I T H E A

International Journal

INFORMATION TECHNOLOGIES & KNOWLEDGE

2007 Volume 1 Number 2

International Journal
INFORMATION TECHNOLOGIES & KNOWLEDGE
Volume 1 / 2007, Number 2

IJ ITK is official publisher of the scientific papers of the members of
the Association of Developers and Users of Intellectualized Systems (ADUIS).

IJ ITK rules for preparing the manuscripts are compulsory.
The rules for the papers for IJ ITK as well as the subscription fees are given on *www.foibg.com*.
The camera-ready copy of the paper should be received by e-mail: info@foibg.com.
Responsibility for papers published in IJ ITK belongs to authors.

General Sponsor of IJ ITK is the Consortium FOI Bulgaria (www.foibg.com).

International Journal "INFORMATION TECHNOLOGIES & KNOWLEDGE" Vol.1, Number 2, 2007

Printed in Bulgaria

Edited by the Institute of Information Theories and Applications FOI ITHEA, Bulgaria,
in collaboration with the V.M.Glushkov Institute of Cybernetics of NAS, Ukraine, and
the Institute of Mathematics and Informatics and the Institute of Information Technologies, BAS, Bulgaria.

Publisher: Institute of Information Theories and Applications FOI ITHEA
Sofia, 1000, P.O.B. 775, Bulgaria. www.foibg.com, e-mail: info@foibg.com

ISSN 1313-0455 (printed) ISSN 1313-048X (online) ISSN 1313-0501 (CD/DVD)

DIMENSIONING OF TELECOMMUNICATION NETWORK BASED ON QUALITY OF SERVICES DEMAND AND DETAILED BEHAVIOUR OF USERS

Emiliya Saranova

Abstract: The aim of this paper is to be determined the network capacity (number of necessary internal switching lines) based on detailed users' behaviour and demanded quality of service parameters in an overall telecommunication system. We consider detailed conceptual and its corresponded analytical traffic model of telecommunication system with (virtual) circuit switching, in stationary state with generalized input flow, repeated calls, limited number of homogeneous terminals and losses due to abandoned and interrupted dialing, blocked and interrupted switching, not available intent terminal, blocked and abandoned ringing (absent called user) and abandoned conversation.

We propose an analytical - numerical solution for finding the number of internal switching lines and values of the some basic traffic parameters as a function of telecommunication system state. These parameters are requisite for maintenance demand level of network quality of service (QoS). Dependencies, based on the numerical-analytical results are shown graphically.

For proposed conceptual and its corresponding analytical model a network dimensioning task (NDT) is formulated, solvability of the NDT and the necessary conditions for analytical solution are researched as well. It is proposed a rule (algorithm) and computer program for calculation of the corresponded number of the internal switching lines, as well as corresponded values of traffic parameters, making the management of QoS easily.

Keywords: Telecommunication Network, Circuit Switching, Network Traffic, Terminal Traffic, Human Factors, Network Dimensioning.

ACM Classification Keywords: C.2.1 Network Architecture and Design (Circuit-switching networks); C.2.3 Network Operations (Network management); H.1.2 User/Machine Systems (Human factors).

1. Introduction

The purpose of the teletraffic theory is to find relation between quality of services and equipment cost [Iversen 2004]. This is very important for a good planning and controlling of telecommunication networks.

The Quality of service (QoS) concept is defined in the ITU-T Recommendation E-800 as: "The collective effect of service performance, which determines the degree of satisfaction of a user of the service".

QoS parameters are administratively specified in Service Level Agreement (SLA) between users and operators. These QoS parameters (from a contract of SLA) are reflecting on GoS parameters.

Network dimensioning is necessary for designing and control of network and its level of quality of services (QoS), in an advance determined level.

Based on a given set of QoS requirements, a set of GoS (Grade of service) parameters are selected and determined as functions of human behaviour characteristics.

2. Conceptual Model

In this paper we consider detailed conceptual and its corresponded analytical traffic model [Poryazov 2005b] of telecommunication system with channel switching, in stationary state, with BPP (Bernoulli-Poisson-Pascal) input flow, repeated calls, limited number of homogeneous terminals and losses due to abandoned and interrupted dialing, blocked and interrupted switching, not available intent terminal, blocked and abandoned ringing and abandoned conversation.

The conceptual model of the telecommunication system includes the paths of the calls, generated from (and occupying) the A – terminals in the proposed network traffic model and its environment (shown on Fig. 1).

The names of the virtual devices used are constructed according to the device position in the model.

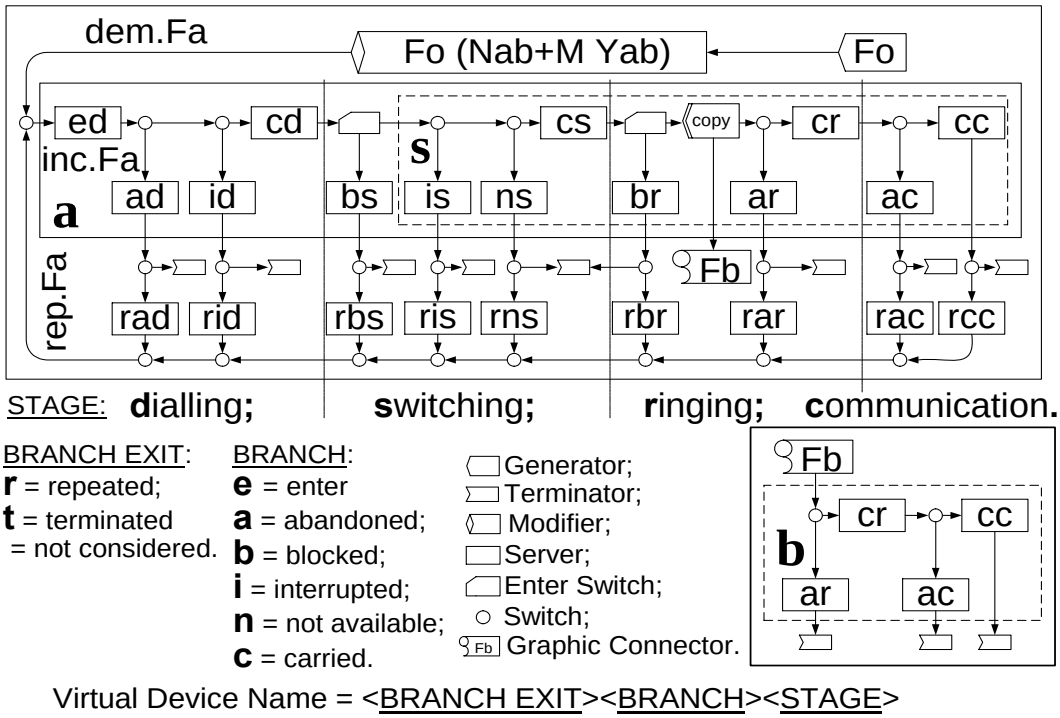


Fig. 1. Normalized conceptual model of the telecommunication system and its environment and the paths of the calls, occupying A-terminals (a-device), switching system (s-device) and B-terminals (b-device); base virtual device types, with their names and graphic notation.

2.1. The Comprising Virtual Devices

The following important virtual devices on Fig.1 are shown and considered:

a = comprises all the A-terminals (calling) in the system (shown with continuous line box).

b = comprises all the B-terminals (called) in the system (box with dashed line).

ab = comprises all the terminals (calling and called) in the system (not shown on Fig.1);

s = virtual device corresponding to the switching system. It is shown with dashed line box into the a-device. *Ns* stand for the capacity (number of equivalent internal switching lines) of the switching system.

2.2. Stages and Branches in the Conceptual Model:

Service *stages*: dialing, switching, ringing and communication.

Every service stage has *branches*: enter, abandoned, blocked, interrupted, not available, carried (correspondingly to the modeled possible cases of ends of the calls' service in the branch considered).

Every branch has two *exits*: repeated, terminated (which show what happens with the calls after they leave the telecommunication system). Users may make a new bid (repeated call), or to stop attempts (terminated call).

2.3. Device Parameters and its Notations in the Conceptual Model:

Letter *F* stands for intensity of the flow [calls/sec.], *P* = probability for directing the calls of the external flow to the device considered, *T* = mean service time, in the device, of a served call [sec.], *Y* = intensity of the device traffic [Erl], *N* = number of service places (lines, servers) in the virtual device (capacity of the device). In the normalized models [Poryazov 2001], used in this paper, every virtual device, except switches, has no more than one entrance and/or one exit. Switches have one entrance and two exits. For characterizing the intensity of the flow, we are using the following notation: *inc.F* for incoming flow, *dem.F*, *ofr.F* and *rep.F* for demand, offered and repeated flows respectively (ITU E.600). The same characterization is used for traffic intensity (*Y*).

Fo is the intent intensity of calls of one idle terminal; *inc.Fa* = *Fa* is intensity of incoming flow of A-terminals and *M* is a constant, characterizing the BPP flow of demand calls (*dem.Fa*). If *M* = -1, the intensity of demand flow

corresponds to Bernoulli (Engset) flow model, if *M* = 0 - to the Poisson (Erlang), and if *M* = +1 - to the Pascal (Negative Binomial) flow model. In our analytical model every value of *M* in the interval [-1, +1] is allowed. The BPP-traffic model is very applicable [Iversen 2004], but in the numerical examples, presented here, *M* = *0,* because the conclusions made are independent of the input flow model.

2.4. The Main Assumptions of the Model:

For creating a simple analytical model, we make the following system of fourteen (A-1 – A-14) assumptions [Poryazov 2005b]:

A-1. (Closed System Structure) We consider a closed telecommunication system with functional structure shown in Fig. 1;

A-2. (Device Capacity) All base virtual devices in the model have unlimited capacity. Comprising devices are limited: ab-device contains all the active $Nab \in [2, \infty)$ terminals; switching system (s) has capacity of *Ns* calls (every internal switching line may carry only one call); every terminal has capacity of one call, common for both incoming and outgoing calls;

A-3. (A-Terminal Occupation) Every call, from the flow incoming in the telecommunication system (*inc.Fa*), falls only on a free terminal. This terminal becomes a busy A-terminal;

A-4. (Stationarity) The system is in stationary state. This means that for every virtual device in the model (including comprising devices like switching system), the intensity of input flow F(0, t), call holding time T(0, t) and traffic intensity Y(0, t) in the observed interval (0, t) converge to the correspondent finite numbers F, T and Y, when $t \to \infty$. In this case we may apply the Theorem of Little (1961) and for every device: $Y = FT$;

A-5. (Calls' Capacity) Every call occupies one place in a base virtual device, independently from the other devices (e.g. a call may occupy one internal switching line, if it find free one, independently from the state of the intent B-terminal (busy or free));

A-6. (Environment) The calls in the communication systems' environment (outside the blocks a and b in Fig. 1) don't occupy any telecommunication systems' device and therefore they don't create communication systems' load. (For example, unsuccessful calls, waiting for the next attempt, are in "the head of" the user only. The calls and devices in the environment form the intent and repeated calls flows). Calls leave the environment (and the model) in the instance they enter a Terminator virtual device;

A-7. (Parameters' undependability) We consider probabilities for direction of calls to, and holding times in the base virtual devices as independent of each other and from intensity Fa = inc.Fa of incoming flow of calls. Values of these parameters are determined by users' behavior and technical characteristics of the communication system. (Obviously, this is not applicable to the devices of type Enter Switch, correspondingly to Pbs and Pbr);

A-8. (Randomness) All variables in the analytical model may be random and we are working with their mean values, following the Theorem of Little.

A-9. (B-Terminal Occupation) Probabilities of direction of calls to, and duration of occupation of devices ar, cr, ac and cc are the same for A and B-calls;

A-10. (Channel Switching) Every call occupies simultaneously places in all the base virtual devices in the telecommunication system (comprised of devices a or b) it passed through, including the base device where it is in the moment of observation. Every call releases all its occupied places in all base virtual devices of the communication system, in the instant it leaves comprising devices a or b.

A-11. (Terminals' Homogeneity) All terminals are homogeneous, e.g. all relevant characteristics are equal for every terminal;

A-12. (A-Calls Directions) Every A-terminal directs uniformly all its calls only to the other terminals, not to itself;

A-13. (B-flow ordinariness) The flow directed to B-terminals (Fb) is ordinary. (The importance of A-13 is limited only to the case when two or more calls may reach simultaneously a free B-terminal. A-13 may be acquitted from results like in (Burk 1956) and (Vere-Jones 1968);

A-14. (B-Blocking Probability for Repeated attempts) The mean probability (Pbr) of a call to find the same B-terminal busy at the first and at the all following repeated attempts is one and the same.

3. Analytical Model

3.1. Some General Equations

For the proposed conceptual model we derived the following system of equations (Poryazov, Saranova 2005):

$$Yab = Fa[S_1 - S_2(1-Pbs)\ Pbr - S_3\ Pbs] \tag{3.1.1}$$

$$Fa\ =\ dem.Fa\ +\ rep.Fa \tag{3.1.2}$$

$$dem.Fa\ =\ Fo\ (Nab\ +\ M\ Yab) \tag{3.1.3}$$

$$rep.Fa = Fa\ [R_1 - R_2\ Pbr\ (1-Pbs) - R_3\ Pbs] \tag{3.1.4}$$

$$Pbr\ = \begin{cases} \dfrac{Yab\text{-}1}{Nab\text{-}1} & in\ \ case\ of\ \ \ 1 \le Yab \le Nab, \\ 0 & in\ case\ of\ \ \ \ 0\ \le Yab < 1. \end{cases} \tag{3.1.5}$$

$$Ts\ = S_{1Z} - S_{2Z}\ Pbr \tag{3.1.6}$$

$$ofr.Fs\ =\ Fa\ (1\text{-}Pad)(1\text{-}Pid) \tag{3.1.7}$$

$$ofr.Ys\ =\ ofr.Fs\ \ Ts \tag{3.1.8}$$

$$Pbs\ =\ Erl\_b\ (Ns,\ ofr.Ys) \tag{3.1.9}$$

$$crr.Ys\ =\ (1-Pbs)\ ofr.Ys \tag{3.1.10}$$

The following notations are used:

$$S_1 = Ted + Pad\,Tad + (1-Pad)[Pid\,Tid + (1-Pid)[Tcd + Pis\,Tis + (1-Pis)[Pns\,Tns + (1-Pns)[Tcs + 2\,Tb]]]] \tag{3.1.11}$$

$$S_2 = (1-Pad)(1-Pid)(1-Pis)(1-Pns)[2\,Tb - Tbr] \tag{3.1.12}$$

$$S_3 = (1-Pad)(1-Pid)[PisTis + (1-Pis)[Pns\,Tns + (1-Pns)[Tcs + 2\,Tb]]] - (1-Pad)(1-Pid)\,Tbs \tag{3.1.13}$$

$$S_{1z} = Pis\,Tis + (1-Pis)[PnsTns + (1-Pns)(Tb + Tcs)] \tag{3.1.14}$$

$$S_{2z} = (1-Pis)(1-Pns)(Tb + Tcs) \tag{3.1.15}$$

$$R_1 = Pad\ \mathrm{Pr}\,ad + (1-Pad)(Pid\ \mathrm{Pr}\,id + (1-Pid)[Pis\,\mathrm{Pr}\,is + (1-Pis)(Pns\,\mathrm{Pr}\,ns + (1-Pns)Q] \tag{3.1.16}$$

$$R_2 = (1-Pad)(1-Pid)(1-Pis)(1-Pns)(Q - \mathrm{Pr}\,br) \tag{3.1.17}$$

$$R_3 = (1-Pad)(1-Pid)\{Pis\,\mathrm{Pr}\,is + (1-Pis)[Pns\,\mathrm{Pr}\,ns + (1-Pns)Q - \mathrm{Pr}\,bs]\} \tag{3.1.18}$$

$$Q = Par\,Prar + (1-Par)[Pac\,Prac + (1-Pac)\,Prcc] \tag{3.1.19}$$

An important assumption for proposed analytical model is:

The intent intensity of calls of one idle terminal is $Fo \ge 0$.

3.2. General Blocking Probability

Based on the conceptual model we define general blocking probability as follows:

Definition: General blocking probability (*Pbl*):

$$Pbl = \{Pbr \oplus Pbs, Pbr \in (0,1), Pbs \in (0,1):$$
$$(1-Pad)(1-Pid)[Pbs + (1-Pbs)(1-Pis)(1-Pns)Pbr]\} \quad (3.2.1)$$

Pad, *Pid*, *Pis*, *Pbs*, *Pns*, *Pbr*, *Par*, *Pac* and *Pcc* are known probabilities (see the conceptual model).

3.3. Probabilities of Blocking Switching (*Pbs*) and of Finding B-Terminal Busy (*Pbr*).

If $Pbr \in [0,1], Pbs \in (0,1]$ then each duple (*Pbr*, *Pbs*) defines a value of *Pbl* throw (3.2.1) and back, each value of *Pbl* defines a set of duples (*Pbr*, *Pbs*).

As GoS - parameter we consider general blocking probability *Pbl* based on (3.2.1).

Analogously, *adm.Pbl* (administratively determined value of *Pbl* in SLA in advance) defines set of duples (*adm.Pbr*, *adm.Pbs*) and back.

We consider general blocking probability (*adm.Pbl*) as a main QoS parameter, administratively determined in advance in SLA.

4. Network Dimensioning Task

4.1. Formulation of a Network Dimensioning Task (NDT):

1. To be dimensioned a network (to be found necessary number of internal switching lines), when in advance level of QoS is administratively determined and the values of known parameters are dimensioned and/ or calculated.
2. To be found the values of the unknown parameters, describing the system state in the upper case. For example, a system parameter, describing macrostate of the system (through the value of *Yab*), a terminal capacity of the system (the maximal number of active terminals *Nab*), intensity of demanded and repeated call attempts (respectively *dem.Fa* and *rep.Fa*), offered to the switching system traffic intensity (*ofr.Ys*) and others.

Parameters in the Network Dimensioning Task:

Administrative determined parameters:

adm.Pbl and *M* (4.1.1)

Known parameters:

$$Fo, Tb, S_1, S_2, S_3, R_1, R_2, R_3, S_{1z}, S_{2z} \quad (4.1.2)$$

Aim: To determine the number of switching lines *Ns*; and the following unknown parameters:

Yab, Fa, dem.Fa, rep.Fa, ofr.Fs, Ts, ofr.Ys (4.1.3)

Condition:

Pbl (*Pbr*, *Pbs*) ≤ *adm.Pbl* (4.1.4)

4.2. Solvability of the NDT:

The traffic intensity *Yab* characterizes the macrostate of the system. In Poryazov, Saranova (2005) is shown that

$$Yab = \frac{F_0(S_1 - S_3 Pbs) - (F_0(S_1 - S_3 Pbs) + F_0 S_2(1-Pbs))Pbr + F_0 S_2(1-Pbs)Pbr^2}{F_0(S_1 - S_3 Pbs) - (F_0 M(S_1 - S_3 Pbs) + F_0 S_2(1-Pbs) - 1 + R_1 - R_3 Pbs)Pbr + (1-Pbs)(F_0 M S_2 + R_2)Pbr^2} \quad (4.2.1).$$

Theorem 1: If $Pbr \neq 0$ and $Fo \neq 0$, then analytical presentation (4.2.4) of *Yab* in the NDT exist.

Proof: Considering the system equations (3.1.1) - (3.1.10) when $Pbr \neq 0$ and $Fo \neq 0$ from (3.1.5) and (3.1.3) follows

$$dem.Fa = \frac{Fo}{Pbr}[Pbr - 1 + (M\ Pbr + 1)Yab] \quad (4.2.2)$$

From (3.1.2) and (3.1.4) follows

$$dem.Fa = Fa\ \{1 - R_1 + R_2\ Pbr + (R_3 - R_2\ Pbr)\ Pbs\} \tag{4.2.3}$$

Then (4.2.2), (4.2.3) and (3.1.2) gives

$$Yab = \frac{F_0(1-Pbr)\{S_1 - S_2 Pbr - (S_3 - S_2 Pbr)Pbs\}}{F_0(1+MPbr)\{S_1 - S_2 Pbr - (S_3 - S_2 Pbr)Pbs\} - Pbr\{1 - R_1 + R_2 Pbr + (R_3 - R_2 Pbr)Pbs\}} \tag{4.2.4}$$

(4.2.4) is new simplified expression of the (4.2.1).

If *Fo* = 0, then obviously *Fa* = 0, *dem.Fa* =0 and *rep.Fa* = 0.

Therefore, when $Pbr \neq 0$ in NDT, on the base of administrative determined values of parameters *Pbs*, *Pbr*, *M* and the known parameters (4.1.2), traffic intensity *Yab* is derivable. The other system parameters in the NDT are depending on the system state (respectively on *Yab*).

We will prove that the values of unknown parameters (4.1.3) in the NDT can be derived (evaluated) through *Yab* and known parameters (4.1.2) in correspondence of determined conditions.

Theorem 2: If

$$Pbr \neq 0 \ and \ \ Pbs \neq \frac{S_1 - S_2 Pbr}{S_3 - S_2 Pbr} \tag{4.2.5}$$

in the NDT, then for each unknown parameter of (4.1.3), an analytical expression for its evaluation exists.

Proof: Using the system (3.1.1) – (3.1.10) and (4.2.1) by (*S*1 – *S*2 *Pbr*) - (*S*3 – *S*2 *Pbr*) $Pbs \neq 0$, follows

$$Fa = \frac{Yab}{S_1 - S_2\ Pbr - (S_3 - S_2\ Pbr)\ Pbs} \tag{4.2.6}$$

For *dem.Fa* from (3.1.3) and (3.1.5) is received (4.2.2).

