gartner analysis – Hype cycle for e-Government

Failure in any of the above areas could lead to diminished use of e-government services and loss of confidence among citizens which would seriously hamper future development. [8]

The analysis discovered four key components that could encompass all initiatives in development of e-government and these initiatives are shown in Figure 2.

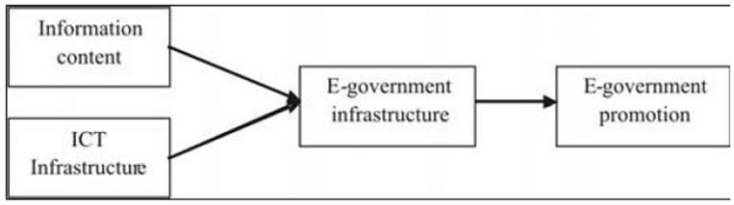

Figure 2. e-Goverment implementation framework [9]

3. CROATIAN E-GOVERMENT STRATEGY

Even so, e-Croatia is not a project that is starting from scratch. Rather, it is building on a rather strong recent history of strategies and programs that the government has implemented in the past few years, including the following:

- Information and Communication Technology Strategy (2002),
- e-Croatia 2007 (2003),
- HITRO.hr Program Strategy (2004),
- National Program on Information Security (2005),
- Open Source Software Policy (2006),
- Broadband Strategy (2006),
- National Program of Digitalization of Archive, Library and Museum Heritage (2006),
- National Strategy for Development of e-Business.

For the purpose of providing proposals, opinions and expert views on different issues important for the state development, the Government of the Republic of Croatia forms its permanent working bodies: *Ministries, Offices of the Government, Central State Administrative Offices, State Administrative Organizations* and *Public Sector*. Unit in place to design eGovernment policies and to coordinate ministries, regional and local governments in this sense is the *Central State Administrative Office for e-* Croatia of the Government of the Republic of Croatia. It is headed by the State Secretary and the two Deputy State Secretaries who are appointed to the Office. Internal structure of the *Central State Administrative Office for e-Croatia* comprises the following departments with a broad scope of activities:

- The Department for the Rationalization of Investments in Information and Communication technologies,
- The Department for the Coordination of the Implementation of the e-Croatia programme,
- The Department of International Cooperation.

The ultimate goal of the Office is to coordinate and implement activities of the e-Croatia 2007 programme to make quicker steps towards the information society following the recommendations of European Union and Lisbon Agenda. [10]

4. IMPROVEMENT OF E-GOVERMENT SERVICES IN CROATIA

The Republic of Croatia has prepared the State Administration Reform Strategy that was adopted by the Government in spring 2008. Basic elements of the Strategy represent the adoption of principles of good governance in line with EU standards, improvement of the legal system, provision of better services due to a modern, professional and trained civil service, simplification of administrative procedures and establishing e-administration. The responsibility for the implementation lays with the Central State Office for Administration. The Strategy has given great importance to the development of e-Government. The responsible institution for the implementation of e-Government is the Central State Office for e-Croatia.

The Central State Office for e-Croatia has coordinated activities in the field of e-Government trough the Program of e-Croatia already from 2003 on. All activities that are prepared in all central state bodies were collected and systemized. That generated a good overview on what was going on in the field of e-activities. It has to be said that Croatia has developed a large number of e-Government services during the past years. Just to mention some of them: e-Company, IS OSG, e-Tax services, e-Pension, e-HealthInsurance, e-Cadastre, e-LandRegister.

At this moment the developed e-services are mostly:

- Developed by one institution and cover the responsibility area of one institution,
- Developed to give information to users.

Central State Office for e-Croatia has organised a working group with the task to analyse the current situation in the specific area and to make a proposition for a strategy that will represent the basis for the future development of integrated functions. The work started in autumn 2008. The Strategy for the Development of e-Government for the Republic of Croatia for the Period 2009-2012 was finished and adopted by the Government of Croatia in January 2009. [10]

The Strategy defines e-Government as an infrastructural transformation of the functioning of state administration bodies by using the information and communication technology to direct the business activities of the public administration versus the users. That requires a permanent adjustment of the legal and technological framework of state administration bodies with the goal to reach better effectiveness, more rational use of the budget and services of higher quality. The Strategy defines ten basic principles for the development of e-Government.

- The first principle is the principle of equality. All services and information have to be accessible to all users having in mind their specific needs,
- The existing services of the administration will not be abolished. The existing services will be reduced as the number of users of the new services increases,
- The existing non electronic services of the administration can not be revoked until all users, even potential users have the access and knowledge to use the services of the electronic administration,
- The development of new services has to be based on ICT and the access to the services has to be based on different easy accessible channels,
- The information that is publicly accessible via electronic services has to be structured and delivered in a way that is simple, understandable and that grants free access,
- All information is entered just once into the e-Government system. All subjects that need the information have to communicate trough an open communication format and will get the specific information if it is in accordance with the legal framework,
- All forms of electronic recording (electronic documents, web based forms, web pages) of the administration have to be in a format that does not oblige the user to purchase commercial products to get a service,
- The basic technology for the development of electronic services has to be independent of the manufacturer of products and services regarding computers, networks and software,
- The security and reliability of developed e-services has to be in line with adopted standards for information security and legal standards for the protection of personal dana,
- The electronic services have to be supported by professionally trained experts. All employees have to have an appropriate level of computer skills in accordance with the European Computer Driving License. [11]

The Strategy has identified:

- The information and communication infrastructure,
- The data/information and documentation basis,
- Electronic services accessibility,
- Human resources,

as key elements for the development of e-Government.

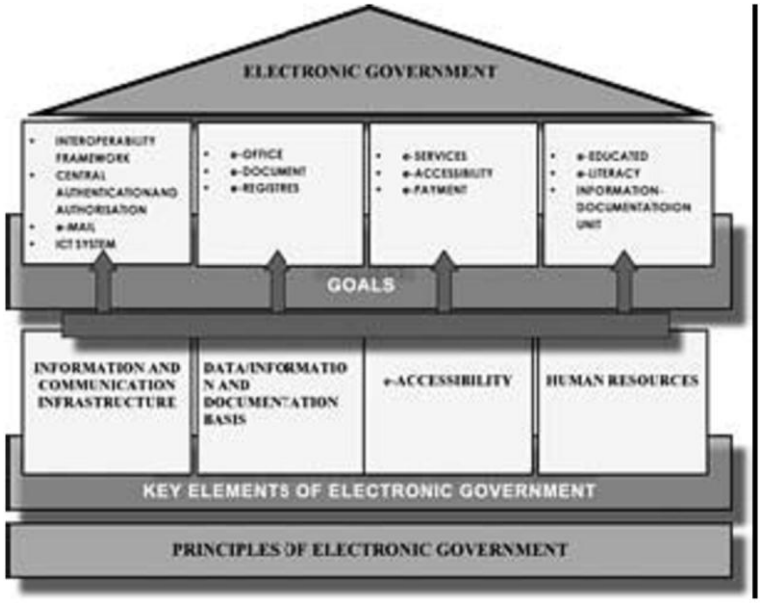

Figure 3. e-Goverment for e-Croatia [11]

4.1. The Information and Communication Infrastructure

The Strategy implies a standardization of the planning, procurement and maintenance process for hardware/network components as well as for program solutions. It will start with a detailed inventory of existing computer systems. The basis of an interconnected electronic administration represents a secure, reliable and effective information and communication infrastructure. The HITRONet network has been developed as the backbone to interconnect all governmental institutions. It is based on the Internet environment with a virtual private network model. The network has to be evaluated especially from the aspect of information security. The interconnected ICT environment enables the exchange of messages and information between users. That requires a secure identification and authentication model where it is possible to uniquely identify all participants in the electronic communication. For that a central identification/authentication and authorization system will be put in place that will enable the identification not only of users but of civil servants too. At this moment the identification/authentication and authorization method is in the responsibility area of each application developer. The interconnected environment requires standards to be put in place. The communication has to be based on interoperability standards defined on the process, semantic and technological level. The interoperability framework will be one of the most important outcomes of the Strategy. It is important to create an efficient, secure, understandable and simple multichannel environment for electronic services of all public bodies including the delivery via internet, mobile phone, digital television and the support for the information agent. The precondition for a successful communication between the state administration bodies as well as the communication with costumers is a secure, reliable and legally defined electronic mailing system.

4.2. The Data/Information and Documentation Basis

Data and information incorporated into documents represent the basis of all activities of the state administration bodies regardless of the way they are conducting their businesses (electronic or non-electronic). To consolidate that area the introduction of an unambiguous system for the management of electronic documents based on the experience of countries with good practises in this field is foreseen. The goal is to prepare a unified system for the exchange of documents between all stakeholders regardless if the document is sent as an electronic form or as a formless document.

The Strategy define all activities needed to build, manage and maintain central basic registers based on ICT. It will define standards for connecting registers with other registers that have a legal basis to use the contents of the central registers. This will ensure not only the accuracy of the registers but also good management of data of the central state bodies. In addition it will be possible to access data needed for the administrative proceedings.

The goal is to enter data just one time and to make it accessible trough ICT solutions to all interested parties that have legal basis. To achieve the one-time data entry principle all personal data like gender, name, birth data and nationality regarding natural persons will be used from one source with a unique record structure and encoding system while the data on birth place will be adjusted with the spatial units register. The same applies to legal persons. Their name and business activity, headquarters (harmonised with the special register), with data on natural persons have to be used from one only source in the state administration.

4.3. Accessibility of Electronic Services

Public information and data as well as services have to be accessible trough all technologically available and acceptable communication channels in any time from any place for all citizens regardless on gender, age, education, social group, financial capacity or any other specificity.

The system of e-Government is visible trough interconnected electronic services open for use trough the Internet and other communication channels. The most important characteristics are accessibility, security and reliability, expandability and interoperability and the technological independence. Every service has to be developed with defined goals that have to be achieved and in accordance with the legal framework that ensures the legal protection of the users of e- Government services.

Standards will be defined that central state bodies will be obliged to use while developing e-services. The whole cycle of e-services, development, introduction and use, has to be measured and indicators have to be set with a goal to analyse the success of the services, the satisfaction of the users as well as the financial efficacy.

4.4. Human Resources

The most important component are people – civil servants that understand on one side the technology and on the other the business rules and requirements. Therefore the success of the electronic administration system depends on the qualification of the civil servants. It is necessary to evaluate the acceptability of existing e-services and to define the needed level of knowledge and skills of the users. Then a unique model for informing the general public on basic rules in using ICT and the Internet in the domain of e- Government has to be prepared. [12]

5. E-CROATIA 2007 PROGRAMME OVERVIEW

The e-Croatia 2007 Programme is divided into two main areas:

- infrastructure, which comprises Broadband, Interoperability, Information security, and HITRONet Network;
- projects, which include eGovernment, e-Justice, e-Business, e-Education and e-Health.