It is resulted from (3.1.4) and (4.2.5):

$$rep.Fa = \frac{Yab\{R_1 - R_2\ Pbr - (R_3 - R_2\ Pbr)\ Pbs\}}{S_1 - S_2\ Pbr - (S_3 - S_2\ Pbr)\ Pbs}\ . \tag{4.2.7}$$

From (3.1.6) and (4.2.5) follows:

$$ofr.Fs = \frac{Yab\ (1 - Pad)(1 - Pid)}{S_1 - S_2\ Pbr - (S_3 - S_2\ Pbr)\ Pbs} \tag{4.2.8}$$

The parameter *Ts* can be calculated from (3.1.7), and from (3.1.3) and (3.1.5) follows:

$$ofr.Ys = \frac{(1 - Pad)(1 - Pid)(S_1 z - S_2 z\ Pbr)\ Yab}{S_1 - S_2\ Pbr - (S_3 - S_2\ Pbr)\ Pbs} \tag{4.2.9}$$

Therefore, the values of the unknown parameters (4.1.3) in the NDT can be expressed and calculated by the conditions of Theorem 1 and Theorem 2.

For the network dimensioning, when the level of QoS is determined administratively in advance (for example blocking probability *Pbs*), Erlangs'B - formula may be used:

$$Pbs = Erl\_b(Ns, ofr.Ys) \tag{4.2.10}$$

$$Erl\_b(Ns, ofr.Ys) = \frac{\frac{(ofr.Ys)^{Ns}}{Ns!}}{\sum_{j=0}^{Ns} \frac{(ofr.Ys)^j}{j!}}\ . \tag{4.2.11}$$

The number of switching lines *Ns* and the values of *ofr.Ys* are calculated by the conditions of Theorem 1 and Theorem 2.

Remark 1-2: *ofr.Ys,* being evaluated on the base of the Theorem 1 and Theorem 2 for *adm.Pbs* and *adm.Pbr*, is resulted in a fixed value. Then *Pbs* = *Pbs* (*Ns,ofr.Ys*) is a function of *Ns* only and *Pbs* = *Pbs* (*Ns*).

Theorem 3: The function *Pbs* = *Pbs* (*Ns,ofr.Ys*), defined through (4.2.11) in the NDT is strictly monotone decreasing according to $Ns \geq 1$, when *ofr.Ys* >0 is a fixed value.

Proof: It can be proved that *Pbs* (*Ns*+1, *ofr.Ys*) < *Pbs* (*Ns, ofr.Ys*). Obviously (see (4.2.11)) $Pbs(0, ofr.Ys) = 1$. Using the recursion Erlangs'B - formula [Iversen 2004]:

$$Pbs(Ns, ofr.Ys) = \frac{ofr.Ys\ Pbs(Ns-1, ofr.Ys)}{Ns + Pbs(Ns-1, ofr.Ys)}. \quad (4.2.12)$$

But *Pbs* (*Ns,ofr.Ys*) > 0 when *ofr.Ys* >0 and $Ns \geq 1$, *ofr.Ys Pbs* (*Ns,ofr.Ys*)+*Ns*+1> 0 and

$$Pbs(Ns+1, ofr.Ys) - Pbs(Ns, ofr.Ys) = Pbs(Ns, ofr.Ys)\frac{ofr.Ys[1 - Pbs(Ns, ofr.Ys)] - (Ns+1)}{ofr.Ys\ Pbs(Ns, ofr.Ys) + (Ns+1)} =$$

$$= Pbs(Ns, ofr.Ys)\frac{crr.Ys - (Ns+1)}{ofr.Ys\ Pbs(Ns, ofr.Ys) + (Ns+1)}$$

Because *crr.Ys* ≤ *Ns* follows *crr.Ys* – (*Ns*+1) < 0.

Therefore, *Pbs* (*Ns*+1, *ofr.Ys*) – *Pbs* (*Ns, ofr.Ys*) < 0 and the function *Pbs* = *Pbs* (*Ns,ofr.Ys*), defined through (4.2.10) is strictly monotone decreasing, when *ofr.Ys* > 0 is fixed value.

5. Analytical Solution

Based on the Assumption A-8 we are working with mean values of the parameters. Various techniques for analyzing complex teletraffic systems require a formulation of the Erlang function that is continuous in the parameter *Ns*. This is done via the integral representation [Berezner 1998].

Theorem 4: There is only one solution in the NDT through the equation

$$Erl\_b\ (Ns, ofr.Ys) = adm.Pbs, \quad (5.1.1)$$

according to the number of switching lines *Ns*.

Adm.Pbs ∈(0; 1] is in advance administratively determined value of blocking probability, providing of QoS.

Proof: Existence: It was proved, that the function *Pbs* = *Pbs* (*Ns,ofr.Ys*), defined through (4.2.10) in the NDT, is strictly monotone decreasing, when *ofr.Ys* >0 is fixed value. The absolute maximum is 1 and 0 is absolute minimum of the function. There is only one solution for equation (4.2.11) for *adm. Pbs* ∈(0; 1], relying of the Intermediate Value Theorem (Dirschmidt, H. Yorg, 1992).

Uniqueness: Admitting that there are two different solutions $N_S' \neq N_S''$ of the equation (3.1.1) – (3.1.19) for *adm. Pbs* ∈(0; 1], therefore they are simultaneously fulfilled $Pbs(N_S', ofr.Ys) = adm.Pbs$ and $Pbs(N_S'', ofr.Ys) = adm.Pbs$, is contradicting to Theorem 3.

It is proved that only one solution of *Ns* exists, fulfilling the equation (4.2.11) and corresponding to the determined administratively in advance value of the blocking probability *adm.Pbs* ∈(0; 1].

6. Algorithm for Calculating the Values of the Parameters in the NDT:

1. From SLA and ITU-Recommendation are specified and determined administratively blocking probability *adm.Pbl*, respectively *adm. Pbs* ∈(0; 1] and *adm.Pbr* ∈[0; 1]:

$$\forall adm.Pbl \Rightarrow \exists (adm.Pbr, adm.Pbs):$$
$$adm.Pbr \oplus adm.Pbs = adm.Pbl:\ adm.Pbr \in [0,1], adm.Pbs \in (0,1] \quad (6.1.1)$$

2. The unknown parameters (4.1.3) in the NDT are evaluated on the base of Theorem 1 and Theorem 2, known parameters (4.1.2), especially *adm.ofr.Ys* (*adm.Pbr*, *adm.Pbs*).

3. On the basis of each calculated value *adm.ofr.Ys*, we evaluate

$$\tilde{Ns} \in R_+ : \{\forall (adm.ofr.Ys, \tilde{Ns}) : Pbs(adm.ofr.Ys, \tilde{Ns}) = adm.Pbs\} \quad (6.1.2)$$

4. If $adm.Ns = \sup \tilde{Ns}$, then

$$Ns = [adm.Ns] + 1 : Pbl \le adm.Pbl. \quad (6.1.3)$$

5. For finding of the number of internal switching lines *Ns,* a computer program is created on the base of the recursion Erlangs'B – formula (4.2.11) [Iversen 2004]. From numerical point of view, the following linear form is the most stable:

$$I(Ns, ofr.Ys) = 1 + \frac{Ns}{ofr.Ys} I(Ns-1, ofr.Ys), \qquad I(0, ofr.Ys) = 1, \quad (6.1.4)$$

where $I(Ns, ofr.Ys) = 1/Pbs(Ns, ofr.Ys)$. This recursion formula is exact, and for large values of (*Ns*, *ofr.Ys*) there are no round of errors.

6. The received results for numerical inversion of the Erlang's formula (for finding the number of switching lines *Ns*) were confirmed with results of others commercial computer programs.

Therefore, it is proved that if $Pbr \neq 0$ and $Pbs \neq (S_1 - S_2 Pbr)/(S_3 - S_2 Pbr)$, then the NDT is solvable and there is proposed algorithm for its solution.

When *Pbr* = 0 the network loading is rather low and it is not of great practical interest, but in this case a mathematical research is made also.

7. Numerical Results

Among the easy computable QoS - parameters in the system (resulted from QoS- strategy of the network operators) is blocking probability *Pbl* in *pie-form model* [Poryazov 2000]. The sum of the loss probabilities due to abandoned and interrupted dialing, blocked and interrupted switching, not available intent terminal, blocked and abandoned ringing and abandoned conversation in *pie- form model* is 1.

For finding of the main teletraffic characteristics in proposed conceptual and its corresponding analytical model, the so called *normal - form model* (see Fig. 1) is used for presentation of blocking switching probability (*Pbs*) and probability of finding B-terminal busy (*Pbr*).

Based on the conceptual and its corresponding analytical model (3.1.1) - (3.1.19), defined general blocking probability *Pbl* is presented in *pie-form model* in (3.1.20) as function of the *Pbr* and *Pbs* (which are presented in *normal- form model* in the same equation*)*.

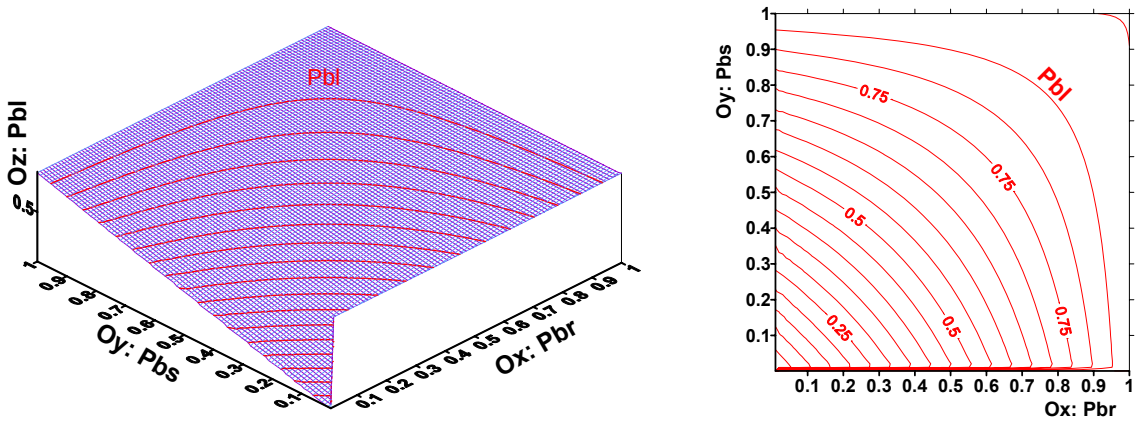


Fig. 2. General blocking probability *Pbl* in *pie-form model* is presented as function of probability of finding B-terminal busy *Pbr* and probability of blocking switching *Pbs* in *normal-form Pbr* and *Pbs* in 3D and contour - map presentation.

On the Fig. 2 blocking probability *Pbl* is shown in *pie- form model*, depending on probability of finding B-terminal busy *Pbr* (Ox – axis) and probability of blocking switching *Pbs* (Oy – axis) in *normal - form model*. *Pbl* increases

when *Pbr* and *Pbs* increase. Therefore, when *adm.Pbl* is predetermined as level of QoS administratively then *adm.Pbr* and *adm.Pbs* can be determined (evaluated) correspondingly.

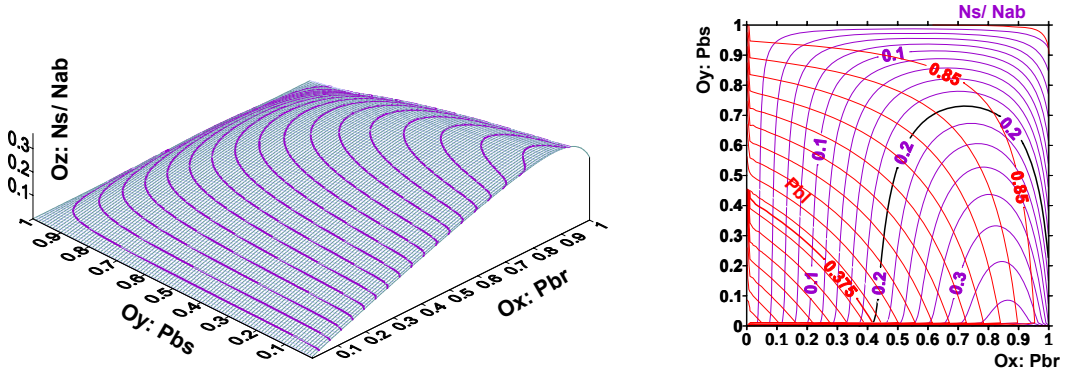


Fig. 3. The number of equivalent internal switching lines *Ns* (as percentage of number of active terminals *Nab*, where *Nab*= 7000 terminals) and general blocking probability *Pbl* are shown as functions of *Pbr* (Ox – axis) and *Pbs* (Oy – axis) in *normal-form model*, as well.

Conclusions of the numerical experiments:

According *Pbr*, *Pbs* and *Pbl*:

If $Pbr \in [0;1]$ and $Pbs \in [2\times10^{-9};0.999917]$ then

1. $Pbl \in [0;0.900896]$.
2. $0.000143 \le \dfrac{Ns}{Nab} \le 0.387857$, $Ns \in [1;2715]$, when $Nab = 7000$;
3. $0.77\times10^{-5} \le \dfrac{ofr.Ys}{Nab} \le 1.728311$, $ofr.Ys \in [0.473782;12098.18]$, when $Nab = 7000$.

   Ofr.Ys may exseed *Nab* by 73% approximately. This is "unproductiveness occupying of resources".
4. Absolute maximum for *ofr.Ys*:

Maximum *ofr.Ys* = 12098.18 and this value is about 4.9 times greater than switching system capacity *Ns* = 2715.

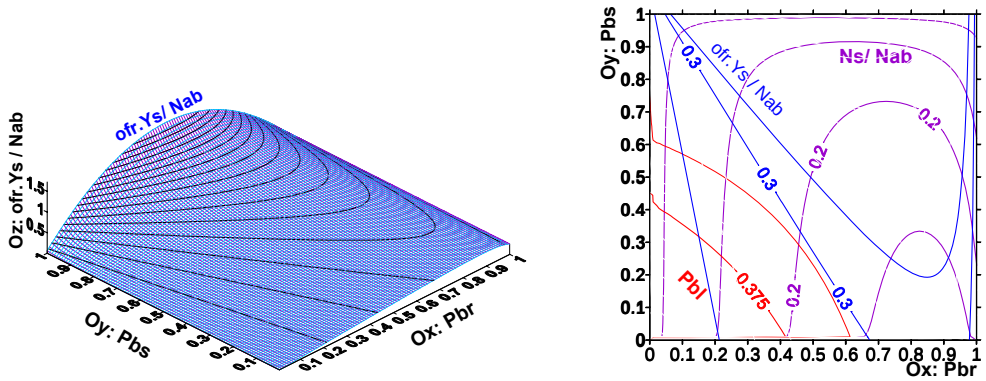


Fig. 4. The offered traffic *ofr.Ys*, the number of internal switching lines Ns and general blocking probability *Pbl* are presented as function of probability of finding B-terminal busy *Pbr* and probability of blocking switching *Pbs* in *normal – form model* in 3D and contour- map presentation.

Absolute maximum for *Ns*:

Ns = 2715 when *Nab*=7000 terminals, *Ns* = 38.79% of *Nab*. This is possible if *Pbl* = 0.900882 ≈ 90% (maximum theoretical value of *Pbl*), *Pbr*= 0.876623 ≈ 87.7% and *Pbs* = 9.28×10^{-9}, *Yab*= 6136.487 Erl ≈ 87.66 % of *Nab* and *ofr.Ys* = 2452.021 Erl = *crr.Ys* ≈ 35.0289 % of *Nab*.

8. Conclusions

1. Detailed normalized conceptual model, of an overall (virtual) circuit switching telecommunication system (like PSTN and GSM) is used. The model is relatively close to the real-life communication systems with homogeneous terminals.
2. General blocking probability *Pbl* as GoS parameter and *adm.Pbl* as QoS – parameter in *pie - form model* are formulated. The offered traffic *ofr.Ys*, the number of internal switching lines *Ns* and general blocking probability *Pbl* are derived as functions of probability of finding B-terminal busy *Pbr* and probability of blocking switching *Pbs* in *normal – form model*.
3. The network dimensioning task (NDT) is formulated on the base of preassigned values of QoS parameter *adm.Pbl* and its corresponding GoS - parameters - *adm.Pbr* and *adm.Pbs*; The NDT is formulated on condition that $Pbl \le adm.Pbl$.
4. The conditions for existence and uniqueness of a solution of the NDT are researched and an analytical solution of the NDT is found;
5. An algorithm and a computer program for a calculation the values of the offered (*ofr.Ys*), carried (*crr.Ys*) traffic and the number of equivalent switching lines *Ns*, are proposed. The results of numerical solution are derived and graphically shown;
6. The received results, in NDT, make the network dimensioning, based on QoS requirements easily;
7. The described approach is applicable directly for every (virtual) circuit switching telecommunication system (like GSM and PSTN) and may help considerably for ISDN, BISDN and most of core and access networks dimensioning. For packet switching systems, like Internet, proposed approach may be used as a comparison basis especially when they work in circuit switching mode.

Bibliography

Berezner S.A., 1998, - On the inverse of Erlang's Function- J. Appl. Prob.35, 246- 252

Dirschmidt, Hans Yorg, 1992 - Mathematische Grundlagen der Elektrotechnik-Braunschweig, Wiesbaden, pp.55

Engset, T., 1918. The Probability Calculation to Determine the Number of Switches in Automatic Telephone Exchanges. English translation by Mr. Eliot Jensen, Telektronikk, juni 1991, pp 1-5, ISSN 0085-7130. (Thore Olaus Engset (1865-1943). "Die Wahrscheinlichkeitsrechnung zur Bestimmung der Wählerzahl in automatischen Fernsprechämtern", Elektrotechnische zeitschrift, 1918, Heft 31.)

ITU E.501. ITU-T Recommendation E.501: Estimation of traffic offered in the network. (Previously CCITT - Recommendation, revised 26. May 1997)

ITU E.600, ITU-T Recommendation E.600: Terms and Definitions of Traffic Engineering (Melbourne, 1988; revised at Helsinki, 1993).

ITU E.800. ITU-T Recommendation E.800: Terms and Definitions related to Quality of Service and Network Performance, including Dependability. (Helsinki, March 1-12, 1993, revised August 12, 1994).

Iversen V. B., 2004. Teletraffic Engineering and Network Planning, Technical University of Denmark, pp.125

Little J. D. C., 1961. A Proof of the Queueing Formula L=λW. Operations Research, 9, 1961, 383-387.

Poryazov S. A, Saranova E. T., 2002. On the Minimal Traffic Measurements for Determining the Number of Used Terminals in Telecommunication Systems with Channel Switching. In: "Modeling And Simulation Environment for Satellite and Terrestrial Communication Networks - Proceedings of the European COST Telecommunications Symposium", Kluwer Academic Publishers, 2002, pp. 135-144;

Poryazov S. A, Saranova E. T. 2005. Some General Terminal and Network Teletraffic Equations in Virtual Circuit Switching Systems. Symposium "Modeling and Simulation Tools for Emerging Telecommunications Networks: Needs, Trends, Challenges, Solutions", Munich, Germany, 8 - 9 September 2005, Institut für Technische Informatik, Universität der Bundeswehr München (in printing in LNCS, Springer).

Poryazov S. A. 2005a. Can Terminal Teletraffic Theory Be Liberated from the Main Illusions? In: Proceedings of the International Workshop "Distributed Computer and Communication Networks", Sofia, Bulgaria 24-25 April, 2005, Editors: V. Vishnevski and Hr. Daskalova, Technosphera publisher, Moscow, Russia, 2005, ISBN 5-94836-048-2, pp. 126-138.; COST-285 TD/285/05/04; COST 290 TD(05)009.

Poryazov S. A. 2005b. What is Offered Traffic in a Real Telecommunication Network? COST 285 TD/285/05/05; 19th International Teletraffic Congress, Beijing, China, August 29- September 2, 2005, accepted paper No 32-104A.

Poryazov S. 1991. Determination of the Probability of Finding B-Subscriber Busy in Telephone Systems with Decentralized Control. Comptes Rendus de l'Academie Bulgare des Sciences – Sofia, 1991, Tome 44, No.3, pp. 37-39.

Poryazov S. A. 2000. Потребительский и терминальный телетрафик- классификация и обозначения , Conference "Telecom'2000" - Varna, Bulgaria, October 2000 – pp. 58-59.

Poryazov S. A. 2001. On the Two Basic Structures of the Teletraffic Models. Conference "Telecom'2001" - Varna, Bulgaria, 10-12 October 2001 – pp. 435-450.

Poryazov, S. A. 2004. The B-Terminal Busy Probability Prediction. IJ Information Theories & Applications, Vol.11/2004, Number 4, pp. 409-415;

Author's Information

Emiliya Saranova – Institute of Mathematics and Informatics - Bulgarian Academy of Science, Sofia, Bulgaria; High College of Telecommunication and Posts, Sofia, Bulgaria; e-mail: saranova@hctp.acad.bg

IMPLICATIONS OF RECENT TRENDS IN TELECOMMUNICATIONS ON MODELING AND SIMULATION FOR THE TELECOMMUNICATION INDUSTRY

Gerta Köster, Stoyan Poryazov

Abstract: With this paper we would like to trigger a discussion on future needs of modeling and simulation techniques and tools for the telecommunication industry. We claim that the telecommunication market has undergone severe changes that affect the need for and type of simulations in industrial research. We suggest some approaches how to address these new challenges. We believe that there is need for intensive research in the area.

Keywords: Telecommunications, Market Evolution, Modeling and Simulation Challenges.

ACM Classification Keywords: D.2.2 Design Tools and Techniques; K.6.1 Project and People Management.

Introduction

Models and simulations in telecommunications fulfill a number of tasks. We build models and run simulations to check the design of emerging products, we model the environment in which such a product is employed and we evaluate and optimize the performance of telecommunication equipment.

In the telecommunications industry the ultimate goals behind these tasks is to save costs and to make money by better or faster design and efficient development support.

We claim that the world of telecommunication has lately undergone severe changes that affect the need for and type of simulations in industrial research. In the course of this paper, we will outline the main trends we observe and discuss the implications on modeling and simulations.

1. Fractalization of the "old" incumbent telecommunication market.

2. Shift of operators' interest from providing a network to providing services and applications.

3. Shortening of development cycles.

4. Telecommunication and telecommunication problems pervade other sectors of private and business life.

Evolution of the Telecommunication Market

2.1. Fractal markets: We observe that a large part of the market has become fractal as opposed to monolithic in the past. The large state owned companies have been replaced by private enterprises. This is true for both, the Western industries, where the telecommunication business has been privatized and the former Soviet block where telecommunication has been denationalized.

This means that a large number of companies are producing and offering new - often small - products, services and devices, among which the customers, operators and consumers, may choose. Often these products complement each other enhancing each other's value. This means that there are no longer single solutions out of one hand.

An example might be the offering of ring-tones that enlivens the world of mobile communications as a complementary business. Another one is the home entertainment sector where networks are closely intertwined with the services, such as TV over broadband, that are offered through the networks and even the content that is offered.

2.2. Enlarged value chain: Operators and manufacturers seek to get a bigger portion of the value chain by offering applications and services or at least middleware to enable applications and services. This can be observed, for example, in the home entertainment market. Broadband internet access opens a new channel to the end consumers of content, e.g. movies, and thus offers an opportunity for direct cooperation between content manufacturers and operators such as the big national telecommunication providers. We may, at some point, see a merger of network provisioning and services.

2.3. Shortened development cycles: Industrial research is always coupled with a certain product that the company wishes to launch. When products become smaller and follow, to a certain extend, the fashion of the day, development cycles shorten and design becomes more volatile. In industry, we need to cope with these shortened development cycles. Otherwise people may cease to use simulations as a decision basis. As a matter of fact, we have the impression that there is already a decline in the use of simulations. We think, that there is a need for "rapid modeling", may be even accepting a degradation of simulation quality.

2.4. Telecommunication pervades other industrial and private sectors: Telecommunication, at the very beginning, dealt with the communication between two people at some distance. Today not only people communicate, but people with devices and devices with devices. We see, for example, that the control units in a truck (for motor, breaks, gearing) exchange information across a mini communication system. Tags are attached to new cloths that communicate with the security system of a store through a small radio system. Infotainment systems in cars communicate with some traffic control network outside the car or with other cars.

There are many more examples. But do we carry over the knowledge of how to build a good telecommunication system to these areas?

3. State of the Modeling and Simulation Art in Industry

3.1. Adaptive development processes: In “old” product design, requirements were supposed to be complete and clear before product development began. With today’s explorative technology and rapidly changing markets (e.g. in telecommunications today), modeling and simulation targets are difficult to fix. A new development process must enable quick requirement adaptation, often without sufficient prior modeling and simulation. Some of the changes in the telecommunication market are discussed in Andersen (2002).

3.2. Volatility of modeling requirements: The lack of stable modeling and simulation requirements makes it difficult to validate and verify product models. Product "verification" usually refers to testing whether a product meets certain specifications. "Validation" seeks to ensure that the customer is satisfied and that the correct specifications were incorporated into the product.

3.3. Availability of software tools: The design and manufacturing processes may be presented as activities, both series and parallel, at several levels of detail over time during the development of a product. As depicted in NRC (2004), software tools are not available for 60% of the required product development activities. For other activities software tools may be emerging or even commonly available. However they rarely interoperate and their use is often inefficient.

3.4. Tool interoperability: Only when tools are available and fully interoperable, designers and engineers can use and link various data and models for a given activity as well as across different activities required for product design and realization. It yet needs to be demonstrated whether modeling and simulations tools can be integrated across multiple domains including geometric modeling, performance analysis, life-cycle analysis, cost analysis, and manufacturing.

3.5. Need for econometric evaluation of modeling and simulation: The critical decisions in the design and manufacturing of a product are taken in the early stages of its life cycle, based on the products’ modeling and

simulation. But the economic usefulness of product modeling and simulation is still difficult to judge. We need methods and tools for the econometric evaluation of product modeling and simulation (Sargent et al 2000). The usefulness of models is also discussed in (Andersen 2002) also.

3.6. Multilayer modeling merging the product, management and market levels: Modeling and simulation play an important role on the product design level, on the management level (Ericsson AB 2005) and on the market level (Andersen 2002). Merging these levels implies multilayer modeling, incorporating different paradigms, languages and methods. Such a methodology is emerging in physics (Fishwick et al. 1992) but a modeling methodology that integrates the languages of physics, management and economics is still a matter of the future.

4. Challenges and Conclusion

On the one hand, the incumbent telecommunication market seems to undergo severe changes towards a fast-living volatile world and on the other hand "classic" telecommunications modeling and simulation problems pop up in new and unexpected areas of life.

We think that modeling and simulation can respond to these changes by tackling the following challenges:

- Develop new paradigms for modeling and simulating emerging New Generation Networks, applications and services that take into account the volatility of product requirements.
- Integrate dynamic models with different levels of abstraction, different paradigms, and different modeling languages.
- Development of meta-modeling methodology and tools to evaluate the technical and econometric usefulness of modeling and simulation in every stage of product life cycle.
- Establish sets of clear criteria for modeling and simulation needs for industrial applications, such as the necessary levels of sensitivity and accuracy and the ability to adapt to requirement changes.
- Establish suitable methods and criteria to verify, validate, and certify model trustworthiness for emerging systems, devices and their environments, especially when we develop adaptable methods.
- Increase the reusability of model components as well as of data of the real systems measurements and of simulation results.
- Integrate the methods and tools for design, creation, management and control of telecommunication products and systems (for a preliminary approach see Caughlin 2000).

5. References

Andersen 2002. Outlook of the development of technologies and markets for the European Audio-visual sector up to 2010. http://europa.eu.int/comm/avpolicy/stat/tvoutlook/tvoutlook_finalreport.pdf

Caughlin 2000. Don Caughlin. An Integrated Approach to Verification, Validation, and Accredition of Models and Simulations. Proceedings of the 2000 Winter Simulation Conference, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

Ericsson AB 2005. Enhancing Telecom Management. White Paper. http://www.ericsson.com/technology/whitepapers/telecom_management.pdf

Fishwick et al. 1992. Paul A. Fishwick and Bernard P. Zeigler. A Multimodel Methodology for Qualitative Model Engineering. ACM Transactions on Modeling and Computer Simulation, 2(1):52- 81, 1992.

NRC 2004. National Research Council of the National Academy of Sciences, 2004. Retooling Manufacturing: Bridging Design, Materials, and Production. http://www.nap.edu/catalog/11049.html)

Sargent et al 2000. Robert G. Sargent, Priscilla A. Glasow, Jack P.C. Kleijnen, Averill M. Law, Ian McGregor, Simone Youngblood. Strategic directions in verification, validation, and accreditation research. http://citeseer.ist.psu.edu/cache/papers/cs/17452/http:zSzzSzcwis.kub.nlzSz~few5zSzcenterzSzstaffzSzkleijnenzSzsargentr.pdf/strategic-directions-in-verification.pdf (Access 23.05.2006)

Authors' Information

Gerta Köster – Siemens AG, e-mail: gerta.koester@siemens.com

Stoyan Poryazov – Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, e-mail: stoyan@cc.bas.bg

REQUIREMENTS MANAGEMENT AND ACQUISITION MANAGEMENT EXPERIENCES IN SPANISH PUBLIC ADMINISTRATIONS

Jose A. Calvo-Manzano, Gonzalo Cuevas, Ivan Garcia,Tomas San Feliu, Ariel Serrano, Magdalena Arcilla, Fernando Arboledas, Fernando Ruiz de Ojeda

Abstract: As a part of the activities of the first Symposium on Process Improvement Models and Software Quality of the Spanish Public Administration, working groups were formed to discuss the current state of the Requirements Management and Supplier Agreement Management processes. This article presents general results and main contributions of those working groups. The results have allowed the obtention of a preliminary appraisal of the current state of these two processes in the Spanish Public Administration.

Keywords: Requirements engineering, subcontracting management, improvement models, process.

ACM Classification Keywords: D.2.9 Management, K.6.3 Software Management

Introduction

Most of the organizations have the same problems in their software production process in spite of the advances in Information Technologies (IT). These problems are:

- The software product is delivered, most of the time, with 15% defects.
- One quarter of software projects are not finished or they are abandoned.
- From 30% - 45% of software costs are dedicated to software rework.
- Only one half of the plans and schedules established at the beginning of the projects are satisfied.

Over the last decade, some of the software engineering research centers began to organize the practices that are used to produce and maintain software and have demonstrated to be effective in some organizations [1].

The Software Engineering Institute (SEI) defines a process as a set of practices to perform and obtain a result, including tools, techniques, materials and people. This set of tools, techniques, materials and people is named "Software Process". The SEI has grouped the effective practices in reference models.

A reference model is a set of processes that helps organizations to know their process status and it is used as a guide to improve them. One of the most important improvement process models is the "Capability Maturity Model for Software (Sw-CMM)", developed by the SEI [2]. Currently, this model is recognized for its integrated version (Capability Maturity Model Integration, CMMI) [3]. The purpose of CMMI is to provide a "road map" for process improvement that works as a framework to improve processes in an effective way. This "road map" will be a useful guideline to improve organizational processes. In addition, the CMMI offers a structured framework to evaluate the organization's current processes and establish priorities for the improvement activities.

Motivation

Due to the increasing use of Internet and the continuous developing of new IT, Public Administrations are focused on improving their services. In this context, Madrid's Region (called "Comunidad Autónoma") through its Informatics and Communications Agency in collaboration with the Polytechnical University of Madrid, organized the 1st Symposium on Process Improvement Models and Software Quality of Public Administrations sponsored by some software companies.

The Symposium Committee invited the Spanish General Public Administration, the 17 Autonomous Regions Public Administration and the some city councils to participate in the symposium. Finally, the Symposium had an audience of 133 participants from three Ministries, eleven Autonomous Regions and two local Councils. Besides, eleven enterprises participated in the Symposium.

The Symposium objectives were:

1. Spread the evolution on Process Improvement Models, mainly CMMIv1.1 [4] [5].
2. Identify the benefits of applying CMMIv1.1 to Public Administrations software process.

3. Take quick view of the current state of the Requirements Management process (RM) and the Subcontracting Management process (SAM) in Public Administrations using CMMI as the reference model.
4. Obtain recommendations about how to begin a process improvement initiative in Public Administrations.

In order to achieve the above objectives, the Symposium was organized in:

- Conferences.

  The current trends of the Process Improvement Models, particularly the CMMI suite developed by the SEI were shown. Also, CMMI implementation experiences were presented [5] [6] [7] [8]. These activities were done to achieve the objectives 1 and 2.

  In order to get a common language about the RM and SAM processes, some brief explanations were presented. This activity was done to achieve the objective 3.

- Focus Groups.

  A quick view of the RM and SAM processes in the Public Administrations using the CMMI as the reference model was obtained. RM and SAM processes were selected because they have a special interest for Public Administration.

  As mentioned earlier, two brief explanations were presented. Then, the RM and SAM processes were discussed in several and separated focus groups.

  Each focus group was formed with ten participants. A moderator managed the meeting and a note keeper took notes about discussions and wrote down the issues raised. A ground rule was not to accept in the same group two or more participants of the same Public Administration.

  The following activities were carried out:

  1. Gathering the current state of the PA processes.
  2. Identifying the potential benefits of implementing CMMI practices.
  3. Obtaining a list of short term recommendations.

  These activities were done to achieve the objectives 3 and 4.

- Workshops.

  Some workshops to help in the implementation of a process improvement were addressed by the symposium sponsors. In addition, several practical demos were showed to the participants.

  These activities were done to achieve the objectives 1 and 2.

Next, findings and proposals of short term actions of RM and SAM focus groups were presented.

Requirements Management Focus Groups

In order to establish a common language, as was indicated earlier, the moderator explained the Requirement Management Process concepts and practices of CMMIv1.1. Then, Public Administrations explained their current stated of RM process.

To facilitate the comprehension of the obtained findings, a brief explanation of each specific practice (SP) of the RM Process is described.

- SP1.1. Develop an understanding with the requirements providers on the meaning of the requirements.

  Description. A requirement provider could be an internal or external user, and it could be any other official source from which to receive requirements (customer, project manager, systems engineers, software suppliers, software engineering group, marketing, higher level management, etc.).

  Findings. There are multiple providers for the Public Administrations. It is true that most of the time these providers could be identified by their name; however the communication process is informal. This means that no documentation exits to identify all requirements providers. Usually, the requirements are communicated to the top level in a horizontal way.

- SP1.2. Obtain commitment to the requirements from the project participants.

Description. The previous specific practice dealt with reaching an understanding with the requirements providers, this specific practice ensures that project participants commit to the current approved requirements and the resulting changes in project plans and activities.

Findings. Almost all of the participants determined that they do not have defined processes for establishing agreements with the requirements providers. The commitments are made informally, mainly orally, through work meetings.

- SP1.3. Manage changes to the requirements as they evolve during the project.

  Description. During the project lifecycle requirements change for many reasons; nevertheless it is fundamental to manage these changes in an efficient way. Documenting the requirements changes and the rationale is the most important activity of the Requirement Management Process.

  Findings. On the one hand, almost all Public Administrations do not have a process for change management or in the best of cases they have a poor process because they do not document their changes. On the other hand, the participants found that it is too difficult to manage the changes because sometimes the requirements providers are not aware of the impact generated by the change.

- SP1.4. Maintain bidirectional traceability among the requirements, project plans and the work products, from their source to a lower level.

  Description. Requirement traceability is the ability to describe and follow the life of a requirement, in both a forwards and backwards direction. The traceability that covers both the horizontal and vertical relationships is called bidirectional traceability.

  Findings. On the one hand, a lot of the participants did not understand the "traceability" concept. This exposed the lack of knowledge of participants because it was the first time that they had heard the word. In this way, many Public Administrations did not perform traceability practices and they do not use a traceability matrix for their requirements. On the other hand, only some of the participants performed a change management process but poorly.

- SP1.5. Identify inconsistencies between the project plans and work products and the requirements.

  Description. This specific practice finds the inconsistencies between the requirements and the project plans and initiates the corrective action to fix them.

  Findings. It was determined that Public Administrations did not perform the revision of their projects plans, activities or work products for consistency with the requirements.

Finally, "process institutionalization" to ensure that the process will be documented, effective, repeatable and lasting is a long term objective. Public Administrations confirmed the use of one or two previous practices, but the institutionalization practices are a concept out of their hands in this moment.

Proposals of Short-term Actions on Requirements Management Process in Public Administrations

The participants of these focus groups identified the need of having an effective, repeatable and lasting Requirements Management Process in order to obtain reliable and controllable requirements. The following short-term actions were identified:

- Involve all the organizations in the process improvement project, mainly the Senior Management.
- Promote RM process training among the Public Administration personnel.
- Sensitize Public Administrations Senior Management with the importance of having an effective, repeatable and lasting RM process, and the benefits that this process brings to the software development process.
- Make the users understand the cost (in time and effort) that any change implies and the importance of having an adequate requirements definition process at the beginning of a development.
- Implement traceability techniques and promote the purchase of tools to make the implementation easy.
- Develop a list of the most common terms used in Requirements Engineering to avoid confusions.
- Develop a guideline of RM practices to obtain a successful process in future projects.

Conclusions about RM process in Public Administrations

Focus Groups allow an active participation of Public Administrations to achieve identify their current issues. Also, with the ideas and concepts expressed in the Conferences, each Public Administration took a quick look of their own current state of the RM process using the CMMIv1.1 as reference model. In this way, Public Administrations identified the gap that they had with respect to the model. Although it is true that each Public Administration use a poor RM process, this Symposium helped them to compare their process with the CMMI and then identify their deficiencies and what they had to do to improve their RM process. In addition, all Public Administrations understood the importance of having processes that allow them to repeat the successes in all their projects.

Subcontracting Management Focus Groups

The Acquisition Process is defined as the process of acquiring partially or totally the Information System (IS) Technologies from an external services supplier [10]. This means to delegate everything or part of the IT work through a contract with an external company that joins in the client organizational strategy and seeks to design a solution to existing software problems inside the latter. In the last years, the SAM process of IT functions has been gained the attention of many researchers and industries.

In order to establish a common language, as was indicated earlier the moderator explained the Subcontrating Management Process concepts and practices of CMMIv1.1. Then, Public Administrations presented the current state of SAM process.

To facilitate the comprehension of the obtained findings, a brief explanation of the specific practices grouped by Specific Goals (SG) of the SAM Process was described.

- SG1. Establish supplier agreements.

  Description. Establish agreements with the suppliers by a formal contract. Also, it is necessary to determine the product to acquire, and consequently identify and select the potential suppliers.

  Findings. Public Administrations usually perform a call for proposals to subcontract their projects. However, these types of projects exceed budget the most of the time. Also, this results in a loss of project control and lack of communication.

  One of the main issues was the poor knowledge that Administrations have about the products or services to be contracted. Also, the supplier has a poor knowledge of the services o products to be offered. With this context in mind, a loss of negotiation capacity is produced and it is necessary to have and independent supervision.

  Another issue appears in the supplier selection activity because the Regional Government is in charge of selecting the providers and does not consider the selection criteria and requirements of the Local Administrations.

- SG2. Satisfy supplier agreements.

  Description. Monitor the supplier agreements for each project.

  Findings. Public Administrations have a subcontracting process, efficient or not, but they do not have a control process for managing the subcontracting project. This loss of control can be due to the lack of knowledge about subcontracting standards and models within the Public Administrations.

Proposals of Short-term Actions on Acquisition Management Process in Public Administrations

The participants of these focus groups expressed the following short-term actions:

- Subcontract in a rational way, maintaining the strategy and project functional analysis.
- In depth knowledge of the product that Public Administration wants to acquire: “If I do not have the knowledge, I do not know what we want to subcontract”.
- Never forget that the subcontracting process does not avoid the work, but generates a new role: the control and monitoring of the acquisition process.
- Establish clear objectives between the Public Administration and the provider.
- Use subcontracting strategies and establish processes to control the required service.
- Subcontract small projects because they are easier to control.

Conclusions about SAM Process in Public Administrations

Public Administrations have certain mechanisms that manage the SAM process. Usually they plan the project without metrics and they do not have monitoring processes.

The lack of project control can be due to the lack of knowledge about subcontracting standards and procedures in the Public Administrations.

Public Administrations use a model of effective practices, like Spanish methodology called METRICA3, to cover partially the SAM process. However their implemented processes are not very efficient although it is not very difficult to align them with CMMI practices.

Conclusions

This first Symposium represents the first stage to initiate an improvement program, taking into account the poor knowledge and initial skepticism. A discussion forum to study and debate how Public Administrations could improve their current states was made. Also, recent studies on improvements models and their applicability in Public Administrations around the world were presented.

With this “quick look” at Public Administration, both in Requirements Management and Software Acquisition Management processes specifically, the lack of control in these processes was expressed. This issue will promote Public Administration’s initiatives to begin an improvement program.

Initial Symposium objectives were accomplished. Firstly, an initial quick look at RM and SAM processes of all Public Administrations was obtained and secondly, the idea that it is possible to improve and obtain the leadership in Public Administrations was promoted among all the Symposium participants.

This Symposium was the first step to begin an improvement program. Now it is the turn of Public Administrations. Public Administrations should begin the formal assessment of their processes to identify the strengths and weaknesses, and prioritize their improvements actions.

Acknowledgements

This work is sponsored by Endesa, DMR Consulting Foundation and SUN Microsystems through the “Research Group of Software Process Improvement in Latin America”

Bibliography

[1] Goldenson, D.R., Gibson, D.L., "Demonstrating the Impact and Benefits of CMMI: An Update and Preliminary Results", CMU/SEI-2003-SR-009, Software Engineering Institute, Carnegie Mellon University, Pittsburegh, PA, 2003.

[2] Paulk, M.C., Curtis, B., Chrissis, M.B., and Weber, C.V., "Capability Maturity Model for Software, Version 1.1", CMU/SEI-93-TR-024, Software Engineering Institute, Carnegie Mellon University, 1993.

[3] CMMI Product Team, "CMMI for Development, Version 1.2", CMU/SEI-2006-TR-008, Software Engineering Institute, Carnegie Mellon University, 2006.

[4] Product Development Team, "Capability Maturity Model Integration (CMMI), Version 1.1, Continuous Representation", CMU/SEI-2002-TR-011, Software Engineering Institute, Carnegie Mellon University, 2002.

[5] Product Development Team, "Capability Maturity Model Integration (CMMI), Version 1.1, Staged Representation", CMU/SEI-2002-TR-012, Software Engineering Institute, Carnegie Mellon University, 2002.

[6] Cuevas, G., "Training on Software Process Improvement Models across University", 1st Symposium on Improvement Process Models and Software Quality of Public Administrations. San Lorenzo del Escorial. Madrid, Spain. November, 2004.

[7] Jordán, A., "SEI's Maturity Models Motivation and Evolution", 1st Symposium on Improvement Process Models and Software Quality of Public Administrations. San Lorenzo del Escorial. Madrid, Spain. November, 2004.

[8] Sivaramakrishnan, G. R., "Software Development in India, Why, For what, and How apply Software Process Models", 1st Symposium on Improvement Process Models and Software Quality of Public Administrations. San Lorenzo del Escorial. Madrid, Spain. November, 2004.