To realize both areas, besides an institutional framework, legislation is required, and in that respect Croatia has already adopted several acts to support the ICT development and its application in private and business sector, such as Financial Agency Act, Electronic Signature Act, e-Commerce Act, Registry Number Act, Act on Personal Data Protection, Telecommunications Act, etc. [11], [12]

6. E-CROATIA 2007 PROGRAMME COMPONENTS

6.1. Broadband

In October 2006, the Government of the Republic of Croatia adopted the *Strategy for the Development of Broadband Internet Access by the year 2008* and *Implementation Action Plan of that Strategy for the year 2007*. Strategy and Action Plan for the development of broadband Internet access create prerequisites for the accelerated development and adoption of this technology. The goal of the Strategy is the reduction of the gap between Croatia and EU countries as well as to encourage an even-paced development through all Croatian regions and population categories.

6.2. Interoperability

In Croatia the *Croatian Standards Institute* is responsible to develop and align its standards to recommendations of international organizations for standardization. So far, some open standards are accepted to be implemented, such as Web Content Accessibility Guidelines 1.0 (implemented on all public administration web sites in 2007) developed and maintained by World Wide Web Consortium (W3C), standards for e-business developed and maintained by the Organization for Advancement of Structured Information Standards (OASIS). Croatia is also started to develop an interoperability framework for several segments of public administration (The Customs Administration, The Tax Administration, Spatial Data national infrastructure) to join national information systems (IS) with ISs of EU countries.

6.3. Information security

The precondition for eGovernment implementation is data interchange, so a security policy should be created and information security standards applied in that environment. Therefore, Croatia adopted *the National Programme for Information Security in the Republic of Croatia* as well as *the Plan for the Implementation of the National Programme for Information Security in the Republic of Croatia for the year 2005*. In the year 2007 the focus is on the strengthening of the appropriate security norms concerning the implementation of information and communication technologies. According to this, two ISO norms were adopted: 'Information technology – Security techniques – Code and practice for information security management

(HRN ISO/IEC 17799:2006)' and 'Information technology – Security techniques – Information security management systems – Requirements (HRN ISO/IEC 27001:2006)'.

6.4. HITRONet Network

HITRONet Network is an information-communication network for state administration interlinked into a unique communication infrastructure. It links various bodies of state administration and provides common Internet access, access by remote users, assistance to users, system security, overseeing and managing the system as well as the subsystem of common network and application services. It is projected that the end of 2007 will have comprised 80% of eGovernment services of the central state administration bodies within HITRONet.

6.5. eGovernment

To enable the development of common electronic services and central access to information resources of the government administration, many projects have been initiated in the area of eGovernment. The main project that incorporates several electronic services is the HITRO.HR service intended for quick communication of citizens and business subjects with the state administration. HITRO.HR is based on the concept of "one-stop-shop" offering citizens and business subjects all the information about the required documentation, as well as the forms and money orders, at the HITRO.HR web site and HITRO.HR counters. HITRO.HR comprises of several services: the establishment of Limited Liability Company, e-Regos (The Central Registry of Insured Persons**)**, e-Tax, e- VAT, e-Pension, e-Craft, e-Cadastre, and e-Corner. Majority of the services mentioned require authorization for accessing the service and authentication of forms by applying smart cards with a digital certificate issued by the Financial Agency. Other projects that are taking place in the bodies of government administration are the e-Registries, the Electoral Register, the Central Database Registry on Personal Data and the Eurovoc Thesaurus.

6.6. e-Justice

e-Justice projects were initiated to introduce information and communication technology into the judicial system. Projects serve to citizens, judges and other judiciaries by enabling access to legal databases and registers. The *e- Portal of the Ministry of Justice* provides information on the activities, reforms and functioning of the judicial system. Some projects that are operational are: Integrated communication system for managing court cases (ICMS), the e-Land-registry Certificates Project, the e-Cadastre Project, the e-Court registry Project, the e-Judicial Practice Database Project, the Judges Web etc.

6.7. e-Business

The goal of e-Business projects is creating favourable conditions for the successful development of electronic business. Legal framework with several laws and decrees already exists, and the Strategy for the development of ebusiness should be adopted by the end of the year 2007. Two e-Business projects are incorporated into the HITRO.HR service (the e-REGOS project and the e- VAT project). Other projects are the e-Crew project and the e-Customs Project. However, despite the predominantly positive attitude towards electronic business, its potentials have not been utilized sufficiently. One study showed that main reasons for not using public services for business were lack of information about the services, habit of

doing business in traditional way, the perception that a company doesn't have sufficient technical conditions for e-business and information security issues.

6.8. e-Education

Owing to the scope of its vision, the e-Education project is probably one of the most ambitious projects among e- Croatia projects. It includes several aspects, such as information-service infrastructure, application of ICT in teaching, and the development of standards and application of certifications in education, necessary for Croatia's further growth as a knowledge country. Some of the projects under e-Education umbrella are the following: the Information System of Elementary and Secondary Education, the Information System of Higher Education Institutions (ISVU), GigaCARNet, Mobile CARNet, free broadband access in student dormitories (StuDOM), the Croatian National Educational Standard (HNOS), the ECDL (European Computer Driving Licence) project, e-Indeks (student's electronic card).

6.9. e-Health

The computerisation and integration of the national health care system is one of the strategic goals of the Government of the Republic of Croatia, which is to provide better service to patients, improve quality of life and decrease the health care system's costs. An integrated computer system for primary health care has already been released in 350 medicine practices and has brought several advantages: doctors have access to patients' medical records, laboratory test results are forwarded to patients' doctors and patients can schedule appointments via Internet. By the end of 2008, all primary health care institutions will have been connected to the integrated computer system. The informatization of the hospital information system (IBIS) as well as that of the system of social welfare is also in progress. The HZZO Portal (the portal of the Croatian Institute for Health Insurance, http://www.hzzo-net.hr/) now offers a possibility to fill in an electronic health insurance form. [11], [12]

7. CONCLUSION

The advent of new communications and information technology is slowly modifying the destiny of democracy. Indeed, what is being witnessed is a change in the very idea of politics, no longer conceived of as the exclusive reserve of professional operators, but rather as a process in which citizens play an active role as parties to an ongoing debate. The web becomes a new metaphor for democracy, guaranteeing heightened availability of information by bringing down barriers of space and time while, at the same time, making it possible to establish collective decision-making procedures by readying new forms of access and spaces in which user can play an active role. But despite the opportunities for the development and consolidation of democratic processes offered by ICT, we still fall far short of being able to define democracy as a policy-making process in which those who govern, considered to be politically equals subject to controls, show themselves to be receptive to the preferences of the governed. The delay, so to speak, can most likely be traced to the fact that, together with a previously unimaginable potential for interaction, the new technology has also spawned a series of concrete risks and limits.

Many developing countries are in the initial phases of adopting electronic government (e-government) programs to improve public services and deliver them as efficiently and conveniently as possible. Experience with a variety of governments throughout the developing world at different stages of implementing e-government programs with citizens (G2C), businesses (G2B), and other entities of government (G2G) suggests that a major reason behind the success or failure of e-government projects is the extent to which, first, the governments address technological infrastructure encouraged by appropriate telecommunications policies; and second, the legal and regulatory instruments required for e-government. Information and communication technology (ICT) infrastructure development is at the heart of successful deployment and sustainability of e-government programs.

REFERENCES

[1] Heeks, R, Bailur, S. '*Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice*', Government Information Quarterly, Elsevier, Vol. 24, pp 243-265, 2007.

[2] Layne, K., & Lee, J. *'Developing fully functional e-government: A four stage model'*, Government Information Quarterly, Vol. 18, pp 122−136, 2001.

[3] Cheng-Yi Wu, R. *'Enterprise integration in e government*', Transforming Government: People, Process and Policy; Vol. 1 Issue: 1, 2007.

[4] Ebrahim, Z., Irani, Z. *'E-government adoption: architecture and barriers'*, Business Process Management Journal; Vol. 11 Issue: 5, 2005.

[5] Lam, W. *'Barriers to e-government integration'*, Journal of Enterprise Information Management; Vol. 18 Issue: 5, 2008.

[6] The Economist, *'A Survey of Government and the Internet'*, June 24, pp 3, 2000.

[7] Hype Cycle Shows E-Government Overcoming Disillusionment; Andrea Di Maio, John Kost – Gartner Group; March 2004, http://www.gartner.com/DisplayDocument?ref=g_searc &id=427810 (accessed 28.01.2011.)

[8] Bertot, J. C., Jaeger, P. T. *'The E-Government paradox: Better customer service doesn't necessarily cost less'*, Government Information Quarterly, Vol. 25, Issue 2, pp 149-154, 2008.

[9] Chang, C. M. L., YiMeng, L., Pan, S. L. *'E government implementation: A macro-analysis of Singapore's e-government initiatives'*, Government Information Quarterly, Vol. 25, Issue 2, pp. 239-255, 2008.

[10] Central State Administrative Office for e-Croatia, Operational Plan for Execution of e-Croatia Program with Overview of Activities in 2007. http://www.e-hrvatska.hr/sdu/hr/ehrv/contentParagraph/011111111111113/document1/Operativni_plan_e Hrvatska_za_2007.pdf

[11] e-Croatia: *Government adopted resolution plan for e-Croatia for 2007*. from http://www.ecroatia.hr/, (accessed 28.01.2011.)

[12] National Competitiveness Council: *Recommendations for increasing ICT and technological competitiveness of Croatia*,http://nvk.multilink.hr/dokumenti/ 119___Preporuke%20za%20povecanje%20ICT%20 konkurentnosti%20Hrvatske.pdf , (accessed 28.01.2011.)

The Data Quality In CRM Systems

Marija Boban[1], Vesna Jevtic[2]
[1] University of Split
[2] University of Novi Sad, Technical faculty Mihajlo Pupin Zrenjanin

Abstract: **This work emphasizes importance of Customer Relationship Management (CRM) data for success in business, especially in marketing, sales and customer services areas. The data quality is essential for deployment of increasingly integrated operational and analytical CRM systems, as well as for meaningful set of customers' metrics maintenance. This work represents goals and strategy, factors and management program of data quality, as well as, data privacy and data security in CRM systems.**

1. INTRODUCTION

CRM data play very important role in marketing, sales and customer services areas of business and like a source of very valuable information are often treated as a business assets. When it comes to customer information, poor data quality can lead to problems that are no less important. Quality of customer data is essential for enterprises and their ability to successfully deploy and maintain increasingly integrated operational and analytical CRM systems, as well as maintain meaningful set of customers' metrics. Therefore, consistency and quality of customer data are critical components for CRM success. The CRM data stored within the CRM system include information contact details as well as other information related to people, organizations and accounts or related to records of orders and communications linked to each of these. Information that are collected by particular person or organization are accurate or true at the moment of recording. These information are prone to become inaccurate during time, and that what holds true for one moment in time may change in future. By integrating customer information from multiple touch points into one, accurate, current record, enterprises can rely on consistent segmentation to drive their treatment of customers and reveal true purchasing behaviors and patterns to forecast future sales accurately. Quality of data is subject by the following key issues:

- How can enterprises create and maintain a database of consistent, high-quality customer information?
- How can consistent, integrated customer interactions be enabled across multiple channels?
- How will enterprises evolve their analytical systems to reveal key customer insights?
- How can customer insights be profitably applied to inbound and outbound customer interactions?
- How to provide privacy for customers' information?