[9] Yard, J., "Maturity Models Implementation in Public Administration", 1st Symposium on Improvement Process Models and Software Quality of Public Administrations. San Lorenzo del Escorial. Madrid, Spain. November, 2004.

[10] Lee, J.-N., Huynh, M., Chi-Wai, R., and Pi, S.-M., "The Evolution of Outsourcing Research: What is the Next Issue," *presented at 33rd Hawaii International Conference on System Sciences*, 2003.

Authors' Information

Jose A. Calvo-Manzano – e-mail: jacalvo@fi.upm.es
Gonzalo Cuevas – e-mail: gcuevas@fi.upm.es
Ivan Garcia – e-mail: igarcia@zipi.fi.upm.es
Tomas San Feliu – e-mail: tsanfe@fi.upm.es
Ariel Serrano – e-mail: aserrano@zipi.fi.upm.es

Universidad Politecnica de Madrid – Facultad de Informatica, Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid, Spain.

Magdalena Arcilla - Universidad Nacional de Educacion a Distancia – Escuela Tecnica Superior de Ingenieria Informatica, C/ Juan del Rosal 16, 28040 Madrid, Spain; e-mail: marcilla@issi.uned.es

Fernando Arboledas – e-mail: fernando.arboledas@madrid.org
Fernando Ruiz de Ojeda – e-mail: frdo2@madrid.org

Informatica y Comunicaciones de la Comunidad de Madrid (ICM), C/ Embajadores 181, 28045 Madrid, Spain.

A PRACTICAL CASE OF SOFTWARE LOCALIZATION AFTER SYSTEM DEVELOPMENT[1]

Jesus Cardenosa, Carolina Gallardo, Alvaro Martin

Abstract: Internationalization of software as a previous step for localization is usually taken into account during early phases of the life-cycle of software development. However, the need to adapt software applications into different languages and cultural settings can appear once the application is finished and even in the market. In these cases, software localization implies a high cost of time and resources. This paper shows a real case of a existent software application, designed and developed without taking into account future necessities of localization, whose architecture and source code were modified to include the possibility of straightforward adaptation into new languages. The use of standard languages and advanced programming languages has permitted the authors to adapt the software in a simple and straightforward mode.

Keywords: Localization, Internationalization, XML.

ACM Classification Keywords: D. Software, D.2.7 Distribution, Maintenance and Enhancement

Introduction

Any technical device devoid of human interaction operates and yields an expected level of productivity regardless of the cultural environment where it is located. The same can be said for software, as long as it does not call for any human interaction. However, many software applications require human interaction for a correct functioning. In this case, the level of productivity of the software will depend not only on software's intrinsic technical characteristics but on external human factors.

When a software application is used in a context with a different cultural environment (like different mother language, different icons, symbols, etc.) from its original one, a process of adaptation into the new work culture is required. This process is known as localization. The adaptation into a new culture not only comprises evident factors like the language of the interface and messages to the user, measure units or data formats (also known as overt factors according to [Mahemoff et al, 1998]); but also other slippery and fuzzy issues that finally

[1] This paper has been sponsored by Universidad Politécnica de Madrid and Autonomous Community of Madrid through the project EXCOM (R05/11070).

distinguish a culture, like mental disposition, perception of the world, rules of social interaction, religion, etc., which are referred to as the covert factors of a culture. More specifically, the process of localization consists on the "adaptation of a product, application or document content to meet the language, cultural and other requirements of a specific target market (a *locale*)", as expressed by the W3C [W3C, 2005].

On the other hand, internationalization refers to the design and development of a product, application or document content that enables easy localization for target audiences that vary in work culture, region, or language. In this sense, it can be said that internationalization precedes and facilitates the task of localization.

Besides, the processes involved in localization of software applications changes significantly depending on whether it is done over a pre-existent application or over a developed application.

The next section sketches the most frequent practices of software internationalization and localization in software design. However, in pre-existent applications, and depending on the system development methodology, the localization process can become very expensive in terms of time and resources. We will show how we internationalized and subsequently localized a pre-existent application in a cheap and quick manner, by means of advanced standard implementations languages like Visual .Net and XML.

Software Architectures for Internationalization and Localization

As we commented in the previous section, the internationalization and localization (I&L) processes deal with more that mere language issues. However and for the purposes of this paper, we will consider only the language adaptation, which is the most prominent and visible aspect of I&L.

Apparently, an internationalized product does not entail structural changes in order to adopt a new language. Internationalization consists on abstracting the functionality of a product of any given language, in a way that the support of the information of the new language can be added afterwards, without facing the source code (dependent of a given language) when the product is localized into a new language. Currently, main development platforms offer support and tools to facilitate the internationalization of over factors of applications [Hogan, 2004], [Huang et al, 2001], in a way that currently problematic questions are centered on the optimization of the internationalization processes within the life cycle of the application.

There are three main approaches for internationalizing an application. The first one is the system where messages, menus and other culture-sensitive factors are embedded in the source code of the application. This approach obliges to develop a different version of the system for each of the target cultural environments. Each version requires independent process of testing, maintenance and upgrading, multiplying the costs of localization.

The second approach consists on extracting messages to the user of a given application into an external library. The application is generated from a common source code that links to the culture-sensitive libraries. Although this architecture resorts on a unique source code, only the languages contained in the external library could be incorporated, and it is required to test and maintained each of the supported languages individually.

The third and last approach consists on an architecture composed of the core of the application comprising all the functionalities but independent of cultural factors, which dynamically access to files of external resources that contain information about the corresponding culture (localization packages). The difference with the previous approach lies in the fact that the culture-independent code dynamically calls to the information of culture, so that only one executable must be tested and maintained. Once the set of supported cultures is tested, the addition of new cultures does not imply modifications. From this general idea, each author develops his/her own way of acting. For example, [Stearns, 2002] describes the process of developing systems sensitive to cultures using JAVA and XML for resources files, whereas the environment GNU/Linux [Tykhomyrov, 2002] and the Free Software Foundation [FSF, 2002] prefer the use of special libraries that facilitates the extraction of the localizable contents of the application and the construction of localization packages.

Regarding the aspects related to the life cycle of the internationalized software, [Mahemoff et al, 1999] presents a methodology for requirements specification to develop culture-sensitive systems. On the other hand, [Huang, 2001] offers a description of the processes to be followed to create culture-sensitive software, emphasizing the fact that the internationalization tasks should be included in the corresponding phases of the life cycle of software.

The work on the area of localization is complemented with research on the problem of localizing software already internationalized. Even when the technical procedure for software internationalization is optimized, the bottleneck lies in the localization processes of a product. The process of internationalized software localization resorts on

the concept of repository and reuse of translation resources. That is, apart from the external file that contains the messages to the user and its translations, there is a repository where translations are stored for their subsequent reuse. In some cases, there are also repositories for terminology.

The following standards have been established to facilitate the task of managing the culture-sensitive resources files and their communication with repositories:

- XLIFF (XML Localization Interchange File Format) defines a standard format for resources files that stores the translated strings, in a way that tools for assisted machine translation can be developed independent of the application to be localized, as well as transporting the translation information from one phase of the process to the following phase [OASIS, 2003].
- TMX (Translation Memory Exchange), allows for the storage and interchange of translation memories obtained after the use of automatic tools for translation [LISA, 2005].
- TBX (Term Base Exchange), defines a standardized model for terminological databases [LISA, 2003].

There are also some common practices among companies that have become a "*de facto"* standard [Hogan, 2004] aiming at minimizing the impact of localization on commercial software products, namely:

- Extraction of the fragment of texts used in the user interfaces into resources files.
- Control of the extracted texts, contexts and their translations.
- Outsourcing of the translation tasks to specialized companies.
- Simplification of the contents of the chains and their contents as a previous step to the sending to translation centres.

However, as can be seen, internationalization architectures and localization standards do not offer a solution for already existent applications that require international dissemination. That is, according to these architectures, localization is a bottleneck and it is only possible with an internationalized architecture. But what happens if we want to adapt a software application into many languages? The next section presents how an architecture can be changed in an afterwards-mode and how we internationalized and subsequently localized a pre-existent application in a cheap and quick manner, by means of advanced standard implementations languages like Visual.Net and XML.

Internationalizing an Existent Application: the Context

The starting point of this work is a software application for multilingual generation that allows for human interaction. It is an interactive application composed of a user interface where the user can manipulate semantic representations of the text to be translated.

The only requirement in the development of this tool was the use of UNICODE files, because of the almost certainty that the tool was going to be used for analysis and generation in a variety of languages. This obviously involved the future necessity of localization of the tool. It seems clear that the internationalization should be foreseen and reflected at the level of requirements specification and that it consists on something more than the mere use of UNICODE files. We will show how, in cases where internationalization has not been taken into account in the development processes, a pre-existent system can be adapted a posteriori for internationalization purposes. That is, our work is framed in the following context:

- There is a need of a future internationalization and localization processes, which is partly reflected on the requisites through the need to work with UNICODE files.
- The system is implemented in a development framework compatible with the use of UNICODE files (VB.NET), which guarantees the strict observation of the previous requisite, but nothing else.
- Apart from UNICODE files, there is not any other feature in the system oriented towards internationalization and subsequent localization.

The result of this situation is an environment able to import and deals with UNICODE files, that works with several languages (it is a translation aid tool) but with the totality of the user interfaces functionalities in just one language (in this case, Spanish).

The internationalization process that we are going to describe has been carried out after the complete development of the tool, proving that at least on of the most important and basic tasks of localization, such as language adaptation, can be done even without having internationalized the system in previous development phases.

Description and Preliminary Analysis of the Pre-existing System

The application is conceived as an environment for linguistic tasks, in which some external resources and components as language analyzers, language generators and dictionaries are integrated, with a powerful user interface and graphics management. From the architectural point of view, there are three main subsystems in the environment, which are:

- Kernel: it is the component in charge of managing most of the information and data flow of the application, as well as integration with external language analyzers, generators and dictionaries.
- Graphic controller: this component is in charge of managing the graphical display of abstract and semantic structures, as well as the correspondence between semantic structures and graphics.
- Interface: this component manages the communication of the application with the user. It mainly consists on the user interface with a few functionalities, which are delegated to the kernel or the graphic controller.

Figure 1 shows the application architecture and information flow graphically.

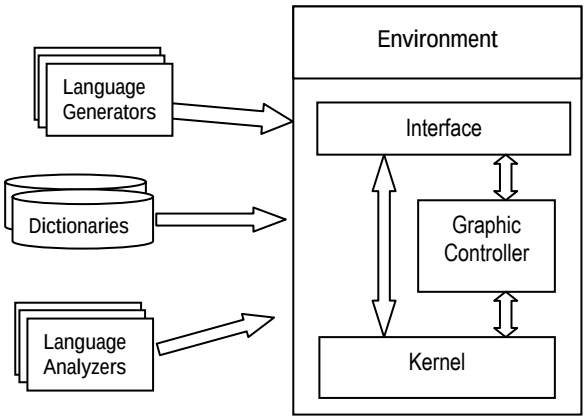


Figure 1. Architecture of the application

The entire interface is in Spanish. It is important to note that there are two types of textual elements in the application interface: "message errors" occurring in unexpected situations (also called emerging messages) and the text of the environment itself.

Each subsystem can generate a given number of emerging messages and windows with their corresponding text elements. In this way, the textual elements are scattered all over the source code.

A preliminary analysis of the source code shows that the textual elements follow two regular patterns. The first pattern corresponds to "emerging messages". These are created with the statement "*msgbox ()*" (an abbreviation for *message box*); the text assigned to the emerging message is written within the parenthesis. As an illustration, a real emerging message informing of a file that is not found will have the following code:

```
msgbox("Fichero no encontrado")
```

The second pattern corresponds to the text used in windows and buttons of the application, which have the general pattern:

```
component.text = "Text associated to this component"
```

Where the expression "component.text" is the convention in VB.NET to note that the string in double quotes is the text that will appear on that specific component. For example, to assign the text "Aceptar" (OK) to a given button, we write:

```
button.text ="Aceptar"
```

Since we are going to restrict the I&L process to just linguistic issues, these textual elements will be the subject of the I&L processes.

Strategy for I&L, Conceptual and Architectural Design

Our specific problem is the need to adapt the environment into the English language. The most obvious and even quick solution is to search for all the text elements in Spanish and create a new version of the application with the interface in English. However, there are some requirements on the I&L adaptation, such as:

a) The localization process should be done by translators / final users.
b) Maintainability of the system and translations should be guaranteed.
c) It is desirable to produce a core application abstracted from the linguistic issues.
d) The pre-existing components must not be functionally modified.

Therefore, the architecture should be modified with the addition of a new component in charge of the internationalization functionalities; so that the textual contents of new languages are stored as a new resource (in the form of an external file, for example) which can be read and processed by the application itself.

The new component is in charge of reading the external files with the translations of the textual elements and imports them into the environment so that messages and interfaces can be shown in different languages. The result is a new software architecture as illustrated in figure 2.

Thus, the global strategy promotes the creation of a new specific component that once integrated in the original architecture is responsible for all the internationalization tasks. The basic functionalities of this new component should be:

1. Identification and labelling of all the text strings written in Spanish language of any kind (emerging messages, buttons, windows, and any other textual elements)
2. Extraction of these strings and generation of a XML file according to a predefined structure
3. Capture of the new XML file, once all the identified strings have been translated into the new language in the XML file.
4. Insertion of the translated strings according to the labelling.

The detailed description of this process is shown in the next section.

The Practical Case

The new component, called "Internationalization Manager", serves a number of functions that guarantee that the required language changes are carried out over the existent environment, while intervening in the current software as less as possible. Figure 2 shows the new architecture of the environment and how the "Internationalization Manager", together with its functional element the Text Management Module (TMM), is integrated in this new architecture. In the remaining, we will describe how the new component works and its main functionalities.

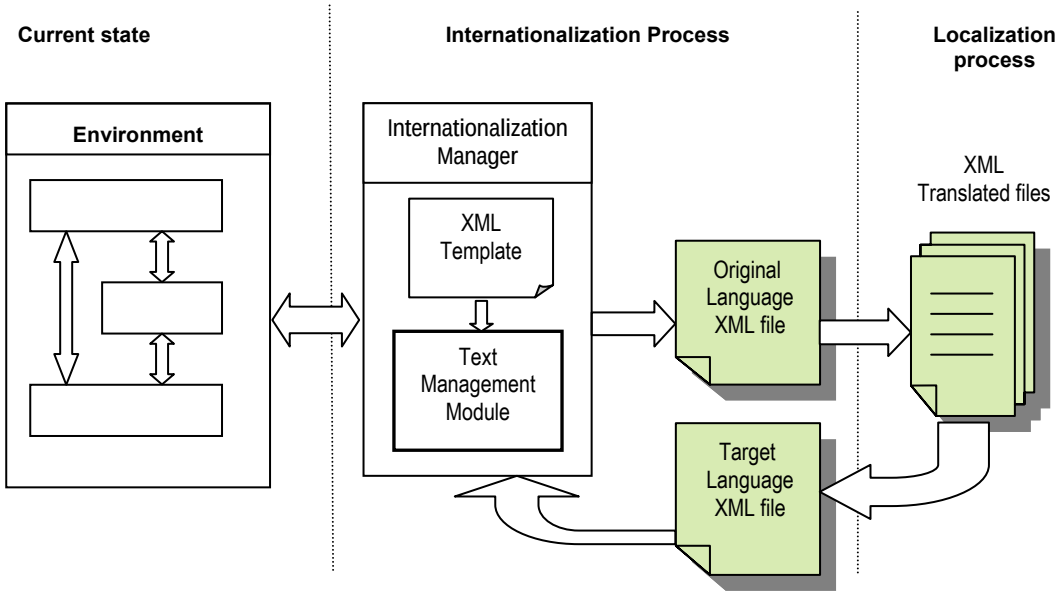


Figure 2. Global process

1. Text string identification and labelling.

This first functionality consists on identifying the textual elements in the original language (in our case, Spanish) following the two aforementioned search patterns, namely `msgbox` and the `component.text`. This function has been carried out by means of a script that identifies these text strings. The result of the script is a file where not only the text string is stored, but also additional information associated to the string, like its location, the component it belongs to, and other information that could be useful. All the information that the script gathers about a text string is labelled with a numeric identifier. The only modification that is done from this moment over

the original software is the substitution of these strings by a function that calls for the identifier in the XML file of the required language and inserts the text string contained in the XML file. Let's see an example of how it works.

Suppose the source code of the application in Spanish contains the following an emerging message:

`msgbox ("Error: Archivo no encontrado")` (English: "Error: *File not found*")

The Spanish text string is substituted by the following:

`msgbox (`**`InternationalizationManager.GetText(57)`**`)`

Where `InternationalizationManager` is the function that calls to the corresponding component of the TMM that executes the instruction `GetText(57)`. This instruction captures and temporally inserts the text string labelled with the identifier `57` in the language selected by the user in its place. Currently there are not text strings of a specific language in the environment anymore but functions like the aforementioned that allow for the incorporation of a new language in the environment without further changes over the original software.

2. XML structure

The information about the text should be structured according to an XML template that permits to save a unique structure but modifiable in the data (in our case the text translations) that guarantees their interchangeability and maintainability. This XML file can be imported by the environment since the programming language (VB.net) is provided with an XML parser. This XML file is delivered to the translators and looks like as shown in figure 3.

The first line of the XML file indicates the version of the XML standard being used and the type of codification of the file (UNICODE in this case). The second line has an empty attribute `langID=""` that will indicate the target language of the translation of the strings. The rest of the XML file is divided in three elements `<userInterface>`, `<kernel>` and `<graphicController>`, each pertaining to the main components of the software. Each component is composed by a number of `<item>`. An `<item>` stores the following elements:

- The attribute `"id"` (in the example of figure 3, one "id" is 56) that uniquely identifies the linguistic text element and its presence in the software component.
- The `"orig"` attribute corresponds to the text string in the original language. One example is the Spanish string *"¿Desea continuar?"*.
- The `<translation>` element which is empty and will have to be filled with the translations into the target language.

This file is distributed to translators so that they can perform the translations tasks in their corresponding working places, allowing for an absolute independence of the translation process and its integration in the software environment. The XML files in the target languages are delivered to the TMM and located in the corresponding directory so that they can serve as the different language options of the environment to be selected by the user. An example of an XML file containing the translations for English is shown in figure 4. This file is the result of the localization process.

```
<?xml version="1.0" encoding="UTF-8"?>
 <localisation langID="">
  <userInterface>
    <item id="56" orig="¿Desea continuar?">
       <translation> </translation>
    </item>
       <item id="57" orig="Error: Archivo no encontrado">
       <translation> </translation>
    </item>   ...
  </userInterface>
  <kernel>
    </item>
       <item id="64" orig="Atributo no válido">
       <translation> </translation>
    </item>
   ...
  </kernel>
  <graphicsController> … </graphicsController>
 </localisation>
```

Figure 3. Original Language XML file

```
<?xml version="1.0" encoding="UTF-8"?>
 <localisation langID="English">
  <userInterface>
    <item id="56" orig="¿Desea continuar?">
       <translation>Do you want to continue?</translation>
    </item>
       <item id="57" orig="Error: Archivo no encontrado">
       <translation>Error: File not found </translation>
    </item>
   ...
  </userInterface>
  <kernel> ... </kernel>
  <graphicsController> … </graphicsController>
                          </localisation>
```

Figure 4. English Language XML file

Finally, the component "Internationalization Manager" is in charge of detecting XML files in the available languages and thus it offers them as options to the user of the environment. Once the user has select a language, the application dynamically imports the XML file that contains the text strings translated into the selected language and shows the environment in that language.

Conclusion

We have presented three approaches for software internationalization and subsequent localization. We have seen how the use of current programming languages which incorporate XML parsers allows the development of the third strategy, which the one that produces more flexible, adaptable and maintainable applications, in a convenient and easy and straightforward manner with a relatively low cost.

This approach also permits that the work of the developers can be initially kept apart from the linguistic questions and permits to maintain a single version of software. Major changes on the original software can be dealt with in the same way even if there appear new items.

Bibliography

[FSF, 2002] [10] Free Software Foundation (2002), "Online GNU gettext manual" http://www.gnu.org/software/gettext/manual/gettext.html. Visited: March 2006

[Hogan, 2004] [5] Hogan M. J., Ho-Stuart C. & Pham B. (2004), "Key challenges in software internationalisation"

[Huang et al, 2001] [6] Huang E., Hsu J. & Trainor H. (2001), "Unicode enabling for software internationalization" www.symbio-group.com/doc/Symbio%20Whitepaper%20on%20Unicode%20Enabling.pdf. Visited: March 2006

[Huang, 2000] [16] Huang E., Haft R., & Hsu J. (2000), "Developing a Roadmap for Software Internationalization" www.symbio-group.com/doc/Developing%20a%20Roadmap%20for%20Software%20Internationalization.pdf

[LISA, 2002] [13] LISA (2002), "TBX Specification" http://www.lisa.org/stantdards/tbx. Visited: March 2006

[Lisa, 2004] The Localization Industry Standards Association (2004), "Lisa Industry Primer 2nd Edition".