2. GOALS AND STRATEGY OF DATA QUALITY

This should form the introduction to the need for a data quality strategy within organization. The strategic objectives should aim to relate data to the success of organization

and goals. It is often best to start by filling in a grid to identify how the results of data strategy will impact organization, as below. Common goals for CRM system are: improving communications, creating multiple touch points to serve the customer, strong relation with Enterprise Resource Planning (ERP), reduced waste of mailing budget, compliance with data regulations. Achieving the level of integrated, consistent customer data necessary to support CRM initiatives does not come easily. Most large enterprises have an average of five to ten operational sources containing customer data (see Figure 1).

Given a single, integrated view of the customer is the cornerstone of CRM, a common pitfall is to attempt to achieve this view through a single, integrated database. For all but the simplest enterprises, however, such as universal database remains elusive — held out of reach by the intricacies of multiple lines of business, geographic diversity and a legacy application mix. Therefore, most enterprise CRM implementations will be forced to use customer information sourced from multiple data stores. In this environment, identifying the most-appropriate operational sources from which customer data elements can be acquired involves significant analysis. Questions to be resolved include:

- Where data reside?
- What format they have?
- Where are duplicated data?
- Do the overlapping data have incremental value?
- Which data sources are the most reliable?

Enterprises must have three types of data to effectively manage their customer relationships: descriptive, behavioral and contextual. The interaction among these information types must be understood to provide a coherent picture of the customer relationship.

- *Descriptive data* focuses on the customer, which could be an individual, a household, a business or some combination of the three. Demographic, lifestyle and psychographic data fit into this category. Much of these data come from the enterprise's operational systems or from external data providers. It is readily available and, therefore, yields little competitive advantage.
- *Behavioral data* includes details on the transactions and interactions that comprise the relationship between the enterprise and its customers. Acquiring relationship data have proven to be the biggest challenge for many enterprises, because they must strike a balance between collecting too much and too little. These data personalize the customer/enterprise relationship.
- *Contextual data* are the least common type of information for an enterprise to have; however, an enterprise is unlikely to maintain strong customer relationships without an understanding of their context. Because contextual data are both, diverse and unstructured, it is difficult to integrate them with operational customer relationship systems.

3. FACTORS OF DATA QUALITY

Industry analysts point the finger at bad data as one of the top three reasons why CRM projects fail. Because bad data leads to misleading, incomplete, and confusing information, it lowers adoption – another major reason why CRM projects fail.

Accurate information and reports are the life blood of an effective sales force. Without it, management does not have the data to make good decisions, sales reps do not have the tools to turn leads into customers, and the company will find it difficult to reconcile CRM data with data in other systems. The results are lost opportunities and revenue, frustrated users and customers, and a lack of user adoption.

To ensure consistent high data quality, the users need to be trained, and companies need to create and implement a data quality process, and use available technologies to automate the process whenever possible. A six-step approach that is working for many companies includes as follows.

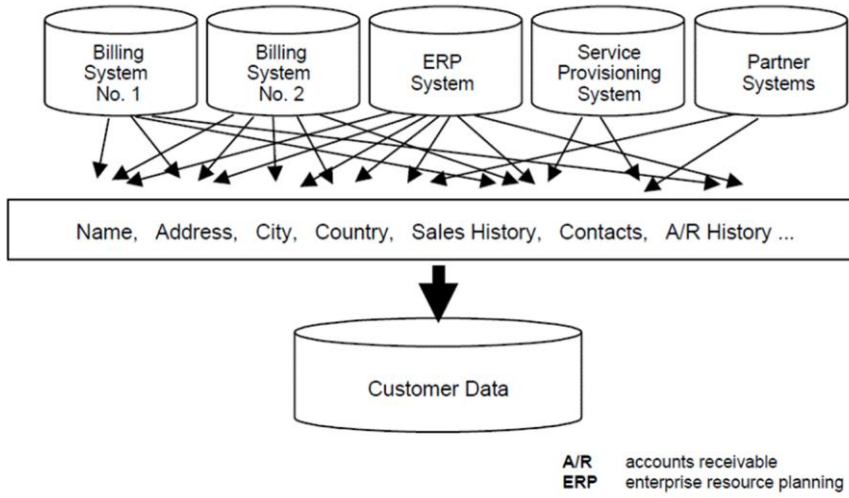

Figure 1. The Quality and Consistency Challenge in a Heterogeneous Data World [2]

3.1. Data Profile and Context

Data profiling is all about understanding data. Companies should know where their data come from: spreadsheets, backend systems, or sticky notes all over reps' desks. An inventory of the data should be taken that includes the following information:

- The data sources should be listed as well as the names of the fields in which data are stored;
- Any potential problems with the data should be noted. Companies should ask questions such as: Do we have automated quality checks before new record can be saved? Are all fields mapped correctly?

Context defines the type of data and how the data are used. Ultimately, the context of your data determines the necessary types of cleansing algorithms and functions in order to raise the level of quality.

3.2. Data Control

Data control includes achieving data accuracy and ensuring the right users have access to the right information, which also means blocking access, as needed. To control data, they should first be cleaned by removing duplicates and errors. Then, processes should be set up and technologies used to keep data clean. Some guidelines can be outlined as follows:

- Use automated routines or tools to clean your data;
- Prioritize your data cleanup effort. First fix data which are highly visibly and frequently used, such as addresses and emails. Fix business-specific information next, such as opportunity types and stages. Finally, remove any duplicated fields; for example, do not repeat account information in the contact object;
- Get your users to fix their data. You can alert users when data records are incomplete or do not conform to data-quality standards;
- Use exception reports and data-quality dashboards to remind users when their accounts and contacts are incorrect or incomplete. Scheduling a dashboard refresh and sending that information to managers is a great way to encourage compliance;
- Review your profiles and role hierarchy and make sure the hierarchy, teams, and groups are kept up to date;
- Meet frequently with management to keep up with organizational changes;
- Define the Create, Read, Update, and Delete (CRUD) rights for each profile to ensure users can work with data, as needed.

3.3. Integrate data and storage

Most organizations have data in more than one system. Whenever there are duplicated data in several systems, it is easy for information to get out of synch. One way to avoid this problem is to integrate systems so that updates in one system automatically update the others, resulting in a single source of "truth" and making it easier for end users to access information. When designing the integration, company should evaluate their business applications to determine which one will serve as system of record for the synchronization process. The system of record (or master) can be a different system for different business processes.

Every data quality strategy must consider where data physically reside. Considering storage as a data quality factor ensures the physical storage medium is included in the overall strategy. System architecture issues – such as whether data is distributed or centralized, homogenous or heterogeneous – are important. If the data reside in an enterprise application, the type of application (CRM, ERP, and so on), vendor, and platform will dictate connectivity options to the data.

3.4. Augment data

To make a CRM system even more valuable, the data should be augmented with information that will give the salespeople and managers an edge. For example, companies should make the most of internal market intelligence, such as purchasing patterns or competitive analyses. To understand what kind of data are valuable, they should survey the sales and marketing users to see what they want the most. Also, an internal information about customers' behavior and buying patterns should be collected and then evaluated, if it is helpful.

3.5. Monitor data

Achieving high-quality data is not a one-shot effort, but requires ongoing vigilance. Establishing policies, processes, and tools for monitoring data is crucial to maintaining data quality. A centralized process for mass data loads and data-cleansing projects should be defined.

3.6. Assign ownership, train users and commit to a data-quality process

Users need to know the importance of data integrity and how to do their part in any data-quality initiative. During users training, companies should show them how data quality directly affects their work. It's also a good idea to assign ultimate responsibility for each region's data to a super user, geographic lead, or other business owner.

4. CUSTOMER DATA QUALITY MANAGEMENT PROGRAM

Best practitioners of Customer Data Quality (CDQ) management program combine vision, technology, culture and business rules to cultivate more intimate and relevant relationships with their customers.

Establishing a successful CDQ program requires more than just applying the right data quality software to the problem. It is imperative that organizations first identify and resolve any underlying business or cultural issues impeding customer information management. Prioritizing objectives to be achieved with the CDQ program is the next order of business.

When those fundamental matters have been adequately addressed, an organization can confidently embark upon the implementation of a CDQ program, which is comprised of the following components:

- Discovery & Analysis
- Data Conversion & Cleansing
- Data Quality Maintenance
- CDQ in Enterprise-wide CRM
- Legality & privacy

The Web presents a special challenge to data quality maintenance because the responsibility for data input lies more with the e-customer and less with the organization. Organizations that value their customer information will place a data quality filter at all customer interaction touch points, including the Web. This filter is the organization's defense against customer data corruption. After the CRM system is populated with cleansed and linked data, organizations will want to focus their preventative data quality measures on the front-line, rather than on more costly and time-intensive back-office clean-up. Just as business rules were critical to the conversion, data quality filters must be flexible and robust enough to support the organization's established business rules. This consistency will ensure the organization's data quality conversion efforts are maintained going forward.

5. DATA PRIVACY IN CRM SYSTEMS

The right to data privacy is heavily regulated and rigidly enforced in Europe. The European Court of Human Rights has given this article a very broad interpretation in its jurisprudence. According to the Court's case law the collection of information by officials of the state about an individual without his consent always falls within the scope or article eight. Thus, gathering information for the official census, recording fingerprints and photographs in a police register, collecting medical data or details of personal expenditures and implementing a system of personal identification have been judged to raise data privacy issues.

Any state interference with a person's privacy is only acceptable for the Court if three conditions are fulfilled:

1. the interference is in accordance with the law,
2. pursues a legitimate goal and
3. is necessary in a democratic society.

Private companies are engaged in threatening activities, especially since the automated processing of data in CRM systems became widespread. As the entire member states of the European Union are also signatories of the European Convention on Human Rights and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, the European Commission was concerned that diverging data protection legislation would emerge and impede the free flow of data within the EU zone.

Therefore the European Commission decided to harmonize data protection regulation and proposed the Directive on the protection of personal data. In full name, Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data contains a number of key principles which must be complied with. Anyone processing personal data must comply with the eight enforceable principles of good practice. The personal data must be:

1. Fairly and lawfully processed;
2. Processed for limited purposes;
3. Adequate, relevant and not excessive;
4. Accurate;
5. Not kept longer than necessary;
6. Processed in accordance with the data subject's rights;
7. Secure;
8. Not transferred to countries without adequate protection

The directive regulates the processing of personal data, regardless if the processing is automated or not. Personal data are defined as "any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;" (art. 2 a). This definition is meant to be very broad. Data are "personal data" when someone is able to link the information to a person, even if the person holding the data cannot make this link. Some examples of "personal data": address, credit card number, bank statements, criminal record.

The notion processing means "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction;" (art. 2 b)

The responsibility for compliance rests on the shoulders of the "controller", meaning the natural or artificial person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; (art. 2 d)

The data protection rules are applicable not only when the controller is established within the EU, but whenever the controller uses equipment situated within the EU in order to process data. (art. 4) Controllers from outside the EU, processing data in the EU, will have to follow data protection regulation. In principle, any on line shop trading with EU citizens will process some

personal data and is using equipment in the EU to process the data (the customer's computer). As a consequence, the website operator would have to comply with the European data protection rules. The directive was written before the breakthrough of the Internet, and to date there is little jurisprudence on this subject. That is why all of the EU countries have developed their own legislative in data protection on the traque of these directives.[4]

In contrast European state-to-state approach, Canada has, through Personal Information Protection and Electronic Documents Act - PIPEDA[5], chosen an organization-to-organization approach that is not based on the concept of adequacy. PIPEDA does not prohibit organizations in Canada from transferring personal information to an organization in another jurisdiction for processing. Organizations not governed by PIPEDA for commercial activities within a province need to be aware that PIPEDA applies to transborder transfers. However, under PIPEDA, organizations are held accountable for the protection of personal information transfers under each individual outsourcing arrangement. The OPC can investigate complaints and audit the personal information handling practices of organizations.

6. INFORMATION SECURITY MANAGEMENT SYSTEM AND THE CREATION OF SECURITY POLICIES IN CRM SYSTEMS

Evidently, in modern digital economy, information is the lifeblood of business and organisations are increasingly dependent on the use of information systems and networks to process information. Computer 'literacy' is now widespread making systems ever more open to abuse, whether deliberate or accidental. Consequently, businesses are increasingly at risk through use of the very tool introduced to increase efficiency, i.e. information technology (IT). Managers must therefore address these risks where they would affect their systems and the information used on them in terms of:

- confidentiality
- integrity
- availability.

as it is shown on the Figure 2[10].

Security requirements are identified by a methodical assessment of security risks. Expenditure on controls needs to be balanced against the business harm likely to result from security failures. The results of the risk assessment will help to guide and determine the appropriate management action and priorities for managing information security risks, and for implementing controls selected to protect against these risks.

[4]For example, in Croatia, the protection of personal data is regulated since 2003. by the Law on protection of personal data, as well as the Law of the right to data privacy. On the traque of European directives, as the crown of this process, information security is regulated and enforced since 2007. by the Law on information security. In Serbia the Law on protection of personal data is also issued in the year 2010. also on the traque of the European Directives.

[5] The Office of the Privacy Commissioner of Canada (OPC) has developed these guidelines to explain how the Personal Information Protection and Electronic Documents Act (PIPEDA) applies to transfers of personal information to a third party, including a third party operating outside of Canada, for processing. The guidelines do not cover transfers of personal information for processing by federal, provincial or territorial public sector entities. Nor do these guidelines deal with any specific rules governing transfers for processing that may be found in provincial private sector privacy laws. For more information look at ◦Federal Privacy Commissioner – Guidelines for Processing Personal Data Across Borders,
http://www.priv.gc.ca/information/guide/2009/gl_dab_090127_e.pdf , 08. 02. 2011.

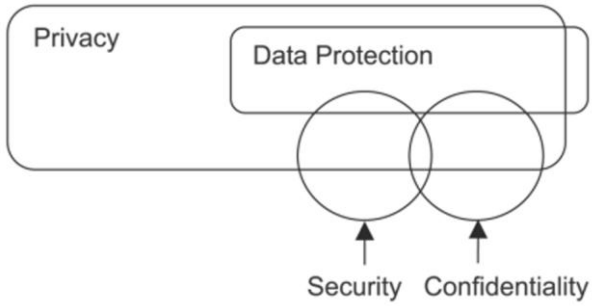

Figure 2. Basics of Information Security Management System

Risk assessment should be repeated periodically to address any changes that might influence the risk assessment results. If managing an organization's information security risks quality, the most important is to choose the best methodology. The model of Risk management is presented at the Figure 3 [11].

In the increasingly interconnected business environment, information is an asset that, like other important business assets, is essential to an organization's business and consequently needs to be suitably protected. As a result of this increasing interconnectivity, information is now exposed to a growing number and a wider variety of threats and vulnerabilities. Information can exist in many forms: it can be printed or written on paper, stored electronically, transmitted by post or by using electronic means, shown on films, or spoken in conversation. Whatever form the information takes, or means by which it is shared or stored, it should always be appropriately protected. It is important to stress that information security is the protection of information from a wide range of threats in order to ensure business continuity, minimize business risk, and maximize return on investments and business opportunities.

Figure 3. Model of Risk Management Method

In organisation environment, information security is achieved by implementing a suitable set of controls, including policies, processes, procedures, organizational structures and software and hardware functions. These controls need to be established, implemented, monitored, reviewed and improved, where necessary, to ensure that the specific security and business objectives of the organization are met. This should be done in conjunction with other business management processes.

Information security policy[6] in accordance with the application of ISO 27001:2005[7] 4.2.1 b must be defined as follows:

"The organization must do the following:"

- Define the scope and limits of ISMS in terms of business features, organization, its location, assets and technology including details and justification of any exclusions from the scope
- Define the ISMS policy in terms of features, business, organization, its location, assets and technology that:
    - includes a framework for setting goals and establishing overall sense of direction and principles for acting in relation to information security,
    - takes into account the business and legal or regulatory requirements and contractual obligations related to security,
    - it is consistent with the organizational context of strategic risk management within which will be made the establishment and maintenance of ISMS;
    - establishes the criteria by which risks will be evaluated (see 4.2.1c)), and
    - is approved by management.

And according to the requirements of Annex A:

A.5.1 Information security policy – the goal: "Provide guidance and support from management for information security in accordance with business requirements and applicable laws and regulations.

A.5.1.1 Information security policy document must be approved by management, must be published and all employees and relevant external parties must be familiar with it.

A.5.1.2 "Information security policy must be viewed at planned intervals or if significant changes to ensure the lasting relevance and effectiveness.

[6] General security policy is binding for the entire organization. With defining of a specific security policies, general security policy is specified and descends to the lower operational levels. Specific security policies regulate individual aspects of business processes that wish to be maintained under control of ISMS. The area for which are defined the specific security policy must be within the scope of the ISMS and the policy must be consistent with the general policy of the ISMS.

[7] ISO/IEC 27001:2005 - Information technology - Security techniques -- Information security management systems - Requirements

7. CONCLUSION

As it is shown, consistency and quality of CRM data are crucial for business success. In order to prevent incorectness of data, enterprises develop strategy and goals of data quality. Enterprises use descriptive, behavioral and contextual data types to effectively manage their customer relationships.

Factors which ensure consistent high data quality are represented by six-step approach including: data profile and context, data control, data integration and storage, data augmentation, data monitoring, assigning ownership, users training and commitment to data-quality process.

Important issues of CRM systems are data privacy and information security policies. Data privacy is very delicate issue and every EU country has developed its own legislative in data protection. Information security in organisation environment is achieved by implementing a suitable set of controls, including policies, processes, procedures, organizational structures and software and hardware functions. These controls are part of information security management system, which ensures accomplishment of specific security and business objectives of the organization.

ACKNOWLEDGMENT

This research is supported by Ministry of Science and Technological Development, Republic of Serbia, under the project number TR32044 “The development of software tools for business process analysis and improvement”.

REFERENCES

[1] George Fratian, Planning your CRM SAP Implementation, Galileo Press, 2008.

[2] CRM Data Strategies: The Critical Role of Quality Customer Information, Gartner, Executive Report Series, 2003.

[3] Russ Lombardo, CRM for the Common Man, PEAK Sales Consulting, LV, USA, 2006.

[4] Jill Dyche, CRM Handbook A Business Guide For CRM Customers, Addison Wesley IT Series, 2004.

[5] M.Ivković and B.Radenković, Internet i savremeno poslovanje, Tehnički fakultet Mihajlo Pupin Zrenjanin, 1998

[6] http://knol.google.com/k/crm-data

[7] http://trilliumsoftware.com/home/solutions/enterprise

[8] http://www.priv.gc.ca/information/guide/2009/gl_dab_090127_e.pdf

[9] www.cihi.ca/partship/conoct99/peladeau.pdf

[10] Wilhen Associates Limited/ Project Frameworks Solutions Limited October 2002.

[11] Miodrag Ivkovic, Dusanka Milanov, Branko Markoski, Organizational readiness for CRM Software Implementation, E-Society Journal Research and Apllications, Volume 1 Number 1, pp 31-37, 2010.

Elearning Scenario in Community of Practice Environment

Mr Željko Džunić[1], prof. dr Leonid Stoimenov[2]
[1]Univerzitet u Nišu,
[2]Elektronski fakultet u Nišu

***Abstract*: The phenomenon of social learning is present in human society since ancient times and scientists in that fact see the main reasons for the development of human society. Despite this, eLearning systems do not take advantage of such learning, but are based mainly on the distribution of learning materials and evaluation of knowledge. On the other hand there is undeniable growing popularity of online social communities as well as the application of social software within them. The concept of Community of practice represents an excellent bridge between the education area and social communities. The implementation of eLearning systems that are based on Community of practice is recognized as significant potential that can lead to effective social learning. For this reason, the project Scope is running, project of development Student Community of practice, in order to investigate the effects of social learning and the relationship between technology of social software and eLearning. Scope should provide a learning environment for students and other members of the community where free exchange of knowledge is and providing help between members is stimulated. Through development of Community of practice we expect very significant effects in other areas, especially in the field of e-lance economy and employment of staff. The first results of research conducted within the project Scope are presented in this paper.**

***Keywords*: Community of practice, eLearning, Online social community, Social Networking**

1. INTRODUCTION

ELearning is a concept describing any type of learning environment that is computer enhanced. As a concept of learning that has been available for a longer period, its advantages and disadvantages are quite enough explored [1]. ELearning systems are very often used in educational institutions as a primer way of education, not only as support of existing classic process of learning. There are a lot of tools which can enable the implementation of eLearning environment, commercial or open source, such as Moodle, Blackboard, etc. Nevertheless, the following question is still actual: is it possible to implement the complete process of education using these systems and tools? The analysis of already implemented eLearning systems and services available at University of Nis, has shown that such systems are used only for learning objects distribution and knowledge evaluation but not for students collaboration using Weblog, forums or other services for collaboration. In that way one very important part of learning, shared learning - learning through practice or experience is missing.

On the other hand, there is a theory according to which the social and cultural factors are most influential in the development process of an individual [2]. People are constantly learning from other people who are in their environment. A similar theory, called 'Situated Cognition' was set by Lave and Wenger [3], in which they presented the concept of Community of practice. Learning, as outlined in this Wenger-vision Community of practice [4] is achieved mainly through social activities. In such an environment the student acquires his knowledge

and competence through connectivity and belonging within communities where he can realize his interests and share knowledge with others. In addition to these communities, there is an interesting community in the area of collaborative learning, identified by Berlanga et all. [5]. They define the so-called ad hoc community as "a community that exists in order to meet the individual requirements in a limited period of time". It should be particularly noted that in these communities, sharing knowledge is not imposed or under pressure, but occurs spontaneously, whereby the application of technology can help speed up the process and the emergence of community itself. Conditions that each community should fulfill so that the sharing of knowledge is enabled are as follows:

- Community has to have a clear goal;
- Community has to have members with different levels of knowledge in different domains;
- Community has to track all member's activity and to measure performances based on community trust upon each member;

Human life as well as social communities have got a new dimension with technology development. Digital community or social community available on Web, are today more and more in focus. Regarding this, Tim Berners Lee notes that the Web is more a social creation than a technical one, and it is designed for a social effect, to help people work together. According to Tim Berners Lee, the ultimate goal of the Web is to support and improve our Web-like existence in the world [6].