[LISA, 2005] [12] LISA (2005), "TMX Specification" http://www.lisa.org/standards/tmx/tmx.html. Visited: March, 2006

[Mahemoff et al, 1998] [3] Mahemoff M. J. & Johnston L. J. (1998), "Software Internationalisation: Implications for Requirements Engineering"

[Manemoff et al, 1999] [15] Mahemoff M. J. & Johnston L. J. (1999), "The Planet pattern language for software internationalisation"

[OASIS, 2003] [11] OASIS (2003), "XLIFF 1.1 Specification" http://www.oasis-open.org/comitees/xliff/documents/xliff-specifications.htm. Visited: March 2006

[Stearns, 2002] [4] Stearns B. et al (2002), "e-business Globalization Solution Design Guide: Getting Started" http://www.redbooks.ibm.com/abstracts/sg246851.html?Open. Visited: March 2006

[Tykhomyrov, 2002] [9] Tykhomyrov O. Y. (2002) "Introduction to internationalization programming" The Linux Journal, December, 2002 http://www.linuxjournal.com/article/6176 Visited: March 2006

[W3C, 2005] [2] W3C, "Localization vs Internationalization".http://www.w3c.org/International/questions/qa-i18n. 2005

[Yeo, 2001] [8] Yeo A. W (2001), "Global-software development lifecycle: an exploratory study" Conference on Human Factors in Computing Systems, 2001

Authors' Information

Jesús Cardeñosa – Department of Artificial Intelligence; Universidad Politécnica de Madrid; Madrid 28660, Spain; e-mail: carde@opera.dia.fi.upm.es

Carolina Gallardo – Department of Artificial Intelligence; Universidad Politécnica de Madrid; Madrid 28660, Spain; e-mail: carolina@opera.dia.fi.upm.es

Álvaro Martín – Department of Artificial Intelligence; Universidad Politécnica de Madrid; Madrid 28660, Spain; e-mail: martin@opera.dia.fi.upm.es

EXPERIENCES ON APPLICATION GENERATORS

Hector Garcia, Carlos del Cuvillo, Diego Perez, Borja Lazaro

Abstract: The National Institute for Statistics is the organism responsible for acquiring economical data for governmental statistics purposes. Lisbon agreements establish a framework in which this acquisition process shall be available through Internet, so each survey should be considered as a little software project to be developed and maintained. Considering the great amount of different surveys and all changes produced per year on each make impossible this task. An application generator has been developed to automate this task, taking as a start point the Word or PDF template of a survey, and going through a graphical form designer as all human effort, all HTML, Java classes and Oracle database resources are generated and sent from backoffice to frontoffice servers, reducing the team to carry out the whole set of electronic surveys to two people from non I.T. staff.

Keywords: Software automation, application generators, CASE.

ACM Classification keywords: D2.2. Design Tools and Techniques

Introduction

Complaining Lisbon agreements concerning e-Government, the Spanish National Institute for Statistics (INE) tackles the problem of translating all economical surveys from paper format into web applications. There exist hundreds of different forms, and for a particular survey, more than one version depending on the kind of target organization, so the required effort to create all infrastructure exceeds not only the capacity of I.T. Department, but the budget to carry out the gigantic task. A previous successful experience on metadata processing from INE and the pilot projects on Java application generation from Technical University of Madrid seem a proper combination to afford the trouble.

The idea consists of taking as a start point the current survey forms in Microsoft Word or PDF format, translating these into a tag based format appropriate for both browser representation and automated processing. This creates some kind of template used as a background for the application. Then a user may define the web form over the background painting components using a designer, and establishes properties for the components from those pre-defined in the designer. Finally only translating these definitions into source code is still to be done.

The technology of generated code shall meet the following requirements:

- HTML 4.01, later substituted by XHTML 1.1 by the research team at UPM, for the web user interfaces.
- XForms 1.0, for the definition of validation rules, with the premise to deploy complete surveys in XForms for future use.
- Java servlets, based on action struts architecture and their corresponding beans.
- Hibernate 3 as database connection tier.
- PDF format as receipt of the answered surveys.

The Architecture

The main goal of the project, beyond any other, was to decrease sensitively the staff, effort and time to market for each survey application, and so, of the whole set of applications. The lack of I.T. professionals in the department in charge of the project also conditions the profile of the target user of the generator.

Four modules were found to be the core of the survey generator:

Format translation tool

As long as the forms corresponding to the different surveys are being created in other departments, the format and composition developed for hard copies is not valid for automated processing, some tool to extract the contents from Word and PDF files and export them into tagged files, closer to web requirements and much more appropriated for processing.

For this purpose several options were evaluated. At first the best choice seemed to develop a specific translator, in Java language. The wide support for PDF processing available supposed a great advantage, but by that time the number of API or information about accessing Word files, especially about the structure of these files, was very poor. Only some arising APIs, such as Apache's POI, were available, so finally the decision went on a third party product, and then develop only the integration to the system.

Survey definition tool: the editor

Once the source document has been translated in a processable format, and before proceeding to its publishing, the system allows users to define the forms for the survey. Of course, there is not enough information in the templates, but visual aspect. In this sense, the captured survey is shown to the user as background in a screen, in which he may add or remove components that will later compose a web form.

For each field in the survey the editor allows to define some specific features, such as data type, length, etc. In case of combo boxes or lists, it is possible to define the valid values list. Also constraints have been implemented, such as date formats, decimal and thousands separator, ranges, allowed data sets, etc.

The components available while designing forms are the following:

Label: consists of a read only text. The user may define component name, text, font type, size and color, bold and italics style and background color.

Text: it is a read/write field containing characters, and it is possible to configure component name, length, maximum capacity, data type (string, date, time, year, day, month, float, double, integer, positive, negative, long) for validation rules to be applied. If the validation finds that the content of the field does not match the data type, the user shall be advised, so the application allows the user to set an error message to be displayed in a dialog. It is possible to apply some modifiers to the text fields, such as mandatory, read only, hidden and calculate value automatically. In case of selecting calculated field, a calculator is shown to define the formula. A formula may contain both values or fields in the same survey. There is also a description of the field to be shown as a hint. For some specific surveys, with repetitive contents, such as tables with a row per city or state, it is necessary to associate the field with a column in the table where data is to be stored.

Text from database: it is a read/write field associated to a column in a table from a database. The purpose is to set a default value for the field when the form is loaded. The field is processed as a text field once the form is submitted. It is very useful when defining, for instance, headers in a form, with the name, address and essential data for the final user. Possible configurations are identical to text fields.

Text area: consists of a text field with several rows. The user may define the size in terms of rows and columns, and data type, read only and mandatory options are available. For this component the validation rules may be disabled.

Radio, check boxes and lists: a new feature is added regarding the previous components: an interface to manage the choices and value for each choice in the available options, and establish a default selection if any.

Figure1.- Toolbar

Buttons: it is possible to draw buttons and associate a number of actions to them. Currently a button can save the form, add or delete items to repetitive contents (such as tables), validate the form and generate the errors list, or open the help page. Also specific separate components are available to create reset and submit buttons.

Of course a survey can be defined in several working sessions, so the functionalities for saving a survey in the current definition state, open a saved survey, and delete are available, as well as preview of the current survey. Finally, in the bottom of the toolbar, generate survey and exit commands are located.

If the user has already defined other surveys, the system enables the import functionality, so that the components from other surveys can be brought. This feature is especially useful to create surveys that are simple modifications from the previous year format, or to create the same survey in a different language, so only some modifications on descriptions and hints are needed.

Most surveys are so long that page breaks are needed. The editor allows defining these line breaks, which also are considered when generating the PDF receipts.

Application generator

The application generator retrieves the form already defined with the previous tool, starting a analysis process, in which the definition data are separated into modules, and afterwards the different application components are generated as a set of XHTML pages that are a composition starting with the original HTML from the PDF or Word, adding the information defined in the editor, and some GEN specific attributes for some tags that allow further processing. These attributes are used mainly to show the specific user information (i.e. tags with organization information, or the form with values that have been saved in a previous session).

Then the XForm files needed to validate the data in the forms are built to be used in web and bulk load processes, plus the resources for XHTML, such as background images, help files, etc.

All these infrastructure is stored in the database, and when a request to fill in a form is produced a set of servlets, action struts and filters that parse the XHTML page to be showed, together with the required resources, get the result.

To generate PDF receipts from the submitted forms XSL:fo is used. As long as the forms are, after the process, HTML pages, a previous transform has to be done. Fo processes basically XML documents, and HTML does not provide the closing tags, for instance
, so the document is analyzed and translated into XHTML before processing. The PDF shall have same number of pages as defined in the survey using the page breaks.

Publishing tool

The first approach to publish all the resources generated by GEN was to send to the production server all classes, JSPs, servlets, images, etc. It is a remote server in a different network, and should be configured for allowing the access to file structures, execute compilations, modify Tomcat settings, … at the same time that the network devices where also reconfigured. Of course this is not an affordable task, obviously a software tool of this kind cannot cause all this changes.

Finally, this led us to refactor all design, and instead generating all the mentioned resources, two meta servlets where produced, so that one of them is capable to draw any form and the other one is capable to receive (after submitting) any kind of form. This reduces sensibly the publishing process, now only access to the database is required, and all resources are stored there. The database is accessed through the corporate intranet.

Case Study

Probably the best way to describe how GEN works and the simplicity of the process to generate the new application to manage a survey is to follow an example.

First step: translating a PDF survey into HTML

The first of all, we need the file containing the survey designed by the corresponding department. Usually surveys have been translated into PDF, anyway this is an immediate process, usually covered by Adobe Acrobat Writer or Distiller, as well as by any tool complaining the standard.

This translation is based on an external tool, so we do not mind how the translation is done, the only important issue is the quality of the HTML we get after the translation. GEN only needs as input a HTML complaining the W3C standard. We always avoid HTML resulting from Microsoft Word exports, FrontPage, etc.

In this case study, we are using a translation tool which converts PDF into an HTML composed of layers. Each component of the document is set into a different layer, and the absolute position for each layer is written in the code. This is very useful for future processing of the template, while painting the components in the proper location over their relative texts and tables.

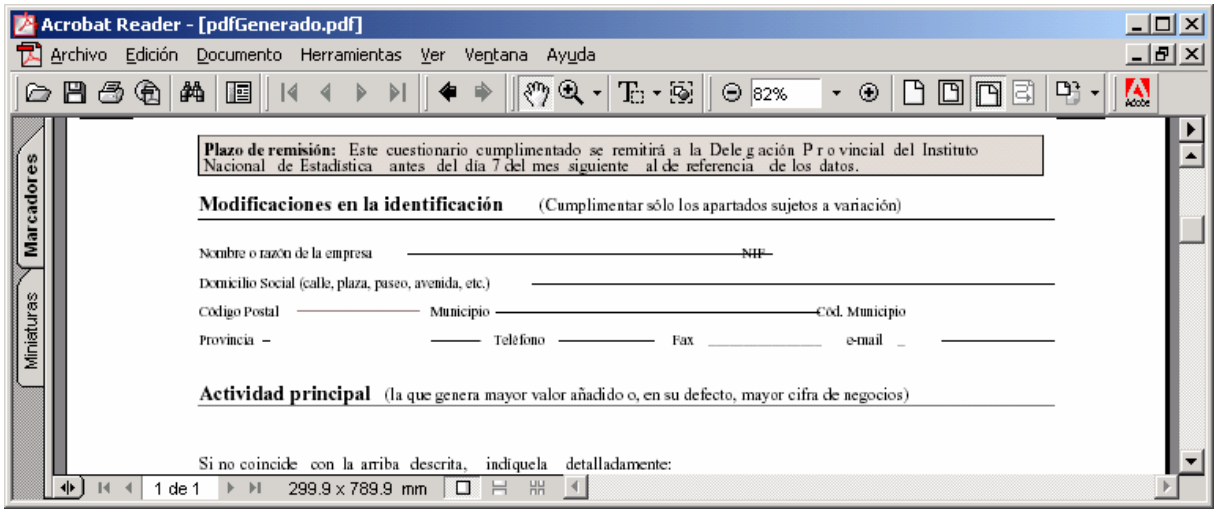


Figure 2.- The original PDF survey

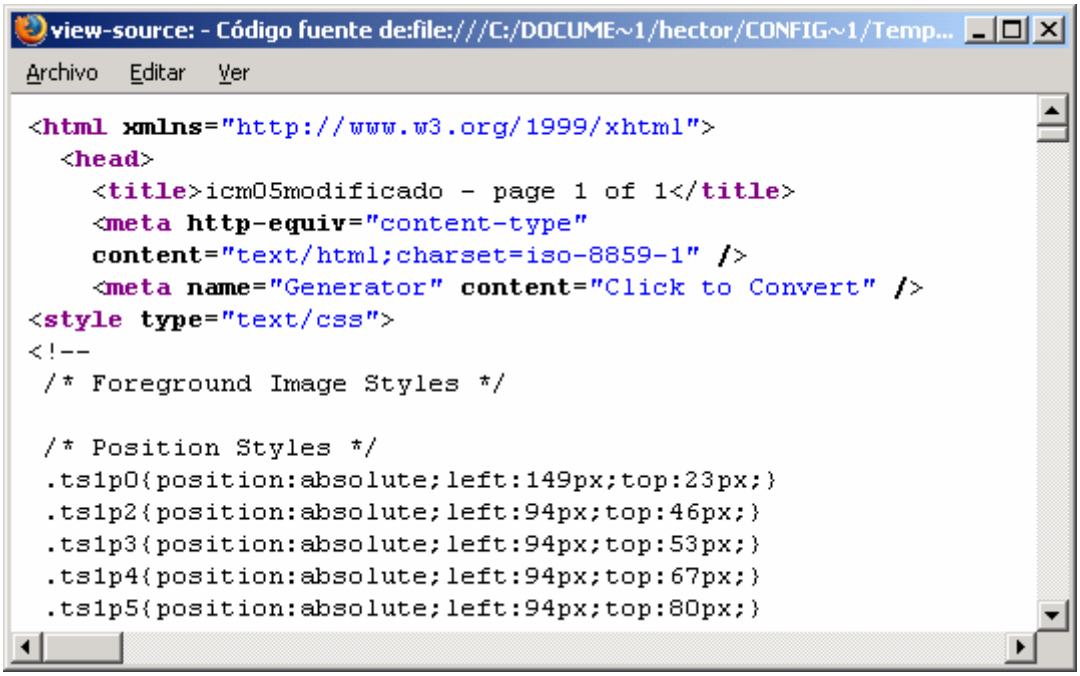


Figure 3.- The HTML obtained from the PDF showing the layer positions

Second step: design the form for a survey

Now GEN user shall define a form to capture the data corresponding to the survey. The HTML from the previous step is considered as a simple background to help the user drawing all components. This is the main task while using the editor, but also some other functions are to be completed: buttons to send, save or step through the forms, page breaks for viewing and generating PDF receipts, or dependencies and constraints specification.

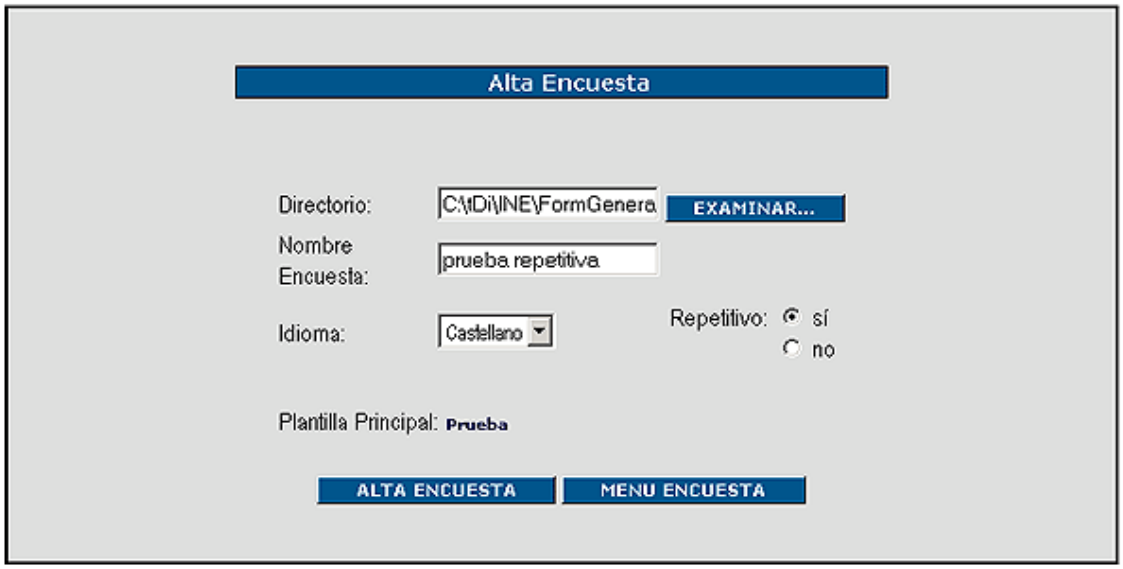


Figure 4.- Creating a new survey

First of all the typical screen where user is going to define the working directory, survey name and type. GEN allows to define multilingual surveys, so that the language specification is also important for further publishing. Each survey may be also converted in a template that may be used to create compositions in which a survey is composed of modules that are really survey templates. This is very useful to define headers or footers, or to include forms common to many applications.

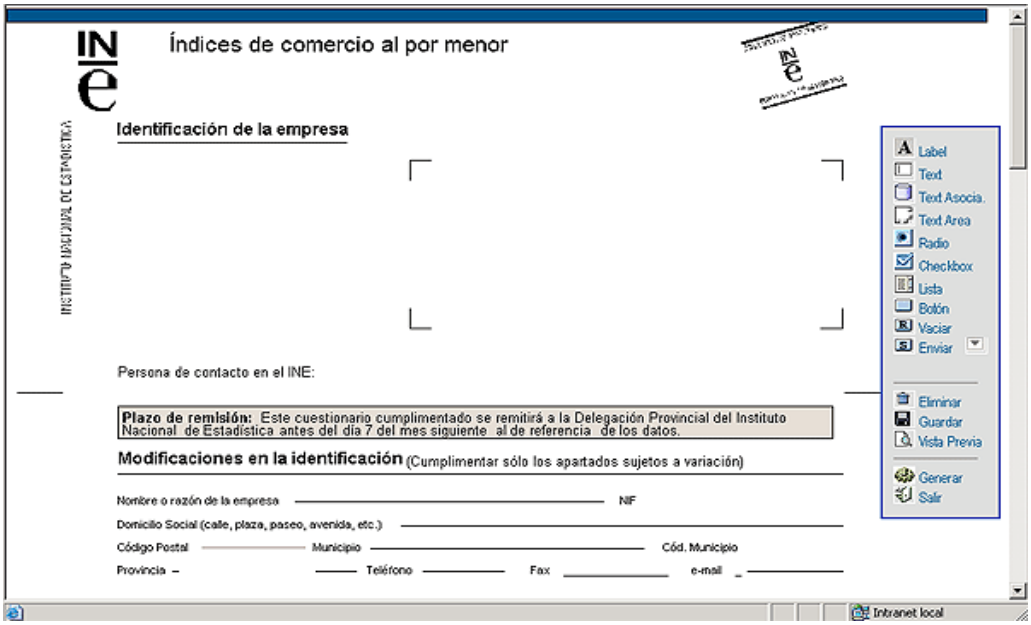


Figure 5.- The survey is loaded in the editor to create the form

Once the survey is loaded in the editor, the user may define the components and its location over the background. The layer information from the HTML obtained in the first step is very important now; when a user draws a component in the form, the editor calculates the position trying to associate the component to a field in the survey, matching them depending on the proximity.

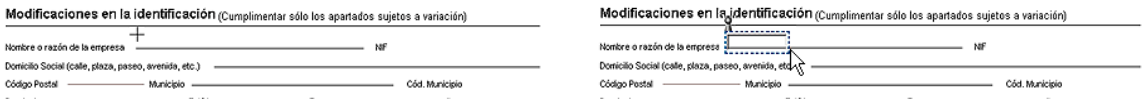


Figure 6.- Drawing a component from the toolbar

Components are managed as usual in an IDE, including the possibility of copy, paste, align vertically or horizontally, size, resize or set equal size for several of them, defining default values for each component type, etc.

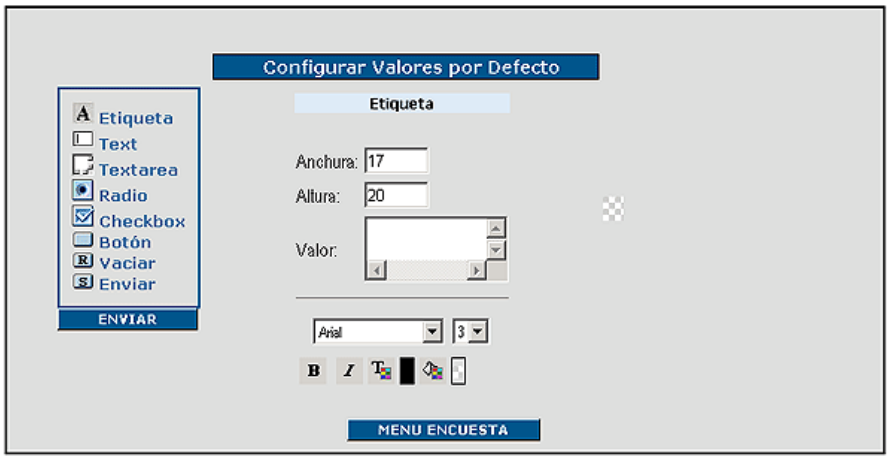


Figure 7.- Defining default values for label components in a survey

In some cases it is very interesting to define a domain of values for some components, specially in case of combo boxes or lists, in the figure below a example on how to define a fixed domain for states is shown. The possibility of recovering dynamically the domain from a table in a database has been also implemented.

Figure 8.- Domain definition window

Another interesting functionality while defining surveys is to set some kind of constraint to the values of the different fields. The user may define both individual or group constraints. Individual constraints are to be applied to the value of a field, while group constraints are applied to the set of values of some fields. In any case, the constraints may obly to fill the field (*mandatory*) or define restrictions on the value of the field if any, through formula specification, as shown in the figure below. In the combo box the user may select some common predefined functions.

After some time defining the form associated to a survey, the user may have added and removed several components (in fact, usually, dozens of them). Each new component is named automatically by the editor and introduced in the tab order. The user may alter both, the given name to make the application comprehensive

(very recommendable) and also the tab order, so that the final user may navigate through the fields with the tab control, which is one of the most accepted (and requested) features from final users.