Based on specified facts about not well researched social character of existing eLearning systems as well as the fact that education process is placed in digital social community, an online community, named „Scope" (Student Community of practice) is developed.

By developing Community of practice for eLearning system we wish to research efects of social connections on education process as well as to find out if the idea of free knowledge sharing is sustainable in university population despite significant differences with existing economical model. Competency will not be based on protecting personal knowledge any more, than quite opposite: trough knowledge sharing and open online collaboration. Community members, who own and share knowledge with other members will have higher rating and thereby higher competency comparing with others who keep their knowledge. The second aspect of our research is finding possibility to use Community of practice and accumulated knowledge about personal knowledge for other kinds of connection between people, firstly on job market and human resources business.

In this paper we put forward the results of the first phase of research, which consists of developing practical community as well as defining functional design of the environment of online social community that can respond to the above requirements. The paper is organized as follows. The first part presents the theoretical basis of Communities of Practice on which developing of Scope - Community of practice is based. In the second part of pape is presented the system architecture and a description of technologies and services implemented within the system. In the context of the architecture, the characteristics of electronic portfolios are presented as well as a description of the role of emerging community members. At the end of the paper, we presented the existing research results of practical application of the concept of online communities and social software in the field of education, in the form of a conclusion. Also we presented the future steps in the development of Scope Community of practice, particularly the implementation of trust on which the whole concept is based and which is a key prerequisite for the use of such systems within the "e-lance" economy.

2. COMMUNITY OF PRACTICE

The term Community of practice has only been recently in use although the phenomenon is present from the foundation of mankind and their need to learn in the community. It turns out that this concept is a useful perspective in the field of knowledge and learning. This part of the paper aims to explore what Community of practice (CoP) is, its theoretical basis and why researchers and experts from various fields and in different contexts see this community as useful in the process of learning and sharing knowledge. In addition to this part of the work, the relationship between online communities and Community of practice are analyzed as well as the possibility that online communities represent a functional environment for the realization of communities of practice.

According to Wenger and his theory of learning within the community, learning is a social process, so it can be seen through the involvement and contribution of each individual to community, which he belongs to. The basic assumption of the theory of Community of practice is, "the engagement and involvement of individuals within the community is the basic process through which we learn and become what we are" [3].

Communities of Practice are formed by people who are involved in the process of collective learning in the shared domains of human behavior, such as, for example, a group of artists seeking new forms of expression, a group of engineers working to solve similar problems, a group of students who prepare a certain exam, a network of surgeons that explore new techniques, a group of inexperienced managers who help each other, etc.. Therefore, Community of practice is a group of people who have common interest, problem or passion in a particular domain and who want to gain knowledge in the appropriate area or expand existing knowledge to specialize in a particular area [4]. Participation in community is voluntary and open to all who are interested in a given area or topic. Community development is based on mutual interest and interaction of participants, which means that it is impossible to create a community without the active participation of people [7].

On the other hand, as is proved by Wenger [7], not every social community or group is a Community of practice, because otherwise this concept would lose its meaning. There are certain characteristics that must be identified, so that a community could be classified as Community of practice.

Community can be considered as Community of practice if it is formed around the corresponding domain, has an interactive community and owns divisible knowledge and experience, while basic features can be explained as follows (modified from [8]):

Domain – Community of practice is not a club or a network of friends. It has an identity defined by divisible domain that represents the interests of all members. Membership in the community implies a commitment to the area and therefore a certain level of competence in a given area that differs members of the community from other people. The aim is to improve community knowledge of the whole community through the exchange of knowledge in a defined area.

Community - to fulfill their interests in the domain, members of the community join together and through their activities and discussions share information and help each other. They build relationships that enable them to learn from each other. Web site is not a Community of practice itself. Also, the same job or the same position does not make

Community of practice, unless members of the community have the opportunity to learn from each other through interaction and thereby advance.

Practice - Community of practice is not only a Community of Interests or people who, for example prefer a certain kind of films. Community members are practical experts. They develop shared resources such as experience, ways to solve specific problems that often occur, in a brief and divisible way. This requires time and sustainable interaction. Exchange of shared experiences should be more or less self-conscious.

The combination of these three elements form a Community of practice.

3. VIRTUAL COMMUNITY OF PRACTICE (VCOP)

Prior to information technology, the term Community of practice related to a group of connected people who usually lived in the same area [9]. With the development of online tools that allow people to exchange ideas in a virtual environment, the concept of face to face community is enriched and expanded with virtual interactions. These online communities could include people who know each other and share the same living space but at the same time are able to communicate on an international level with anonymous participants. Such communities are called online or virtual Communities of Practice and include the online platform in which people share their knowledge and interests in a virtual basis or through online communication in the appropriate domain. In this case, communication and sharing of knowledge is supported by software tools, which are often called social software. These tools enable cooperation and collaboration without time and geographical constraints, which is considered a key factor for Learning on Demand and Just in Time learning as the characteristics of Communities of Practice.

As already emphasized in the paper, members of online Community of practice must not know each other but their activities still adhere to the basic concept of Community of practice defined by Wenger [7]. Considering the theory of social learning and characteristics of the social software it is obvious that there are great similarities between these two concepts. First of all, both concepts are directed towards people and require their active participation and engagement on which their success directly depends on. In addition, both support the concept of shared interests of the people. It should be emphasized that the success of any technology depends on whether it is supported in the respective community as it is spotted in Wenger [10]. Social software is essentially entirely oriented towards the community and therefore has broad support within it.

From the standpoint of technological requirements, virtual Community of practice can be realized through the appropriate Web site with implemented services for collaboration and administration of members, shared work space, shared document repository, search and the creation and management of communities. Considering these requirements, the conclusion is that by social software, almost all the listed requirements for implementation of VCoP could be realized.

Because of these similarities it is interesting to study whether the virtual communities based on social software solution can be considered as Community of practice and support the learning process organized within it. In the literature there are different interpretations of these concepts. Given that the majority of research is still in the description phase, it is of great

importance to precisely define these concepts and relations between them. In relation to the first part of the question, dominant idea is that social software is generally treated as an additional channel for communication and not as a Community of practice itself. A similar conclusion was reached by Johnson [9] who believes that virtual communities can only represent practical support to communities instead of being Community of practice itself, which implies that the technology used in virtual communities is only the means for their implementation. Regarding the second part of question, undeniable fact is that social software is still supporting learning in practice.

The general conclusion about online communities and social software on the one hand and online Community of practice on the other is the following:

- Technologies of virtual communities are developed from existing tools through their active use by community members
- While there are tools that support the Community of practice, there is no technology with which it is possible to entirely realize Community of practice
- The main potential of the Community of practice are not or should not be tools to support them but the people who belong to the virtual community.
- Efficient technology is only part of development process of successful online communities [11].

It can be assumed that, within the Community of practice, emphasis is on connecting people and their active contribution to network development and not on technology.

4. SCOPE OVERVIEW

Project of student Community of practice (SCoPe) has been developed in order to support eLearning at the University of Nis. Community on the one hand supports formal learning through a link with the existing systems of eLearning (in this case it is about Moodle software platform for eLearning) while on the other hand, the emphasis is on informal learning through the development of online communities and access to learning material that can be imported from other media community (Youtube, Wikipedia, etc.). Development of such a designed community allows the exchange of knowledge and the accumulation of knowledge in certain areas, which following the applied social network analysis, represents a huge potential for the entire university, especially in the connecting staff and companies from around the world that have a need for them.

With Scope, students have free access to knowledge and can realize simple communication with other students, professors or institutions within the particular domain. Through Community of practice they share ideas, cooperate with each other and are engaged in one or more groups. In addition to learning within the community, students have at their disposal different resources for learning, which can change in a way that suits them and thus promote the knowledge of the whole community.

The main features of Scope community are the following:

- Activities of the community are directed to knowledge sharing and acquiring;
- Classical relation between teacher and student is changed with community members collaboration no metter which is his role in education process (student also could be the source of knowledge or to transfer his experience in one domain and consume knowledge in other);

- Activities in community are focused to problem solution (Problem Based Learning - PBL) that is learning is happening during the process of solving the problem which student found.

All activities in the Scope, such as sending responses to student questions, writing comments on blogs, initiating research, accumulate in the shared and evolutionary online portfolio. Students who freely share the knowledge and wish to demonstrate and improve their skills, can meet such requirements through their activities in one or more communities. The specific knowledge could be gained on the Web through access to available educational resources such as Wiki, Weblog, learning together, sharing and exchange of knowledge in the community.

The choice of technologies for making Scope Community of practice is managed by the following requirements: ease of use, flexibility, the ability to adjust to the demands of users and a simple and effective communication between users of the system. We considered several solutions, from the development of completely new software product to the implementation of the system using ready-made open source solutions. After the analysis of available products in the field of course management system as Sakai and content management systems such as Joomla and TikiWiki, we decided to use Elgg platform for the development of online communities. The choice is entirely logical if one takes into account that the objective of the whole system is support of informal learning through learning in the community.

Using Elgg platform we have developed a community that is oriented towards the solution of concrete problems. Students are organized into small groups specialized in specific domains within which they exchange the materials, ideas and experiences or ask specific questions and present problems that are encountered in their work. They use mail, chat and other communication tools such as wiki, blog and forum for the solution of concrete problems. In addition, Scope platform is completely open ended to allow communities to connect with existing systems for electronic learning, such as Moodle which was one of the conditions for implementation. The main feature of Scope Community of practice is full orientation towards the users and support of all the requirements of shared learning through the development of a personal environment for learning. This environment can be defined as a system or concept that helps students to control and manage their own learning process. This way, students can:

- define learning objectives;
- manage the learning process and necessary facilities;
- communicate with other participants in the learning process,

and thus achieve the planned objectives.

Personal learning environment is not a type of software but a new approach to using technology for learning [12] or a collection of free, distributed, Web-based tools, mainly concentrated around the blogs, which are interconnected and which group content using RSS feeds and simple HTML scripts [13]. The four basic characteristics could be distinguish in almost all definitions of a personal learning environment:

- Individual control tools and content;
- Content aggregation and collection;
- Service integration;
- No spatial and temporal constraints.