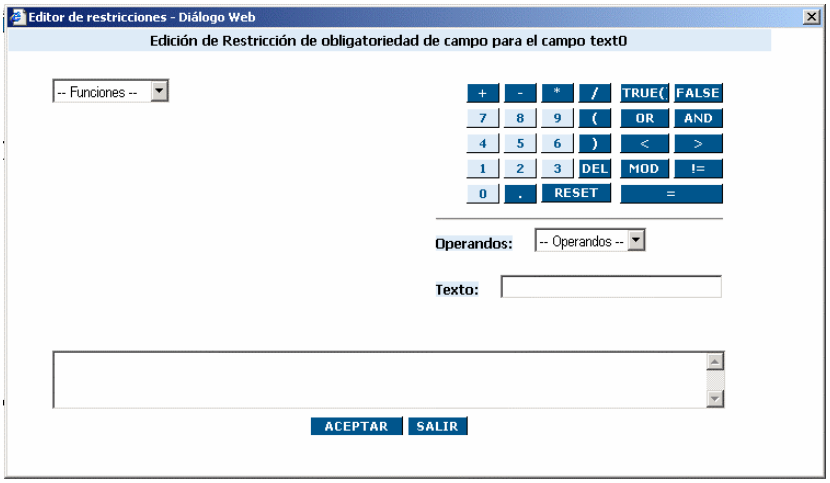


Figure 9.- Constraint definition

Not only the form to publish the survey is important, but also to provide the users an identical aspect and information to reduce the difference to the hard copy surveys. In the same terms the background is the same that in that case, the designer provides the chance to add help buttons to the form, then these buttons may be linked to a help file in HTML format that usually contains the same help provided to non Internet users.

The results: an external user fulfilling a survey

When the process is finally finished, the user gets a result which may be considered in two perspectives:

- An application that supports the publication and collection of data from the web, on a surprisingly low cost and effort, and which is integrated in the portal of the organization.
- A service for the information providers, those are obliged by law to compliment the surveys, and may now carry out this task easily.

Figure 10.- The web site allows not only filling real surveys, but making tests

The applications generated by GEN are integrated in the web site through shared tables in an Oracle database, and the development of a section in which the user may select the survey in which he is interested. The access to the section is based on the typical user and password, that has been replaced by an order number and control digit. This information is sent to each organization together with customized instructions and paper surveys via mail as official notifications, reusing the existing infrastructure and the old procedure, which is mandatory to maintain.

When a survey is generated, the user may define whether the survey shall exist also in test mode. This allows final users to practice with real surveys without writing results or submitting information. The only difference between tests and real surveys is a parameter switching between them.

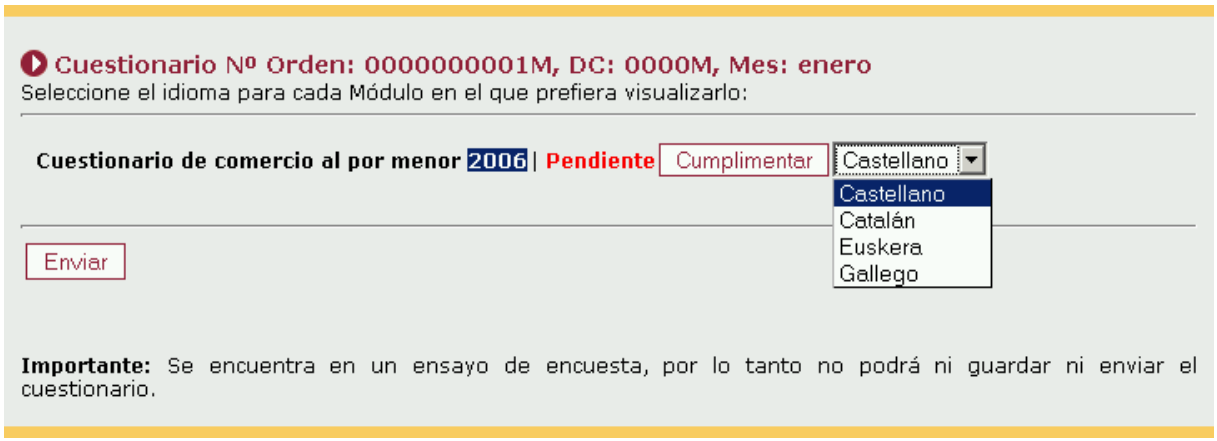


Figure 11.- Language selection. GEN generates multilingual applications

The surveys can be displayed in all Spanish official languages as mandatory by law, but the data are considered to correspond to the same survey, this is logical until a problem arises: the staff in charge of collecting data and extracting statistical information do not use to speak all of them. It is considered a great lack not for the application but for the legal procedure.

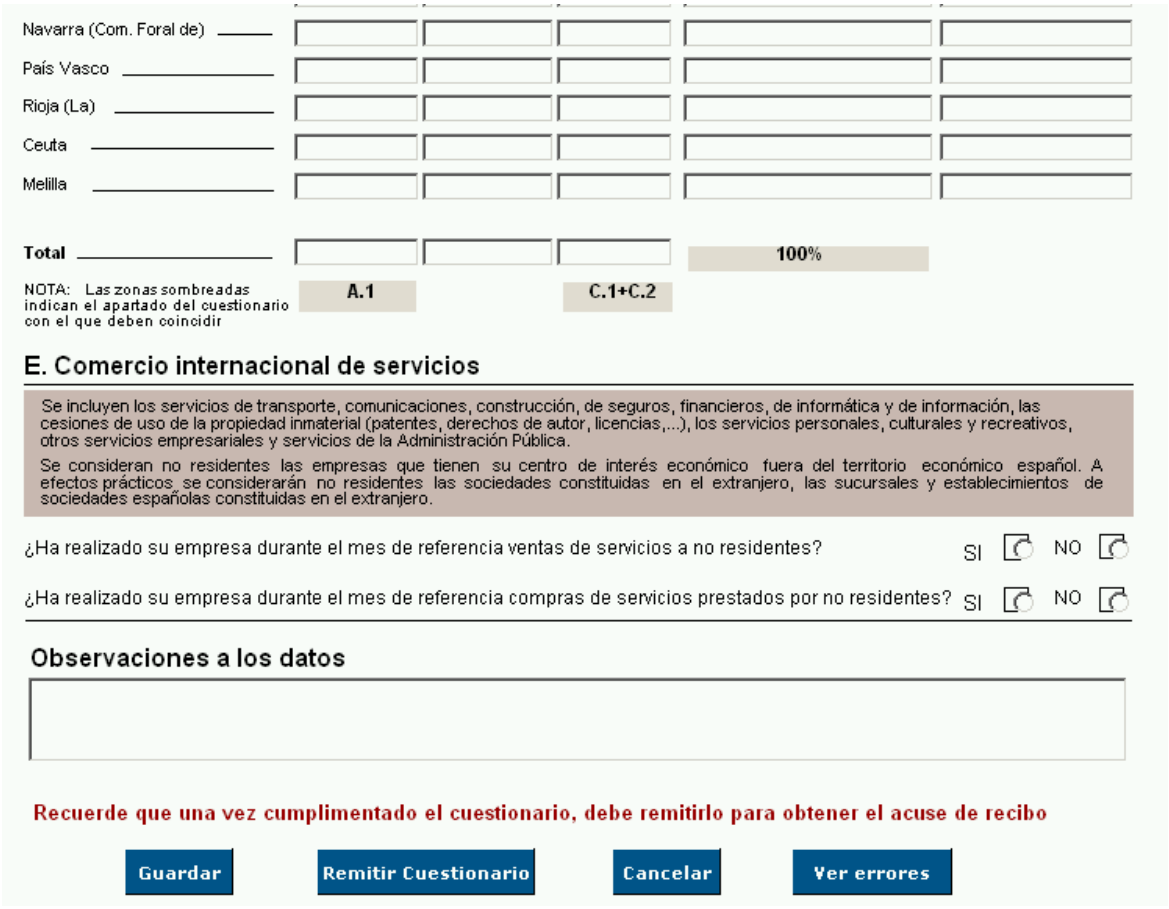


Figure 12.- A piece of the final presentation of a survey on commerce

In figure 12 a fragment of the web showing a survey is presented. Note the final alignment of the fields, the different component types and the buttons in the bottom, offering the functions of *Save*, *Send form*, *Cancel* and *Show errors*.

Before sending the final answer to a survey, the user may save any number of times, and before submitting may query the errors in the different fields in order to make corrections. Sending a correct answer is mandatory by law.

The final reply of a user to a survey is stored in XML format into the database, and will be later exported to the analysis application.

This is the common process for most users, however, a number of enterprises, mainly holdings, do need to fill lots of surveys that they may automate by using their information systems. A complementary application has been developed allowing this kind of users replying massively to the surveys through bulk loads. This application, called G2G, receives XML files containing several surveys, validates the data using the XForms from GEN, after formatting incorrect data or truncating data too long for the precision in database.

Conclusion

Sometimes the public organisations shall face political commitments with little help from the Government, same budgets and staff, allowing the development of beautiful projects in which technology makes life easier. Probably this project had lost the chance if there where enough staff to develop each application as usual software projects.

A tool has been developed that allows a non technical user, with no knowledge on software architectures, design, HTML, nor Java, struts or databases, carry out the task of creating applications. Of course this does not replaces technical staff, but get them to develop only technical high level tasks.

The fact is that a complex technology, that has been tested by Technical University of Madrid, and that shuns from the typical theoretical aspects from code generation, trying to open a new spectrum of possibilities, has been applied successfully to a specific application domain demonstrating that the approach is good and feasible.

For public administrations, the project allows a rise in the corporate image, and the observance of the Lisbon agreements in one of the most difficult aspects it was facing to.

The Spanish National Statistics Institute has been capable to meet the goals described in its services white-paper, minimizing the impact in the staff and budget, and reducing sensitively the time to market for each survey, which is available in Internet in less than a week, so that the complete process increases a minimun percentage from the previous one, with no digital survey. Also the time in which all data has been collected is lower, because there is no need to transcript all data to a computer.

Collaborators that fill in the forms, now have the chance to select the preferred choice: digital or paper surveys, knowing that digital ones provide the proper mechanisms to automate the calculations, report possible errors reducing the time to finish the job, etc.

The success in figures: in the first two months more than 10% of the surveys where submitted via Internet, more than 600 per day are being received, and 7 multilingual surveys have been generated; 168.433 organizations are working with the generated applications.

Bibliography

[Alves2002] Alves, L., von Staa, A., 2002. A construction process for artifact generators using a CASE Tool. *Proceedings of Workshop on Generative Programming 2002.* Austin, Texas, USA, pp. 7-10.

[Arisholm1998] Arisholm, E. et al. 1998. Incorporating Rapid User Interface Prototyping in Object-Oriented Analysis and Design with Genova. *The Eighth Nordic Workshop on Programming Environment Research.*

[Batory 1994] Batory, et al. 1994. The GenVoca Model of Software-System Generators. *IEEE Software*, 0704-7459/94.

[Boehm2000] Jongmoon, B., Boehm, B., 2000. Empirical Analysis of CASE Tool Effects on Software Development Effort. *Technical Report University of South Carolina. Center for Software Engineering.*

[Cuvillo2004a] Cuvillo, C. et al. 2004. Generation tool for DBMS focused applications. *Applied Computing 2004*. Lisbon. Portugal.

[Cuvillo2004b] Cuvillo, C. et al. 2004. Multiplatform web applications generated from relational data models. (In spanish). *CISCI 2004*. Florida, USA.

[Calejo2002] Calejo, M. et al. 2002. Web Application Maker - A model based approach to web database development. *6th International Conference on Enterprise Information Systems.*

[Grønbæk 1991] Grønbæk, K.et al. 1991. ApplBuilder - an Object-Oriented Application Generator Supporting Rapid Prototyping. *The 4th International Conference on Software Engineering & Its Applications*.

[Schmidt2000] Schmidt, D., et al., 2000. POSA2: Pattern -Oriented Software Architecture: Patterns for Concurrent and Networked Objects.

Authors' Information

Hector Garcia - Contrated Professor. Technical University of Madrid. E.U. Informática. Ctra. de Valencia Km. 7. E28031 Madrid. e-mail: hgarcia@eui.upm.es

Carlos del Cuvillo - Associate Professor. Technical University of Madrid. E.U. Informática. Ctra. de Valencia Km. 7. E28031 Madrid. e-mail: ccuvillo@eui.upm.es

Diego Perez -. Consultant. Technical University of Madrid. E.U. Informática. Ctra. de Valencia Km. 7. E28031 Madrid. e-mail: dperez@tdi.eui.upm.es

Borja Lazaro - Technician, Group leader. Technical University of Madrid. E.U. Informática. Ctra. de Valencia Km. 7. E28031 Madrid. e-mail: blazaro@eui.upm.es

MODELING AND ANNOTATING THE EXPRESSIVE SEMANTICS OF DANCE VIDEOS

Balakrishnan Ramadoss, Kannan Rajkumar

Abstract: Dance videos are interesting and semantics-intensive. At the same time, they are the complex type of videos compared to all other types such as sports, news and movie videos. In fact, dance video is the one which is less explored by the researchers across the globe. Dance videos exhibit rich semantics such as macro features and micro features and can be classified into several types. Hence, the conceptual modeling of the expressive semantics of the dance videos is very crucial and complex. This paper presents a generic Dance Video Semantics Model (DVSM) in order to represent the semantics of the dance videos at different granularity levels, identified by the components of the accompanying song. This model incorporates both syntactic and semantic features of the videos and introduces a new entity type called, Agent, to specify the micro features of the dance videos. The instantiations of the model are expressed as graphs. The model is implemented as a tool using J2SE and JMF to annotate the macro and micro features of the dance videos. Finally examples and evaluation results are provided to depict the effectiveness of the proposed dance video model.

ACM Classification: J.5 [Arts and Humanities]: Performing Arts; H.5.4 [Information Interface and Presentation]: Hypertext/Hypermedia.

Keywords: Agents, Dance videos, Macro features, Micro features, Video annotation, Video semantics.

1. Introduction

Dance data is essentially multimedia by nature consisting of visual, audio and textual materials. Dance video modeling and mining depends significantly on our ability to recognize the relevant information in each of these data streams. One of the most challenging problems here is the modeling of the dance video semantics such that the relevant semantics are consistent with the perception of the real world.

The classical and folk dances are the real cultural wealth of a nation. In India, the most important classical dances are *Bharathanatyam, Kadak, Kadakali, Kuchipudi* and *Manipuri* (Saraswathi, 1994). Traditionally, dance learners perform dance steps by observing the natural language verbal descriptions and by emulating the steps of the choreographers. Therefore, the properly annotated dance videos will help the present and future generations to learn dance themselves and minimize the physical presence of the choreographers.

Notations are used everywhere and are most important for the dancers to communicate the ideas to the learners. They use graphical symbols such as vertical lines, horizontal lines, dots, triangle, rectangle etc, to denote body parts' actions on paper. Labanotation (Hutchinson, 1954) and Banesh (Ann, 1984) have been the frontier notational systems to record the dance movements or dance steps. Many western dances are using Labanotation to describe dance steps. However, many choreographers still follow the traditional way of training their students using natural language descriptions, because of the very few recording experts and inherent complexity of reading and understanding the symbols. Moreover, all Indian dances have unique structure and no common notational structure exists, apart from wire-frame stick diagram representing a dance step. Due to lack of notations, it is evident that the complexity of modeling the dance video semantics is relatively high.

Since the dance steps were archived in paper form and many classical dances lack notations, this kind of archival of dance becomes impossible even today. With the advances in digital technologies (Dorai, 2002) nowadays, magnetic tapes and disks record dance presentations efficiently. But, searching a dance sequence from these collections is not efficient, because of the huge volume of video data. The solution is to build a dance video information system so as to preserve and query the different dance semantics like, dance steps, beyond the spatio-temporal characteristics of the dancers and their dance steps.

The dance video database system requires an efficient video data model to abstract the semantics of the dance videos. To be more precise, the dance video data model should:

- abstract the different dance video semantics such as dancers, dance steps, agents (i.e., body parts of the dancers), posture, speed of dance steps, mood, music, beat, instrument used, background sceneries and the costume. More importantly, the spatio-temporal characteristics of the dancers must be incorporated in the model;
- capture the structure of the dance videos such as shot, scene and compound scene abstracting the different components of the accompanying song.

This paper addresses two related issues: modeling the semantics of the dance videos and annotating the dance steps from the real dance videos. The dance video semantics model represents the different types of dance semantics in a simple, efficient and flexible way. The annotation tool manually annotates the semantics (as macro and micro features) for further query processing and video mining. The main contributions of this paper are as follows:

- We propose a generic video data model to describe the dance steps as video events;
- We introduce the Actor entity in order to store the event specific roles of a video object. That is, an actor entity describes the context dependent role of the video object;
- We introduce the Agent entity to describe the context dependent action that is associated with the actor entity;
- We develop a tool that implements the dance video model in order to annotate the different dance semantics.

The rest of the paper is organized as follows: Section 2 presents some related works on video data models. Section 3 describes the different semantics of the dance videos. The DVSM for the dance video is introduced in Section 4. Section 5 illustrates the implementation of the DVSM using Java technologies. The proposed video model is evaluated against a set of conceptual and semantic quality factors in Section 6. Finally, Section 7 concludes the paper.

2. Related Work

Video data modeling is an important component of the dance video database system, as it abstracts the underlying semantics of the dance. This section briefly reviews some of the existing video modeling proposals and discusses the applicability to dance videos.

Colombo(1999) classifies the content-based search as semantic level search (e.g. objects, events and relationships) and low level search (e.g. color, texture and motion). They call the corresponding systems as first and second generation visual information systems. Several key word based techniques are applied to semantic search models, such as OVID (Oomoto, 1993), AVIS (Adali, 1996), Layered model (Koh, 1999) and Schema less semantic model (Al Safadi, 2000). Second generation systems provide automatic tools to extract low level features and subsequently semantic search is performed. Some of these systems include, but not limited to QBIC, Virage, VisualSEEK, VideoQ, VIOLONE, MARS, PhotoBook, ViBE, and PictHunter (Smeulders, 2000;

Antani, 2002). However, these systems are either based on textual annotations or purely low level features, but not incorporating the other one.

In (Shu, 2000), Augmented Transition Network based semantic data model is proposed. The ATN models the video based on scenes, shots and key frames using strings as a sequence of characters. The string representation is used to model the spatial and temporal relationships of each object (moving and static) in a shot of the traffic video. Since the semantic features of dance videos are complex, the entire scene or shot cannot be abstracted in a single string.

Translucent markers, reflector costumes, special sensors and specialized cameras are used to capture and track human body parts' movements in some applications such as aerobics, traffic surveillance, sign language, news and sports videos (Vendrig, 2002). In order to record and analyze the dance steps of a dancer, based on this technique requires a special translucent markers or reflector costumes for the dancers. However, dancers do not prefer to use these costumes as these costumes hide the dancer's make-ups and costumes. Moreover, these markers and reflectors prohibit the realism, affect dancer's comfort as well as reduce the focus or concentration of the dancers. Hence, automatic analysis of dance steps to extract the semantics of the dance steps is very complex.

Recently, the extended DISIMA(Lei Chen, 2003) model expresses events and concepts based on spatio-temporal relationships among salient objects. However, the required dance video database model has to consider not only salient objects, but all objects such as instruments, costumes, background and so on.

Event based syntactic-semantic video model (we call it as, ESSVM) (Ahmet, 2004) proposes Actor entity to specify the context dependent role of a player in soccer sports. This model represents the events such as free kick, goal, penalty etc, in which player assumes different roles such as scorer, assist-maker etc. But in dance videos, the contextual information of the dance events has to be described at multiple levels like actor and agent, rather than at a single granularity of actor entity.

COSMOS7 (Athanasios, 2005) models objects along with a set of events in which they participate, events along with a set of objects and temporal relationships between the objects. This model does not model the temporal relationships between events and the contextual roles. It models the events at a higher level only like speak, play, listen etc, whereas dance video model needs more detailed level of event representation such as agents, their action, speed of action, associated song and so on.

3. The Semantics of Dance Videos

Generally, dance information is dominated by visual content such as steps, posture and costume and the accompanying audio such as song and music. Hence, dance videos are rich in semantics and provide ample scope for the efficient semantic retrieval and dance video mining. This section illustrates the song that accompanies the dance performance, the different dance video types and the features of the dance videos in detail.

3.1. Song Granularity

Dance video contains several dance steps representing each song. In the case of classical dance, it is simply a collection of songs choreographed on the stage or theatre with a single start-stop (Cheng, 2003) camera operation. On the other hand, in a movie dance, a movie contains several songs and for each song dance steps are choreographed by the dancers. A song in a movie may be recorded with multiple start-stop camera operations. For instance, an Indian movie will normally contain about five to six songs. Here, each dance step may represent a step from any of the Indian dances or a new step innovated by the choreographer. Further, it includes the presentation aesthetics such as mood, feelings, emotion and so on.

Song is composed of four parts: Introduction, Additional Introduction, Chores and Stanzas (Web of Indian Classical Dances, 2003). Depending on the type of a song, Additional Introduction and Chores may be optional. Each part has few lines of lyrics for which dance steps are choreographed. In the dance video hierarchy, a shot represents a dance step, a scene represents dance steps of any of the song parts which are recorded in the same location and a video clip represents dance steps of a song. Our DVSM will represent the semantics of one dance step as a dance event. Dance step is the unit of analysis in this paper.

3.2. Features of Dance Videos

There are two types of dance video features- macro features and micro features and are annotated by the human annotators at macro and micro levels (Forouszan, 2004) accordingly. Macro features are general properties of the dance that are event independent and micro features are the properties of the dance step. That is, micro features are spatio-temporal characteristics of the dancers while rendering the dance steps. Micro features can also be called as event dependent features.

Macro features(or Bibliographic features): Date of recording, time of recording, geographic origin of the dance, geographic origin of the dancers, sex, age, number of dancers in a dance, type of performance venue (such as theatre, open-air, etc), type of the accompanying song, type of accompaniment, type of musical instrument used and types of dance videos. The different dance videos are movie dance video, theatre dance video, folk dance video, classical dance video, street dance video and festival dance video. These macro features are independent of the dance steps and are common to all dances.

Micro features (Dance step specific features): Spatio-temporal features classify dance movement behavior which include: movement of one dancer in relation to another dancer, movement of a specific body part (such as eye, leg etc. Refer Appendix-A for a complete list) of a dancer in relation to another part of the body, movement path of the dance (such as circular, linear, serpentine and zigzag), distance between body parts of a dancer while performing a dance step and distance between dancers.

Hence, the proposed video model has to characterize a set of macro features and micro features that exist in the dance videos.

4. The Dance Video Semantics Model

Conceptual model abstracts the dance video data into a structure for later querying its contents and mining some interesting patterns. For efficient conceptual modeling, one should know how choreographers demonstrate a dance to the learners. They are the experts in describing the rhythmic steps to the audience. This section presents a generic dance video model that efficiently describes the dance steps. Every dance step is called as an event and the model represents dance events by a set of micro features. The model is generic in the sense that it is applicable to any type of dance videos. DVSM is an extension of ER (Chen, 1976) with object oriented features. The goal of the model is to describe a dance step as an event.

The main entities of the model are events, objects that participate in these events, actor entities that describe contextual roles of objects in the events, agent entities that represent the action of the actor and concept entities that model the cognitive and affective features of the dancers.

For example, consider a dancer object with name, age, address and all other event independent attributes. The same dancer assumes different roles throughout the dance video. That is, he becomes hero in one dance step, lover in another dance step and so on. Roles are defined as attributes of Actors. Some other examples of actors are heroine, leader, follower, group dancer, friend etc. These context specific object roles form separate actor entities, which all refer to the same dancer object. Although one would say that actor performs the action in an event, finer granularity is necessary as far as dance videos are concerned. Therefore, contextual data of the dancers have to be described in two levels. A particular dance step is characterized by the actions of the agents who belong to the actors. Spatio-temporal characteristics are part of the actors as well as agents. Hence, they are described as attributes of actors and agents.

Apart from the dancer object, DVSM may also represent the ordinary objects with a standard UML class diagram. Some of them are: speed of the action of an agent, instrument used and the posture of the actor. The graph meta-schema of the DVSM is depicted in Figure1.

The graphical notations used in DVSM are described as follows. A rectangle node refers to an entity or an object. A round rectangle node refers to a concept. A dotted rectangle node denotes an actor entity. A thick rectangle node shows an agent object. Event entity is modeled with a trapezoid. Attributes of entities and relationships are represented with oval nodes. Relationships are denoted with directed lines on which the name of the relationship is denoted. Relationships without their names represent the containment type.

The model is instantiated as a directed acyclic graph. The reason for choosing graphs is that it elegantly models repetition of dance steps and has matured as a graph database. If a dance step repeats after some time, it just requires another edge to point to the same node. A graph is formally defined as follows: Let G = (V, E) be a

directed acyclic graph, where V denotes set of vertices and E denotes set of directed edges. The different entity classes, events, actors, agents, concepts, and other basic classes become vertices of the graph. Similarly, the set of interaction relationships will be denoted as directed edges of the graph.

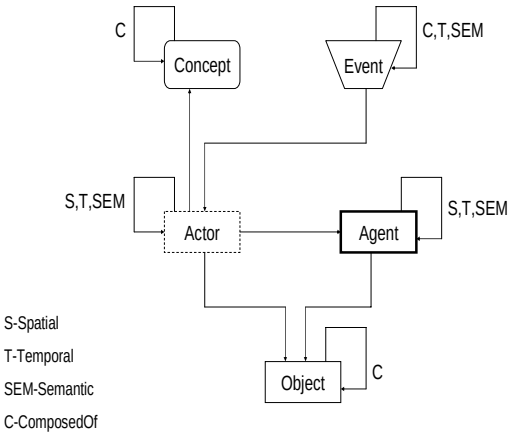


Figure1: Graphical representation of *DVSM*

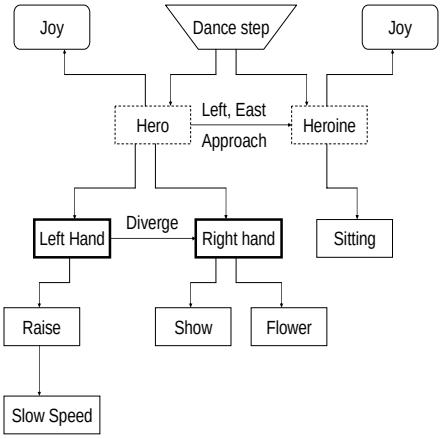


Figure2: Graph of dance step containing actors and agents.

The conceptual representation of an event highlighting a dance step as an instance of the graph, is depicted in Figure2. In this figure2, the dance event consists of two actors whose roles are hero and heroine. Hero is standing left to the heroine initially and facing east. Heroine is sitting and facing west. Now hero approaches the heroine. These spatio-temporal semantics are stored as relations. Event independent characteristics of the actor are stored as video objects separately (not shown in the figure). The actors express joy and it is initialized as emotion. Hero raises his left hand to chest level with medium speed and displays a flower to the heroine with his right hand. Heroine remains idle without performing any action. This dance step is choreographed as part of one line of lyrics of a song. Due to overflowing of nodes, attribute nodes are not shown in this figure. The entity classes and relationships of the model are formally defined as shown below:

4.1. Event Entity Class

Dance step of a song is known as a dance event. For instance, consider a Bharathanatyam step Samathristy (Saraswathi, 1994). It is performed with the eyes by keeping them static without blinking. This step represents a thought, firmness, surprise or an image of an angel. Also, dance events can be combined to form a composite dance event. As an illustration, consider a dance step, Chandran. This step represents a moon and is a combination of two other steps: Pathagam and Lola pathmam and should be rendered concurrently. Pathagam is performed by keeping the thumb closed and the other four fingers straight and denotes clouds, air, sword and blessing. Similarly, Lola pathmam is performed by keeping all the fingers open and stretched and represents a sun. Composite dance event many represent events which are rendered concurrently or sequentially by a dancer.

Dance events are composed of actors, posture of the actors, cognitive state of the actors and the interactions in space and time between agents and actors. Formally, a dance event is described as a tuple,

$$\text{Event} = \{ \text{EID}, \text{D}, \text{AL}, \text{ND}, \text{ML} \}$$

where, EID is a unique identifier of the dance step, D is the description of the dance step, AL denotes the list of actors, ND is the number of dancers who are performing steps in the event and ML is the media locator of the video clip. Here, this Event tuple corresponds to Event object of Figure1.

4.2. Basic Entity Class

A dance video object refers to a meaningful semantic entity of a dance video database. It can be described using attributes which can represent macro and micro features. Formally, it is defined as shown below:

$$\text{Object} = \{ \text{OID}, \text{V}, \text{TY} \}$$

where OID is a unique object id, V = {a1:v1, ..., an:vn } is *n* event independent or dependent attribute value pairs and TY = { AID, AGID } denotes the dependency of the object, either actor or agent (to be defined later).

For example, hero is showing a flower in his right hand. Here, *ShowFlower* is the object in which V denotes attributes action and instrument with values show and flower respectively. TY holds ID of the agent, Right Hand which belongs to the actor Hero. In this case, V represents event dependent (i.e., dance step dependent) values. Similarly, object may also represent any of the macro features.

4.3. Actor Entity Class

Actor is a spatiotemporal entity in dance videos. So the existence time (Vijay, 2004) can be associated with the entity and it represents the life span of it. Actor is also a spatial entity. Therefore actor's displacement in space is modeled using Trajectory Points as in MPEG-7 (Martinez, 2003). Hence, actors are spatio-temporal entities playing context dependent roles in the events. Actors can have spatial, temporal and event specific semantic attributes describing their roles. The roles can be linguistic roles (Martinez, 2003) as in MPEG-7 or any semantic roles, such as loves.

The existence time predicate ΦACTOR, which is associated with the actor entity class, defines life span of the actor in terms of the existence time granularity (e.g. min and sec). $\Phi ACTOR: S(ACTOR) \times Z \rightarrow B$. This predicate takes a particular actor entity and a particular granule (denoted by an integer; say sec) and evaluates to a Boolean. If it is true, then that actor exists in the modeled reality at that granule (sec).

Constraint.1: *Life span of an actor can exist only within the defined lifespan of the event to which it belong*.

Formally, an actor entity can be described as follows:

Actor = { AID, EID, DID, R, L, T, P }

where EID is the event id, DID is the corresponding dancer id, R denotes either semantic or linguistic roles of an actor, L is the existence time or lifespan, T represents the trajectory points(Point Set) as in Mpeg7 and P is the posture of the actor, which is a basic entity.

4.4. Agent Entity Class

Agent entity class represents the finer spatio-temporal semantics of the actions. The agent entity is the one which is most important in dance videos. The essence of a dance step is the actions done by the actors and it is the agent that performs the action. This is an exclusive feature of the dance videos. All other video types possess just one or two agents, which are fixed and do not play any significant role at all. For example, legs are agents in soccer sport videos, bat and ball are agents in cricket sport videos. Agent entity elegantly models the action of the agent which belongs to an actor. For instance, left eye and right eye of a heroine are agents. Formally, it is defined as:

Agent = { AGID, AID, EID, L, T, X, S, I }

where AID and EID denote the actor id and event id respectively, X is the action agent performs, S denotes speed of X and I is the instrument held by the agent. Also, L and T depict the lifespan and spatial trajectory, similar to actor objects. Here, X, S and I are all basic entity types as defined earlier.

4.5. Concept Entity Class

The cognitive and affective content of an actor is modeled as a concept object. The concept is modeled as a separate entity type because of its ontological nature, thereby improving the semantic search. Formally, a concept entity can be defined as

Concept = { CID, AID, EID, T, D }

where T = { Emotion, Feeling, Mood } and D denote type of the concept and description as a string using natural language respectively.

4.6. Interaction Relationships

An interaction relationship relates members of an entity set to those of one or more entity sets. The DVSM employs the following set of relationships between the different entity sets. They are Composition (C), Spatial(S): which are topological (Egenhofer, 1994) and directional (Li, 1996), Temporal(T): Allen's interval algebra (Allan, 1983), Spatio-temporal, Motion(M): such as approach, diverge, stationary which are defined over the basic temporal relations (Athanasios, 2005), Semantic(SE) and Ontological(O).

The following section describes the set of relationships that occur between the various dance video entity sets.

4.6.1. Event Relationship

The relations between events are composition and temporal. Intuitively, a dance step of an actor may be followed by another actor immediately. Similarly, a dance step of an actor may be repeated by another dancer some time later. These follows and repeats relations are cues for later retrieval and mining operations. For example the query, find the set of dance steps done by a dancer, that is repeated by another dancer, can be processed by checking the life spans of the corresponding events.

Suppose E1, E2, ..., En are dance events participating in a temporal relationship. Let a1 and a2 be the actors, x1 and x2 be the actions of agents present in E1 and E2 respectively. Then, the predicate $\Phi REPEATS:S(X)\times S(A)\rightarrow E$ can take an actor and action and can return a set of events in which the action is performed. There is a constraint on the REPEATS predicate.

Constraint.2. *Let LS1 and LS2 be the lifespan of E1 and E2 respectively. Then*

$$(x1 = x2\ )\ V\ (LS1 < LS1) \Longrightarrow (E1 = E2)$$

Similarly, the other predicates such as *performSameStep, performDifferentStep*, and *observe* can be formulated, apart from *follows* and *repeats* predicates. Event relationships are formally defined as follows:

$$EE = \{ SRC, TAR, LST \}$$

where SRC and TAR denote the source and target event ids and LST is the set of composition and temporal relationships which hold between source and target events.

4.6.2. Object Relationship

Objects can be composed of other objects. For example, consider Figure2 where hero holds a flower in his right hand. Here, flower is an example of an object. Formally, the relationship between objects can be represented similar to event relationships with a restriction that the SRC and TAR can be basic entities and LST will contain only composition relations.

4.6.3. Actor Relationship

Actor relationship represents the relationship between the roles of the objects, such as relation between hero and heroine who are dancer objects. Spatial, temporal and semantic relationships exist between the actors in a particular dance event. For instance, hero standing left to the heroine initially, may approach the heroine. This dance semantic contains spatial and motion relationships left and approach respectively. The actor relationship is formally defined as shown below:

$$AA = \{ AID1, AID2, O1, O2, LST \}$$

where AID1 and AID2 are roles of the dancers O1 and O2 respectively and LST is now the set containing spatial, temporal and semantic relationships. Note that O1 and O2 are basic entity types.

4.6.4. Agent Relationship

Agent relationship is a second level semantic relation that describes the spatial and temporal characteristics of the agents. That is, agent relationship represents the finer semantics between the body parts of an actor. For instance, heroine is touching her left cheek with the index finger of her right hand. So, left cheek and right hand are the agents and finger can be the instrument used in the semantic relationship touch. Agent relationship is formally defined as,

$$AGAG = \{ AGID1, AGID2, AID, LST \}$$

where AGID1 and AGID2 are agentIDs of an actor AID and LST is similar to actor relationships.

4.6.5. Concept Relationship

Concept relationship is an ontological relationship (O) between concept entities. Typical ontological relationships (Guiness, 2004) are subClassOf, cardinality, intersection and union. This relationship is similar to event relationship with a modification that the source and target ids represent concepts and LST holds only ontological relations. All other types of relationships between the different dance video entities are either semantic relationships or composition relationships such as partOf, composedOf, memberOf and so on. Table 1 summarizes the semantics of the kinds of relationships that exist between the dance video entities.

Table 1. Semantics of Relationships.

	Event	Object	Actor	Agent	Concept
Event	C,T,SE		C		C
Object		C	C	C	
Actor	C	C	S,T,SE	C	C
Agent		C	C	S,T,SE	
Concept	C		C		O

5. Implementation of DVSM

We have implemented the model in order to annotate the macro and micro features that are associated with the dance video. The tool has been developed using J2SE1.5 and JMF2.1.1 under Dell workstation. The tool is interactive as it minimizes the hard coding.

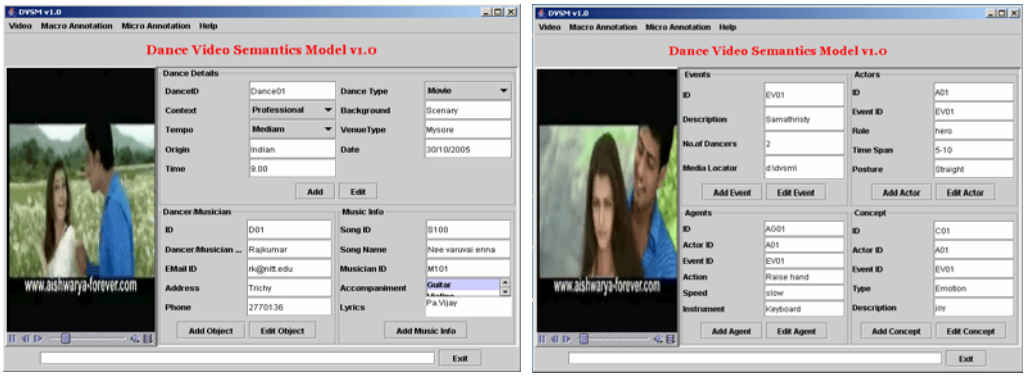


Fig.3 and Fig.4. Macro annotation and Micro annotation of the event semantics.

The dance video can be annotated by looking at the video clips that is running. Macro features can be annotated initially. The details of the dancers, musician, music, song, background, tempo, dance origin, context (whether live, rehearsal, professional play, competition etc), date and time of recording, type of performance venue and type of dance video are annotated. The screen shot depicting the rendering of the dance and interactive annotation of macro features is shown in Figure 3.

Then, micro features of every dance step of a song have to be annotated. The screen shot depicted in Figure 4 represents events, actors, agents and concepts. The annotator, by looking at the video, annotates the different information pertaining to these entity types in the order: event, actors of this event, agents of the actors, concepts revealed by the actors. But, one can swap the annotation of agents and concepts depending on his interest. The user interface has been carefully designed such a way that it minimizes the hard coding, as many of the graphical components will be populated automatically.

The second part of the micro features annotation involves the description of the various relationships between the entity types. For instance, event relationships, actor relationships, agent relationships and concept relationships describe the spatial, temporal, motion and semantic relations that exist between the entity types. The annotated data are stored in a backend database.

6. Evaluation

Batini(1996) posits that conceptual model should possess the basic qualities: expressiveness, simplicity, minimality and formality. Additionally, Harry(2001) outlines other semantic qualities for video types. They are: explicit media structure(M), ability to describe objects (O), events (E), spatial relationships(S), temporal relationships(T) and integration of syntactic and semantic information (I). This paper introduces another factor, contextual description (Actor(A) and Agent(G)) to evaluate the proposed model.

The DVSM satisfies all the semantic quality requirements. Moreover, DVSM is unique in modeling the finer spatio-temporal contextual semantics of the events at finer granularity, with the help of agent entity type. Table 2

contrasts the existing semantic content based models against the semantic quality factors. The table illustrates that some models lack semantic features, some lack syntactic features and only few models integrate both syntactic and semantic relationships. Some applications, like soccer sports video, require the model to represent the contextual features of objects (called actors). The ESSVM proposed by Ahmet et al, describes contextual information at actor level. However, dance videos require contextual description at multiple granularities (called agent), beyond the actor level. Our proposed semantic model possesses both contextual abstractions-actor and agent, apart from the other semantic qualities. Hence, with the agent based approach, the paper claims to have achieved conceptual, semantic and contextual qualities in dance video data modeling.

Table 2. Comparison of semantic video models.
Legend: M-Media structure, O-Object, E-Event, S-Spatial, T-Temporal, I-Syntactic-semantic info, A-Actor, G-Agent.

Model	Semantic Qualities							
	M	O	E	S	T	I	A	G
AVIS	×	×		×	×	×		
OVID	×	×		×	×	×		
QBIC	×	×		×				
DISIMA	×	×	×	×	×	×		
COSMOS7		×	×	×	×	×		
ATN	×	×		×	×	×		
ESSVM	×	×	×	×	×	×	×	
DVSM	×	×	×	×	×	×	×	×

7. Conclusion

Data semantics provides a connection from a database to the real world outside the database and the conceptual model provides a mechanism to capture the data semantics (Vijay, 2004). The task of conceptual modeling is crucial and important, because of the vast amount of semantics that exist in multimedia applications. In particular, dance videos possess several interesting semantics for modeling and mining. This paper described as agent based approach for elicitation of the semantics such as macro and micro features of the dance videos. An interactive annotation tool has been developed based on the DVSM for annotating the dance video semantics at syntactic, semantic and contextual levels. Since dance steps are annotated manually, it is somewhat tedious to annotate dances by the dance expert.

Further work would be useful in many areas. It would be interesting to explore how DVSM can be used as a video model for exact and approximate query processing. As MPEG-7 is used to document the video semantics recently, it is valuable to employ MPEG-7 for representing dance semantics to enable better interoperability. Finally, it will be useful to explore how video mining techniques can be applied to dance videos.

Acknowledgements

This work is supported fully by the University Grants Commission (UGC), Government of India grant *XTFTNBD065*. The authors thank the editors and the anonymous reviewers for their insightful comments.

Bibliography

Adali, S, Candan, K.S, Chen, S.S, Erol,K & Subramanian, V.S.(1996). The Advanced Video Information System: Database structures and query processing. Multimedia Systems. 4, 172-186.

Ahmet Ekin, Murat Tekalp, A & Rajiv Mahrotra.(2004). Integrated semantic-syntactic video event modeling for search and browsing. IEEE Transaction on Multimedia, 6(6), 839-851.

Allen, J.F.(1983). Maintaining knowledge about temporal intervals. Communication of ACM, 26(11), 832-843.

Al Safadi, L.A.E & Getta, J.R.(2000). Semantic modeling for video content based retrieval systems. 23rd Austral Asian Computer Science conference, 2-9.

Ann Hutchinson, G.(1984). Dance Notation: Process of recording movement. London: Dance Books.

Antani, S, Kasturi, R & Jain, R.(2002). A survey on the use of Pattern recognition methods for abstraction, indexing and retrieval of images and video. Pattern Recognition, 35(4), 945-965.

Athanasios.C, Gkoritsas, & Marios.C.Angelides.(2005). COSMOS-7: A video content modeling framework for MPEG-7. MMM05, 123-130.

Batini, C, Ceri, S & Navathe, S.B.(1992). Conceptual database design: An entity relationship approach, Benjamin/Cummings publishing company.

Chen, P.P.(1976). The Entity-Relationship model-Towards a unified view of data. ACM Transaction on Database Systems, 1(1), 9-36.

Cheng Youg & XU.De.(2003). Hierarchical semantic associative video model. IEEE conference NNSP'03, 1217-1220.

Colombo, C, Del,A, Bimbo & Pala, P.(1999). Semantics in visual information retrieval. IEEE Multimedia, 6(3),.38-53.

Dorai,C, Manthe,A, Nack,F, Rutledge,L, Sikora,T & Zettl,H.(2002). Media Semantics: Who needs it and why?. ACM conf. Multimedia, 580-583.