The following services are used in the personal learning environment of Scope Community of practice (Figure 1):

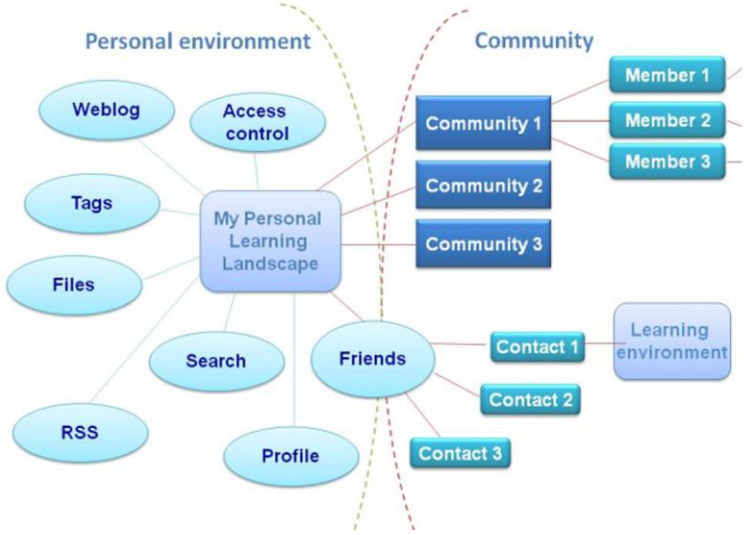

Figure 1. Personal learning environment in the Scope Community of practice

Profile (User profile) is the basis for the creation of user digital identity and e-portfolio which will be discussed in more detail later in paper. In addition, the user profile is the basis for the realization of economic vision of Scope Community of practice since it is based on shared and evolutionary user profiles. Profile reflects the status and needs of each individual within the domain where he belongs. It is essential that every member can easily find the appropriate domain within which he wants to progress and achieve full functionality. User profile within the Scope Community of practice is realized using Foaf standards for the description of user profiles and connections with other people based on the RDF.

Blog (Weblog) is a form of a Web page that contains articles similar to posts in chronological order. Blog can be related to an individual or small group of authors who are in some of the online communities, formed around a particular domain.

File repository allows storage of different types of files that can be support in the exchange of knowledge and ideas.

Tagging is a process in which users assign tags to objects in order to share content with other members of the community. The process is also known as folksonomy and is directly related to the social bookmarking, method that allows storing, organizing, search and management of meta data (tags) by which users mark a Web page. Search tags in Scope Community of practice make an excellent mechanism to find community members or groups with similar interests as well as monitoring their activities in the long run. To achieve the planned goal of learning within the Scope Community of practice it is necessary that the system is developed using technology of Semantic Web. One of the planned mechanisms is tagging objects for learning and user communities with ontological-based tags. This way, it would be possible to connect formal knowledge represented by domain ontology with informal knowledge that is gained through the process of social labeling.

RSS (RDF Site Syndication) is a technology that allows users to search a list of changes in blogs, tags, communities and other services from Community of practice. A user who is logged

on to the appropriate RSS feed, gets information about the names of new items, their short summary and the URL of any changes.

Access Control allows different levels of access to content for individuals or groups within the Community of practice. In Scope Community of practice, users manage their own security because there are mechanisms for defining access to each level of community organisation. It is possible to create an unlimited number of communities which can define public access to their contents such as documents, discussions and other activities, or to keep them away from the public eye.

Online community enables students to connect to the exchange of knowledge around common interests and domains.

Search is one of the key mechanisms of Scope Community of practice that allows users to search resources whether it is a learning material or community members or groups that are formed around certain domain.

A special feature, which is enabled using Elgg, is the possibility of environment adjustments according to the needs of each individual or group.

5. SCOPE MEMBERS ROLE

Implementation of environment for the Community of practice represents only the beginning and prerequisite of its establishment. Community of practice begins its life only when people connect through the domain community and begin to actively work within them. Survival and maintenance of community is possible only if all members have clear and achievable interests and if they are focused on achieving common goals. Also, for the success of community of crucial importance is the manner in which members experience the community, how much time they spent in community activities and whether the community is able to evolve over time.

Members of the community can roughly be classified into those that provide knowledge (Knowledge Providers) and those who use or consume knowledge (Knowledge Consumers). One person can have different roles in different domains: for example to be knowledge provider within one domain while a consumer in the second. On the basis of activities that members achieve within a certain domain, we can classify the members of Community of practice as active, casual, peripheral and external subjects

On the other hand, according to one's knowledge in certain domains of community, members can be observers, beginners, experts, leaders, coordinators and freelancers. The leaders direct other members to be focused on short-and long-term goals and at the same time, if necessary, raise the energy of community organizing new events and activities. Coordinators are tasked to assist members of the community, advise them, connect with other members of the community and constantly stimulate and encourage interaction between members. Experts are experts in the appropriate field. External entities are usually agents of the company, HR managers who are in charge of finding adequate personnel according to the customer requirements, who would be involved in specific projects or activities.

The basis of the community are of course all of its members, regardless of their roles, because without them, there is no community. Starting as the newcomers on the outskirts of the community, members of the community evolve to average users who thus gravitate toward the

center of the community and that at some point on the basis of their activities become experts in the appropriate domain and thus gain a central role. Process of the student evolution in the community is a learning process in which students become more aware of the facts about the community.

6. CONCLUSION

The paper presents the first results of research within the Scope project, developing Community of practice of students. In this phase, we implemented personal learning environment that is based on the Elgg software, and which aims to improve the process of social learning among students and other users of the system. Different services are available to students in the Scope Community of practice such as Weblog, chat, forum, online community that can be formed around different domain, the digital identity of each user, a mechanism for searching for learning materials as well as other members of the community based on social tagging. The system was developed so that it can identify, search and recommend relevant materials as well as individuals and communities that exist within the system that can help in solving of certain learning tasks, based on user profiles and the specific task of learning. It should be emphasized that the focus of developed practical community is oriented towards customers and the most important effects of the system are expect from improvement of social learning. However, Scope Community of practice cannot be considered as complete environment for eLearning since it doesn't contain a system for course management. Considering it's openness, connection to other standard course management systems in the area of e-learning is not a problem, as demonstrated by the Scope Community of practice integration with existing Moodle course management systems. The conclusion is that Scope is the missing link to complete eLearning process, since in the existing systems, the concept of social learning does not apply to a sufficient extent. It is important to say that Scope has just begun its development and the positive effects of the system implementation should yet be expected.

The next phase in the implementation of Scope Community of practice means expansion and provision of community sustainability, as a very important step in our opinion. First thing is the implementation of confidence in the Scope Community of practice which means testing the key factors that customers consider when making decisions whom to believe in the learning process? User profile will be extended with the identified attributes and the trust will specifically refer to each domain in which the user is active.

REFERENCES

[1] Anderson, T., (2008), *The Theory and Practice of Online Learning*, Published by AU Press, Athabasca University

[2] Vygotsk, L., (1986), *Thought and Language. Rev. & ed. A. Kozulin*, Cambridge, Mass: Massachusetts Institute of Technology Press

[3] Lave, J., Wenger, E., (1991), *Situated learning: Legitimate peripheral participation*, Cambridge: Cambridge University Press.

[4] Wenger, E., McDermott, R.E., & Snyder, W.M., (2002), *Cultivating communities of practice*, Boston, Harvard Business School Press

[5] Berlanga, A., Sloep, P.B., Kester, L., Brouns, F., Van Rosmalen, P., Koper, R., (2008), *Ad hoc transient communities: towards fostering knowledge sharing in Learning Networks*, International Journal of Learning Technology, 3(4), 443-458.

[6] Berners Lee, T., (2000), *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*, HarperCollins Publishers Inc.

[7] Wenger, E., (1998), *Communities of Practice: Learning, Meaning and Identity*, Cambridge University Press, New York, USA

[8] Wenger, E., (2006), *Communities of Practice: a brief introduction*, By Etienne Wenger,

[9] Johnson, C. M. (2001), *A survey of current research on online communities of practice*, Internet and Higher Education, 4:45–60,

[10] Wenger, E., White, N., Smith, J. D., and Rowe, K. (2005), *Technology for communities*, CEFRIO Book Chapter, pages 1–15

[11] Garber, D., (2004), *Growing Virtual Communities*, Athabasca University, Centre for Distance Education, Online Software Evaluation Report

[12] Attwell, G., (2007), *The Personal Learning Environments - the future of eLearning?*, eLearning Papers, vol. 2 no. 1. ISSN 1887-1542.

[13] Fitzgerald, S. (2006), Creating your Personal Learning Environment, A workshop presented for the August 3rd LearnScope Workshop. Australian Technology Park, Redfern,

The Implementation of the IMS LD E-Course Generator

Goran Savić, Milan Segedinac, Zora Konjović
Universisty of Novi Sad, Faculty of Technical Sciences

***Abstract*: This paper presents the implementation of an e-course generator. Courses are generated based on three components: learning goals, learning objects and instructional design. The system's architecture is extensible -it is possible to extend it in order to generate courses in different output format. The result is an automatically generated IMS LD compliant course. The system is implemented in Java programming language.**

1. INTRODUCTION

Creating an e-course may be a very time consuming job if it is done completely manually. Also, there is a need to periodically change the created course. A part of this job may be automated using course generator systems. These systems automatically create an e-learning course using different input parameters – learning objects, domain knowledge, student model, pedagogical knowledge, etc. The paper [1] presents a system DCG which generates a sequence of learning activities using domain knowledge, instructional plan, learning material and student's pre-knowledge as the input parameters. This sequence is personalized for each student. Domain knowledge is represented using a concept map. A learning environment for teaching mathematics is presented in [2]. For each student, the system generates a personalized course. The input parameters are learning objectives, learning material (represented by semantically annotated XML documents), student's profile (knowledge and preferences), teaching scenario (there are 6 predefined global pedagogical strategies) and pedagogical rules (sequencing and selection of learning objects defined using if-then rules). The paper [3] presents the system for automatic generation of a set of individualized hypermedia documents. A course is generated using student's model, domain knowledge, learning objects and a set of rules defined in JESS rule engine. Student's model is also used in [4] where a sequence of learning objects is created using planning mechanism and PDDL plan. This language is used in [5] for course generation, too. Learning activity is chosen in the real-time depending of a learning objective and student's profile. The paper [6] presents the system for the automatic generation of a learning path using ontology of learning objectives and student's model. A similar approach is used in [7], but without the formal representation of learning objectives. A course is generated directly using learning objects and relations are directly defined among learning objects.

This paper presents our system for automatic generation of e-learning courses. As we can see above, all similar systems have two basic components: learning objects and domain knowledge. Our system contains these two components, too. In addition, our system is focused on the instructional design in the course. The system contains a component which formally specifies instructional design used in the course. By changing the instructional design definition, different pedagogical strategies may be easily applied in the course. Then, these pedagogical strategies may be evaluated in order to find the most appropriate strategy for the course. This is the current purpose of our system. Currently, our system doesn't have a student's model as a component and it doesn't create a personalized course.

2. SYSTEM ARCHITECTURE

In this paper, an e-learning course is modelled using the model presented in [8]. The model is based on classical Tyler's rationale [9] and it specifies four modules in a curriculum: learning objectives, learning objects, instructional design and testing strategy. In this paper we are concerned about first three modules. E-courses (in popular e-learning systems or in globally adopted e-learning standards) mostly contain these modules. But, the course is represented as a monolithic unit and the modules are not explicitly represented as separated units. Using such representation, it is not possible to independently change only one module, which is practical demand very often.

Hence, we have decided to represent each module as a distinct component. That way, each component may be independently changed and the components are the input parameters for the automatic generation of an e-learning course. Our system generates a course using three input parameters: learning objectives (goals), learning objects and instructional design. Learning objectives may be defined as learning outcomes that a learner has to achieve. In this paper, we have decided to use ontology for representing learning goals which is an approach used in [4] and [6]. Our ontology of learning objectives is represented in OWL [10] language. In this paper we use the term "learning object" to refer to any digital content that helps a student to achieve a learning objective or evaluates if a learning objective is achieved. Since our system is intended for generating an e-course for web environment, we need learning objects represented in a browser-readable format. So, we have chosen HTML format for learning objects, packed in IMS Content Packaging [11] format. As mentioned, the usage of a learning object is always related to a specific learning objective (for learning or evaluation). Therefore, we created a mapping input component that maps learning objects to learning objectives. We use an XML document to specify this mapping. For generating a sequence of learning activities, it is necessary to define an instructional design used in the course. We have created a special-purpose XML-based language for describing the instructional design. The language specifies the sequencing and selection of learning objects. An XML schema for this language is presented in [8]. The result of our system is a formal description of an e-learning course. It may be defined in various formats, and we have chosen IMS Learning Design [12] format. A detailed description of all components may be found in [8]. The global architecture of our course generator is depicted in Figure 1.

3. SYSTEM IMPLEMENTATION

The system is implemented in Java programming language. The system parses the input files and creates the in-memory representation of data read from the input files. Firstly, the ontology of learning objectives and learning objects specification (in the IMS CP manifest file) are parsed. Then, learning objects are linked with learning objectives by parsing the mapping component. The last input file that is parsed is the XML document that describes course instructional design. Using all these data, a sequence of learning activities is generated. Finally, this sequence is a base for generating an IMS LD manifest file, which represents a formal description of our e-learning course. The object models of each component are described below.

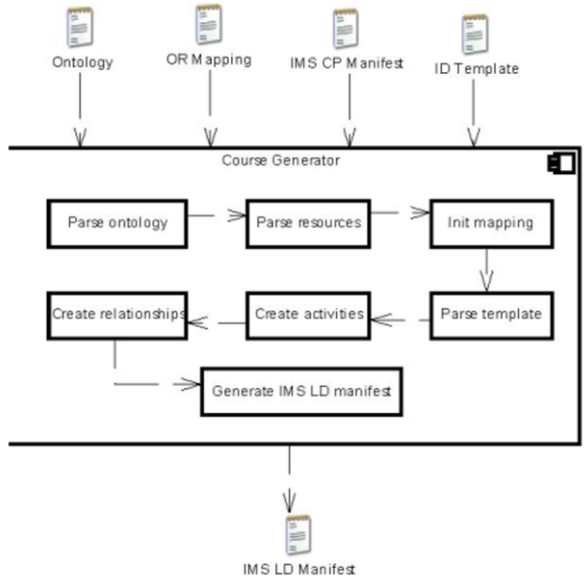

Figure 1. System architecture

3.1. E-learning course

An e-learning course has been modelled using an abstract class `CourseModel`. The course contains learning objectives, learning objects and information about the instructional design. These attributes are initialized from the system inputs. `TemplateParser` class parses the input XML document that describes instructional design. Using initialized attributes, learning activities are created. The course (in the concrete output format) is generated from the sequence of learning activities. `IMSLDCourseModel` class represents the course in the IMS LD format. The methods `createActivities` and `createActivity-Relationships` create the IMS LD representation of learning activities (see section 3.4). The overridden method `generateCourse` creates an IMS LD manifest file. The e-learning course model is shown in Figure 2.

3.2. Learning objectives and learning objects

For in-memory representation of learning objectives and learning objects, we have created the model shown in Figure 3. `Resource` class represents a learning object. For each learning object, the name, id and file path are managed. In addition, this class has a list of metadata that closely describe the learning object. For a learning objective (class `LearningObjective`), the system specifies the name and hierarchical level. Since learning objectives are hierarchically organized, each learning objective has a reference to its parent learning objective (attribute `parentObjective` in the figure). Likewise, a learning objective has a list of its children. Our ontology defines a relation `surmises` which specifies that a learning objective may be a precondition for other learning objectives. This relation among learning objectives is modelled using `ObjectivePrecondition` class. The attribute `source` in this class represents a learning objective that is a precondition for the learning objective defined by the `destination` attribute. `LearningObjective` class has the list of all `ObjectivePrecondition` objects where that learning objective is the source. Similarly,

it contains the list of `ObjectivePrecondition` objects where it is the destination. Mapping of learning objects to learning objectives is modelled in `ObjectiveResource` class.

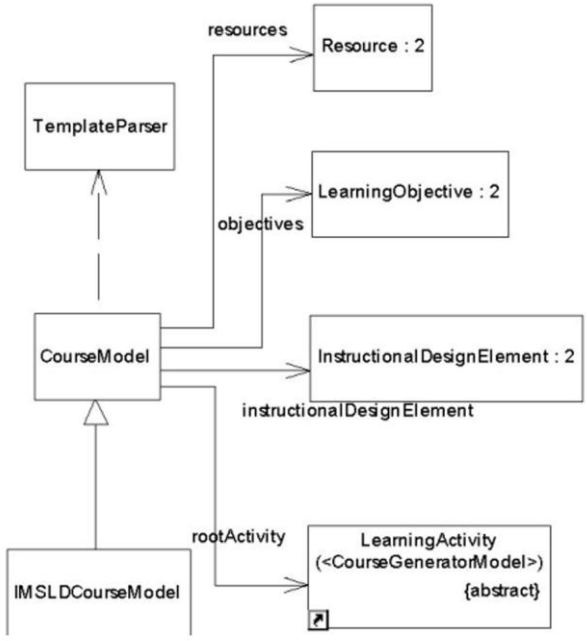

Figure 2. E-learning course model

The class contains a reference to a learning object and to an appropriate learning objective.

3.3. Instructional design

Based on the XML scheme for describing the instructional design in the course, we have created a corresponding object model shown in Figure 4. `InstructionalDesignElement` is a container class and it describes the organization of learning elements in the course. `LearningElement` is a generic learning element in the organization. Learning elements are hierarchically organized, so each learning element contains a reference to its parent learning element and a collection of child learning elements. Also, generic `LearningElement` has a unique identifier represented with `elementId` attribute. There are three different types of learning elements – element group, sequence and learning object. Element group is just a container for other learning elements. Sequence represents a chain of other elements. Its role is similar to the role of loop statements in programming languages. For each sequence, the system iterates through the elements specified with `element` attribute. Learning object is a learning element on the lowest hierarchical level. It actually represents a concrete learning resource. For sequences and learning objects, we should define a specific strategy for selecting learning resources. This strategy is defined in `SelectionRule` element. `SelectionRule` aggregates two lists of `ObjectSelection` elements. First list contains objects that will be included in the course. The second one is for excluded objects. Object selection element specifies learning

objects for including or excluding. Learning objects selection is done by specifying the values of its metadata attributes. Also, using the `priority` attribute we specify the order of included learning objects. Learning object may be used for evaluating student's knowledge. In such case, grading information is modelled using `Grading` class.

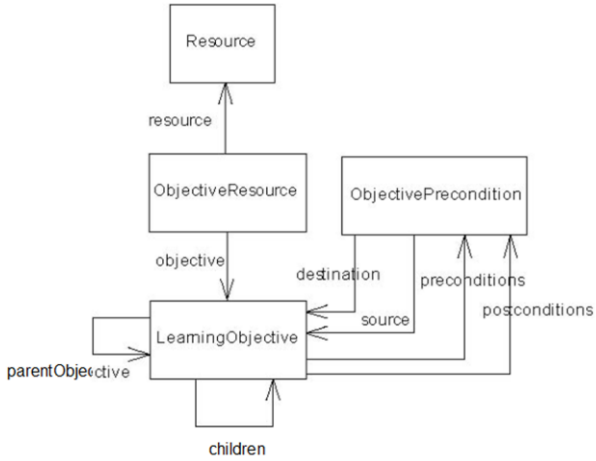

Figure 3. The model of learning objectives and learning objects

Beside the course structure, sometimes it is necessary to define relationships between learning elements. For example, in mastery learning student can't proceed to the next lesson if he hasn't completed the previous one. So, we need to define the relationship between two lessons. These relationships are modelled using `Element- Relationship` class. Learning elements which participate in the relationship are modelled using `ConditionElement` class. Learning elements creates the relationship only when a specific condition (modeled with `ElementJoin` class) is satisfied. `Element- Relationship` class contains the `ifCondition` reference which represents a condition for performing specific actions in the course.

Conditions are modelled using `ConditionExpression` class. When the condition is satisfied, specific actions defined in the `thenAction` list are performed. Otherwise, actions defined in the `elseAction` list are performed. Actions are modelled using `ConditionAction` class (see below).

`ConditionExpression` is an abstract class that represents a certain condition in the learning process. Condition may contain sub-conditions, which is represented with `childExpressions` attribute. `ConditionExpression` class has successors that represent concrete logical expressions (and, or, equals...). The successors override the method `calculate` that evaluates the value of the logical expression (true or false). `ConditionExpression` class and its successors are shown in Figure 5.

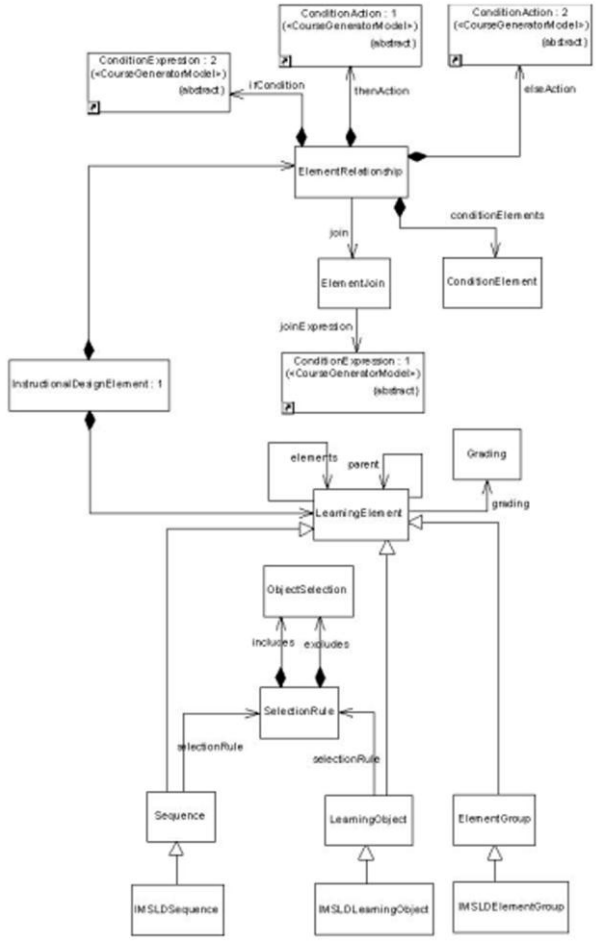

Figure 4. Instructional design model

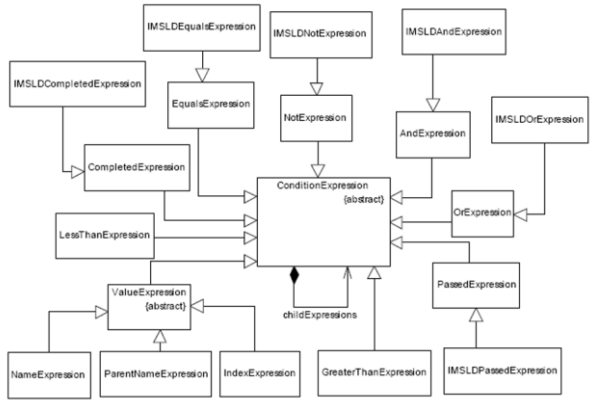

Figure 5. The model of logical expressions in the instructional design description

`ConditionAction` is an abstract class that represents a certain action that should be done in the learning process. Two types of action are supported: showing a learning element (`ActionShow` class) and hiding a learning element (`ActionHide` class). `ConditionAction` class and its successors are shown in Figure 6.

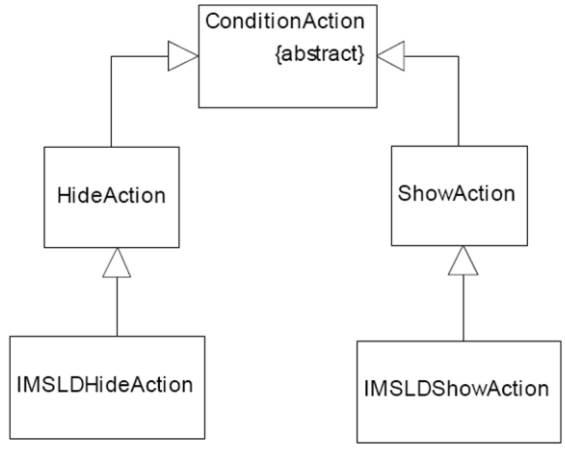

Figure 6. The model of course actions in the instructional design description

3.4. Learning activities

On the basis of input parameters our system creates learning activities in the course. The model of learning activities is shown in Figure 7. An abstract learning activity is represented with `LearningActivity` class. Activities are hierarchically organized. Each activity has a reference to its parent activity (`parentActivity` attribute) and a list of its child activities (`activities` attribute). Grading is a distinct activity and therefore represented with `GradingActivity` class. Grading is always related to a standard learning activity (`learningActivity` attribute), e.g. after the test (standard activity), the teacher gives the grades (grading activity). The relations among learning activities are modelled using `ActivityRelationship` class. The relationship contains a list of activities that form the relation (`learningActivities` attribute). When the `ifCondition` attribute is satisfied, the course performs the actions defined in the `thenAction` attribute. Otherwise, the actions specified in the `elseAction` attribute are performed.

3.5. IMS LD support

Described classes represent a general model of an elearning course and consequently they not define any specific output format. In order to generate a course in the specific output format, it is necessary to create successors of the described classes. We have created successors that generate a course in the IMS LD format. For each learning element in the instructional design model, we have created an appropriate successor class (`IMSLDElementGroup`, `IMSLDSequence` and `IMSLDLearningObject`). Likewise, for conditions and actions, we have created successors that generate conditions and actions in the IMS LD format. All successors override `exportToXML` method and create the XML nodes in the IMS LD manifest file.

`LearningActivity` and `ActivityRelationship` are abstract classes and they represent activities in general. We have created successors – `IMSLDLearning- Activity` and `IMSLDActivityRelationship` in order to generate activities in the IMS LD format. These classes export their attributes to the IMS LD format.

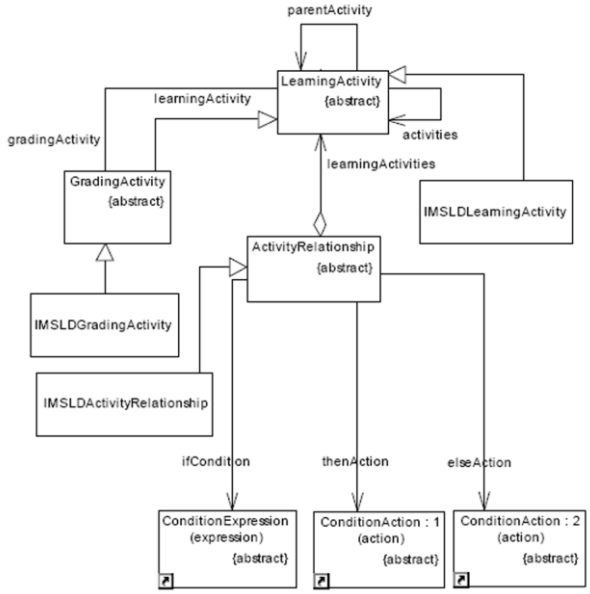

Figure 7. The model of learning activities in the instructional design description

4. SYSTEM OUTPUT

In this paper we have shown how to generate the course in the IMS LD format. But, the model enables generating a course in any XML-based format. It is only necessary to implement appropriate successor classes that export the content to desired format.

So far the system is used for generating the e-course *Numerical Algorithms and Numerical Software in Engineering* at Faculty of Technical Sciences in Novi Sad. We applied three instructional strategies and generated three versions of our course. A part of the one of the generated IMS LD manifest files is shown in listing 1.

The created course may be shown in any IMS LD player and we used Reload LD Player [13]. A screenshot of the generated course is shown in Figure 8.

```
<imsld:learning-activity
        identifier="Iterativni_postupci_1">
   <imsld:title>
        Iterativni postupci
   </imsld:title>
   <imsld:activity-description>
      <imsld:item identifierref =
                  "iterativni_postupci_res">
        <imsld:title>
              Iterativni postupci
        </imsld:title>
      </imsld:item>
   </imsld:activity-description>
</imsld:learning-activity>
```

Listing 1. A part of the generated IMS LD manifest file

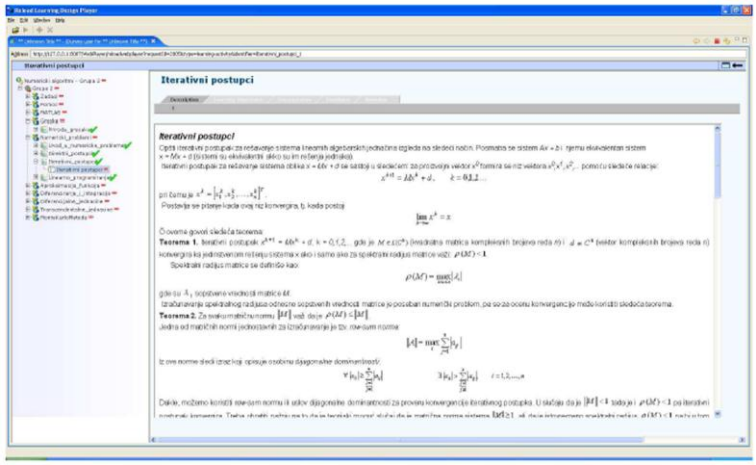

Figure 8. Screenshot of the generated course in the Reload LD Player

5. CONCLUSION

This paper describes the implementation of the system for utomatic generation of e-learning courses. The system architecture, object model and functionalities are presented. The model contains two layers. First layer models an abstract course. The second one models a course in the concrete output format. Our system generates courses in the IMS LD format. The model contains following subcomponents: the model of learning objectives and learning objects, instructional design model and the model of learning activities. System output is the IMS LD manifest file. The implementation is done in Java programming language.

By changing the input parameters, our system may generate different versions of an e-learning course. Still, the generated course is static – it contains a sequence of predefined learning activities. This could be taken as a drawback of our system. The system could be improved if the generated course chooses a next learning activity in the real-time. A student would dynamically get a learning activity depending on various parameters (instructional design, student's knowledge state, personal preferences ...). However, in this phase we have

chosen to generate only a static course because most of the popular elearning systems and e-learning standards don't have support for dynamic courses. Another disadvantage still remaining in our system is that it doesn't consider a student model, so the generated course is not personalized.

The system has been used at Faculty of Technical Sciences in Novi Sad for generating the course *Numerical Algorithms and Numerical Software in Engineering* which is presented as an illustrative example in this paper. In addition, we have generated the course *Web programming*. Our plan is to apply this e-course in the summer semester 2011.

Future work is concerned with analysis of the data collected while the generated course is used in the teaching process. Our short-term goal is to find the most appropriate instructional strategy for the *Numerical Algorithms and Numerical Software in Engineering* course. Also, the plan is to create graphical tools for defining input parameters. So far, we have created a graphical editor for defining course instructional design and we are implementing a graphical editor of learning objectives management. Long-term goal is to develop a system supporting dynamic courses and containing a student model as the input parameter. Using student's knowledge state and personal preferences, the system would generate a personalized e-course.

ACKNOWLEDGMENTS

Results presented in this paper are part of the research conducted within the Grant No. III-43007, Ministry of Science and Technological Development of the Republic of Serbia.

REFERENCES

[1] Brusilovsky, P., Vassileva, J., *Course sequencing techniques for large-scale web-based education,* Int. J. Cont. Engineering Education and Lifelong Learning, Vol. 13, Nos. 1/2, 2003.

[2] Melis, E., Andres, E., Budenbender, J., Frischauf, A., Goguadze, G., Libbrecht, P., Pollet, M., Ullrich, C., *ActiveMath: A Generic and Adaptive Web-Based Learning Environment*, International Journal of Artificial Intelligence in Education, (12), 2001.

[3] Kettel, L., Thomson, J., Greer, J., *Generating Individualized Hypermedia application,* Proc. Of Workshop on Adaptive and Intelligent Web-based Education Systems at 5th International Conference on Intelligent Tutoring Systems, Montreal, Canada, 2000.

[4] Kontopoulos, E., Vrakas, D., Kokkoras, F., Bassiliades, N., Vlahavas, I., *An ontology-based planning system for e-course generation*, Expert Syst. Appl. 35, no. 1-2, 2008.

[5] Hernandez, J., Baldiris, S., Santos, O.C., Fabregat, R., Boticario, J. G., *Conditional ims learning design generation using user modeling and planning techniques*, Proceedings of the 2009 Ninth IEEE International Conference on Advanced Learning Technologies, IEEE Computer Society, 2009.

[6] Capuano, N., Gaeta, M., Micarelli, A., Sangineto, E., *An integrated architecture for automatic course generation*, Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT 02), 2002.

[7] Lluvia, M., Luis, C., Juan, F.-O., Arturo, G.-F., *Automatic generation of user adapted learning designs: An AI-planning proposal*, Proceedings of the 5th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer-Verlag, Hannover, Germany, 2008.

[8] Segedinac, M., Savic, G., Konjovic, Z., *Knowledge representation framework for curriculum development*, International conference on knowledge engineering and ontology development, Valensia, Spain, 2010.

[9] Tyler, R., *Basic Principles of Curriculum and Instruction*, Chicago: University of Chicago Press, 1949.

[10] W3C, *OWL Web Ontology Language Overview*, 2004.

[11] IMS, *IMS content packaging specification*, 2007.

[12] IMS, *IMS Learning Design Information model*, 2003.

[13] Reload, Learning design player v. 2.1.3, 2010, Internet site: www.reload.ac.uk/ldplayer.html