Egenhofer, M & Franzosa, R.(1994). Point-set topological spatial relations, Conference of GIS, 5(2),161-174.

Forouzan Golshani, Pegge Vissicaro & Yound Choon Park.(2004). A multimedia information repository for cross cultural dance studies. Multimedia Tools and Applications, 24, 89-103.

Guiness, D.M, Van, F & Harmelen(2004). OWL: Web Ontology Language Overview. W3C Recommentation, http:// www.w3.org/tr/owl-features/.

Harry.W.Agius & Marios.C.Angelides(2001). Modeling content for sematic level querying of multimedia. Multimedia Tools and Applications, 15, 5-37.

Hutchinson, A.(1954).Labanotation: A system for recording movement. London: T Art Books.

Koh, J.L, Lee,C.S &.Chen, A.L.P.(1999). Semantic video model for content based retrieval. IEEE Conference MMCS'99, 2, 472-478.

Lei Chen & Tamer Ozsu, M.(2003). Modeling video data for content based queries: Extending DISIMA image data model. MMM'03, 169-189.

Li, J.Z, Ozsu, M.T & Szafron, D.(1996). Modeling of video spatial relationships in a object oriented database management systems. IEEE Workshop on MMDMS, 124.

Martinez, J.M.(2003). MPEG-7 Overview, ISO/IEC JTC1/SC29/WG11-N4980.

Oomoto,E,& Tanaka, K.(1993). OVID: Design and implementation of a video object database system. IEEE Transaction on. Knowledge and Data Engineering, 5(4), 629-643.

Saraswathi.(1994). Bharatha Nattiya Kalai.Chennai: Thirumagal Nilayam.

Shu Ching Chen, Mei Ling Shyu, R.L.Kashyap.(2000). ATN as a semantic model for video data. Journal of Network Information Management, 3(1), 9-25.

Smeulders, A.W.M, Worring, M, Santini, S, Gupta, A &.Jain, R.(2000). Content based image retrieval at the end of early days. IEEE Transaction on. Pattern Analysis and Machine Intelligence. 22(12),.1349-1380.

Vendrig, J.W.(2002). Interactive adaptive movie annotation, Multimedia and Expo conference, 1, 93-96.

Vijay Khatri, Sudha Ram & Richard.T.Snodgrass.(2004). Augmenting a conceptual model with geospatiotemporal annotations. IEEE Transaction on Knowledge and Data Engineering, 16(11), 1324-1338.

Web of Indian Classical Dances, http:// www.narthaki.com/index.htm.

Appendix-A: List of Agents

Head	Hand	Knee	Leg	Foot	Arm
Finger	Ankle	Elbow	Heel	Lower Leg	Wrist
Toe	Hip	Shoulder	Waist	Back	Torso
Forearm	Palm	Pelvis	Thigh	Ball of Foot	Chest

Authors' Information

Balakrishnan Ramadoss - Assistant Professor in the Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India (Deemed University).

email: brama@nitt.edu.

Kannan Rajkumar - M.Phil, Lecturer of Computer Science at Bishop Heber College, Tiruchirappalli. Currently he is pursuing Ph.D. in Computer Applications at National Institute of Technology, Tiruchirappalli. e-mail: rajkumarkannan@yahoo.co.in

DIGITAL ART AND DESIGN

Khaled Batiha, Safwan Al-Salaimeh, Khaldoun A.A. Besoul

Abstract: The desire to create unique things and give free rain to one's imagination served as a powerful impetus to the development of digital art and design software. The commoner was the use of computers the wider variety of professional software was developed. Nowadays the creators and computer designers are receiving more and more new and advanced programs that allow their ideas becoming virtual reality. This research paper looks at the history of the development of graphic editors from the simplest to the most modern and advanced. This brief survey includes the history of different graphic editors' creation, their features and abilities. This paper highlights the two basic branches of graphic editors – these that are in free use and commercial graphic editors design software. The researcher selected the most powerful and influential graphic editors design software brands like Paint.NET and GIMP among free software and commercial Adobe Photoshop. This paper also dwells upon the way digital art transferred from the exclusively professional business into the hobby for ordinary users. This research paper bears implications for those who are interested in features and potentiality of most popular graphic editors design software.

Keywords: Digital Art, Graphic information, DPaint, Image Manipulation Program, Paint Shop Pro, Photopaint, Photoshop.

ACM Classification Keywords: I.4 Image processing and computer vision, !.4.1 Digitization and image capture, I.4.10 Image representation, I.5.2 Design methodology.

1. Introduction

Imagination is extremely refined work of the human mind. It is the easiest medium for to creation out of nothing. Human mind constantly works on creating something that has never existed before and does not now exist. This is the approach with which any professional creator of Digital Art, or in other words, creator of design will gain success. Digital culture is neither new nor determined by technology, but rather that technology is a product of digital culture. The term "digital" originally referred to data organized in discreet units in any system, linguistic, and numerical systems included.

Since the use of computers became an everyday occurrence the wide variety of software has been emerging to assist designers. From the simplest and primitive up towards professional graphics editors computer software has undergone the complicated evolution and development. I shall not mention vector graphics editors; however I'd prefer to concentrate on bitmap graphics editors, which are mainly used to produce images.

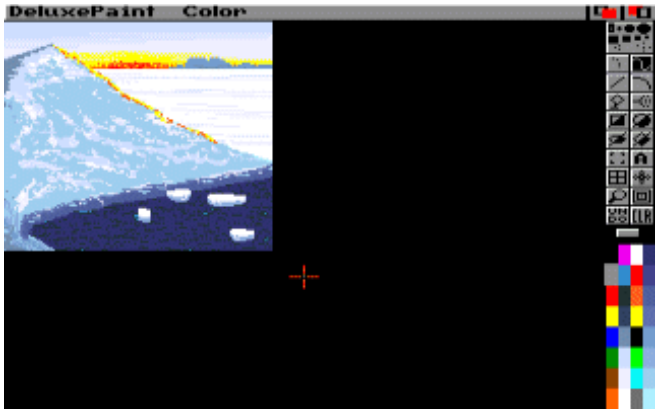
Graphic information is stored in computer memory in "bitmap" or "raster" formats such as JPEG, PNG, GIF and TIFF. Besides that, every company that creates graphics editor sets up its own format of storing raster graphics.

I would like to start the review of graphics editors with RIP editors, those that currently represent solely historical interest. Among the first editors the following deserve mentioning Deluxe Paint, Personal Paint and Photogenic.

2. Deluxe Paint

Deluxe Paint (DPaint) is a bitmap graphics editor created by Dan Silva for Electronic Arts (EA) [1]. The original version was created for the Amiga OS and was released in November 1985.

DPaint was the product of an in-house art development tool called Prism. As Silva added more features to Prism, it started to have market-place potential. When the Amiga was released in 1985, DPaint was quickly released for it. It was quickly embraced by the Amiga community and became the standard graphics development tool for the platform. Amiga manufacturer Commodore International later struck a deal with EA to have DPaint (and later its four "sequels", versions 2, 3, 4 and 5) bundled with every new Amiga sold. This deal lasted until Commodore's bankruptcy in 1994.



Screenshot and image designed in Deluxe Paint.
Taken from http://amiga.emucamp.com/dpaint4.htm.

The program DPaint enables us to create gradients, draw in anti-alias mode, change the palette, make "stencils", and transform any group of pixels into a "brush." It also allows special brush techniques "smooth" and "smear," features that are also found on Adobe Photoshop. The maximum number of colors we can work with is 256, which makes it satisfactory program for altering GIF images.

Other two programs Personal Paint and Photogenics had similar characteristics. Thus I will not dwell on them.

3. Free Graphics Design Software

Some significant position is occupied by graphics editors considered as free software. One can mention here Paint.NET i GIMP.

Paint.NET [2] is a project developed at Washington State University and mentored by Microsoft. It is a free graphics editing program for use on Windows XP and 2000 based operating systems, with the source freely available for download. It is programmed in C# and is released under the open source MIT License. Paint.NET is the unofficial successor to the older Microsoft Paint graphics program.

Graphics editors GIMP [3] deserve more particular attention. At the same time I am going to draw up some comparison of its abilities with those of Adobe Photoshop. The GNU Image Manipulation Program or The GIMP is a bitmap graphics editor and also has some support for vector graphics. The project was started in 1995 by Spencer Kimball and Peter Mattis and is now maintained by a group of volunteers; it is licensed under the GNU General Public License.

Overview

GIMP originally stood for General Image Manipulation Program; in 1997, the name was changed to GNU Image Manipulation Program. It is an official part of the GNU project.

The GIMP can be used to process digital graphics and photographs. Typical uses include creating graphics, resizing and cropping photos, changing colors, combining images using a layer paradigm, removing unwanted image features, and converting between different image formats.

The GIMP is also notable as perhaps the first major free software end-user application. Previous work, such as GCC, the Linux kernel, and so on, were mainly tools by programmers for programmers.

Features

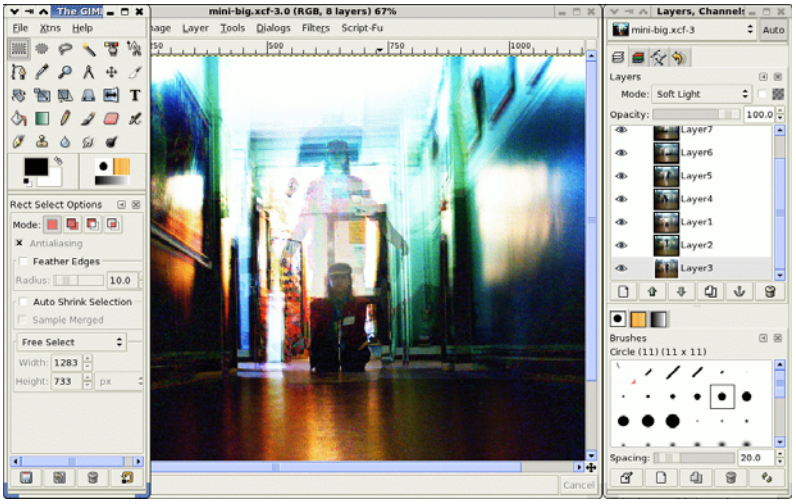
The GIMP was intended as a free (as in speech) alternative to Adobe Photoshop, but the latter still dominates in the printing and graphics industries:

- Photoshop includes licensed support for the Pantone color matching system.
- The number of plugins and other add-ons available for Photoshop is larger.
- GIMP has only experimental CMYK separation support
- GIMP has almost no spot color support.
- GIMP has limited gamma support.
- GIMP has limited color management through LCMS

There is a plugin called PSPI for the Microsoft Windows version of the GIMP only, which allows the use of the 8bf Adobe Photoshop filters in the GIMP.

The peculiarity of graphic design is the ability of graphics to interact with different projects. It is necessary to mention that such feature became indispensable at Internet technologies development. As well as interactive use, the GIMP can be automated with macro programs. The built-in Scheme can be used for this, or alternatively Perl, Python and Tcl can also be used. This allows the writing of scripts and plugins for the GIMP which can then be used interactively; it is also possible to produce images in completely non-interactive ways (for example generating images for a webpage on the fly using CGI scripts) and for batch color correction and conversion of images.

The current (as of March 2005) stable version of the GIMP is 2.2.7. Major changes compared to version 1.2 include a more polished user interface and further separation of the user interface and back-end. For the future it is planned to base GIMP on a more generic graphical library called GEGL, thereby addressing some fundamental design limitations that prevent many enhancements such as native CMYK support.



Screenshot of The GNU Image Manipulation Program 2.0.0 running on XFce on Linux

4. Adobe Photoshop

The most outstanding graphic image editor is Adobe Photoshop [4]. Adobe Photoshop is a bitmap graphics editor (with some text and vector graphics capabilities) developed and published by Adobe Systems. It is the market leader for commercial bitmap image manipulation. As with most of Adobe's other applications, Photoshop is available for Mac OS and Microsoft Windows; versions up to Photoshop 7 can also be used with operating systems such as Linux using software such as CrossOver Office. Past versions of the program were ported to the SGI IRIX platform, but official support for this port was dropped after version 3.

Features

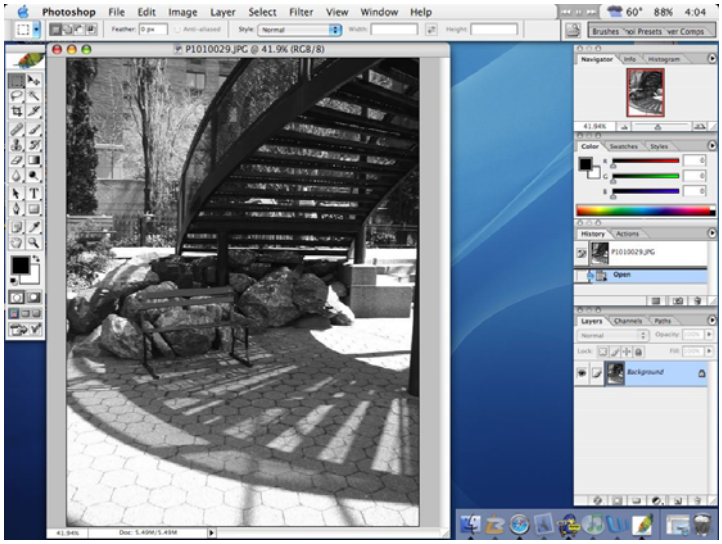
Although primarily designed to edit images for paper-based printing, Photoshop is used increasingly to produce images for the World Wide Web. Recent versions bundle a related application, Adobe ImageReady, to provide a more specialized set of tools for this purpose.

Photoshop also has strong links with software for media editing, animation and authoring. It works with Adobe Illustrator, Adobe Premiere, Adobe After Effects & Adobe Encore DVD to make professional standard DVDs, provide non-linear editing and special effects services such as backgrounds, textures and so on for television, film and the web. Photoshop's native file format (PSD or PDD) can be exported to and from Adobe Illustrator, Adobe Premiere, After Effects and Adobe Encore DVD. Photoshop CS broadly supports making menus and buttons for DVDs. For PSD or PDD files exported as a menu or button, it only needs to have layers, nested in layer sets with a cueing format and Adobe Encore DVD reads them as buttons or menus.

PSD or PDD is a widely accepted file format. Competing bitmap image editing programs (such as Macromedia Fireworks, Corel Photo-Paint, Discreet Combustion, WinImages, GIMP, etc.) can import and edit layered PSD or PDD files.

The most recent version, as of 2006, is version 9. This iteration of the program is marketed as "Photoshop CS2". In an effort to break away with previous versions of the application and to reinforce its belonging with the new line of products, Photoshop even dropped one classic graphic feature from its packaging: the Photoshop eye, which was present in different manifestations from versions 4 to 7. Photoshop CS versions now use feathers as a form of identification.

Photoshop CS features a revolutionary command: 'Shadow/Highlight' which allow user to 'suppress' highlights and/or 'push out' shadows while maintaining most of the 'image details' (i.e. the histogram would remain virtually unchanged).



Screenshot of the Photoshop CS under Mac OS X

5. From Novelty To Everyday Use

Only one and a half decade ago paint programs were treated like novelty and something very exquisite. But the development of such software turned out to be extremely rapid. Thus for comparatively short period they became an every-day need for representatives of a fine art community. Even more, the development of graphics editors caused the emergence of a new term “photoshopping”. The term photoshopping is a neologism, meaning "editing an image", regardless of the program used. The name comes from Adobe Photoshop, the image editor most commonly used for the practice, although other programs, such as *Paint Shop Pro*, *Photopaint*, or the *GIMP* may be used.

The practice of photoshopping is possible because modern image editors made the work of altering images extremely easy, particularly with the clone tool. Nowadays actually anyone who possesses elementary computer skills can use photoshopping to edit his photographs.

Professional photographers also use photoshopping in their work. Thus, practically, no photograph in the magazines Popular Photograph, Nature Photographer, Close Up and others is free of retouch. The example of professional photograph, retouched with the help of Photoshop is presented below:

Photographer Richard Seiling captured this image in Yosemite Valley, as sunlight broke through the clouds of a winter storm. Taken on 4x5 transparency film, Rich scanned the image into the computer, and performed traditional darkroom techniques, such as dodging and burning, using Adobe Photoshop. This image comes matted and overmatted on 8-ply white Archival Mat Board.

It also should be mentioned that today no well-colored magazine can do without the imaged edited in graphics editors. The results of such work are apparent everywhere – from ads in the magazines to the billboards.

Fine Art Photograph by Richard Seiling Breaking Clouds, Half Dome (20 x 24 inches)
Image is take from http://www.yosemitestore.com/

Conclusion

The present study presents a concise review of historical development of graphic editors with the particular consideration of the most representative examples. The comparative approach to the most powerful graphic editors that represent two different, in principle, branches of software - free (Paint.NET and GIMP) and commercial (Adobe Photoshop) revealed that potentialities of commercial software is still leading on software market due to its advanced and newest features that satisfy the most refined aspirations of professional users. That's why the professionals prefer Adobe Photoshop; however, the amateurs may be well satisfied by Paint Shop Pro, Photopaint, or the GIMP. The last but not the least point is that when making a choice for a particular digital art and design program one should remember the rapid progress of this kind of software and the abilities that used to be pertinent to expensive commercial products now are the characteristics of more simple free software products.

References

[1] http://amiga.emugaming.com/dpaint.html official site of Electronic Arts Company

[2] http://www.eecs.wsu.edu/paint.net/ official site of Paint.NET

[3] http://www.gimp.org/ official site of the GIMP

[4] http://www.adobe.com official site of the Adobe Company

In the research the material from the following sites was used:

http://www.guerillapixel.com/pages/digital_ill_pages/software_mainr.htm

http://screamdesign.daz3d.com/

http://www.photoshopcafe.com/PhotoshopCS2.htm

Authors' Information

Khaled Batiha – e-mail: batihakhalid@yahoo.com

Safwan Al-Salaimeh – e-mail: safwan_71@yahoo.com,

Khaldoun A.A. Besoul – e-mail: dr_khaldoun69@hotmail.com

Irbid National University, Faculty of information technology, Irbid, The Hashemite Kingdom of Jordan

CONSCIOUSNESS, SUB-CONSCIOUSNESS AND EMOTIONS, SOUL AND KDS

Zinoviy Rabinovich, Yuriy Belov

Annotation: The report discusses some neurophysiologicai and other phenomena, interpretation of which persuasively supports, although indirectly, coceptual views developed by the authors, concerning memory organization in human brain and the processes that occur in it.

Keywords: memory, soul, consciousness, subconsciousness, emotions, problem solving.

Introduction

First of all, it should be mentioned, that the terms used in the report title are meant by the authors as encyclopaedic (borrowed from the "Encyclopaedic dictionary" of 1981 edition), i. e. usual, colloquial

not subject to any modern interpretation.

This is especially important, for, as it often happens, all arguments concerning nature as natural mechanisms of subconsciousness, consciousness and soul (which is often idealistically separated from its natural essence), arise because of different understanding and interpretation of the terms. In this meaning the brain itself is the concentration of the notions expressed with the above mentioned terms. Scientific literature qualifies brain research as one of three main scientific problems of today (the other two being the State and the Universe). This research is subject to two main objectives: many-sided cognition and, accordingly, applied use (including its use in artificial intellect). Our point of view is that such bionic use is of most interest within the problems of KDS (Knowledge, Dialog, Solution) -because it consists of rational adoption of certain qualities, peculiarities and mechanisms of thinking, viewed as information processing in the brain memory (hereafter – Memory), for information mechanisms and technology. Besides, the problem of KDS in itself is of great interest for the cognition of the mechanisms of human thinking. Indeed, human thinking is defined as both processing of sensory information and as work at knowledge, represented in the language, which is the main attribute of conscious thinking. This issue will be discussed later, now we will only mention that the processes taking place in human soul, which according to the encyclopedic definition is the inner world of a person, consist exactly of a combination of processes of unconscious and conscious thinking, i. e. work at knowledge, which is represented by the first symbol K in abbreviation KDS. Furthermore, thinking as such takes place in double-sided communication of the thinking substance with the outside environment, in other words, in a dialogue with it, which is represented by the second symbol D in KDS. And, finally, the third symbol S, which implies problem solving, is the essence of human thinking (including the main component for the artificial intellect directed at solving particular problems). Thus, studying the mechanisms of thinking as one of the main problems of cybernetics (in its classical, traditional sense) in relation with the problems of KDS, turns out to be rather useful, taking into account that extraordinary circumstance,that the problems of KDS are already equipped with effective mathematical formal descriptions – the theory of representation and knowledge processing, the theory of semantic networks, the theory of probability and its application for solving the problems of model–creation and making defective decisions, defective optimal evaluation, theoretical analysis of possibilities, processing and interpretation of experiment, etc.

The authors' reports at the previous KDS conference contain the conceptual views which deal with the images recognition and problem solving in human memory [1], which are hypothetical, but at the same time supported by the mathematical analogue-computation research of cognitive processes in brain memory[2].This research was based upon the experimental data which was kindly provided by the neurophysiologist, Doctor of Biology, assistant–professor of Moscow State University, named after M. V. Lomonosov, G. S. Voronkov with whom the authors have a long–standing scientific cooperation.

In the report the authors list in addition some neurophysiological phenomena (among which are those discovered quite recently), the interpretation of which, though indirectly, but persuasively supports our conceptual views on memory organization in human brain and the processes that occur in it. This makes it possible, in our opinion, to transfer them from the category of probable and hypothetical to the category of authentic.

Consciousness, unconsciousness, emotions and soul

Without reiterating the aforementioned conceptual ideas we will recapitulate just its main essence to an extent which is necessary for further understanding and expounding the subject of this report. So, what is the brain memory (Memory) as the receptacle of stored knowledge, subconscious, conscious and soul as a whole? As it is known, Memory is only a part of the brain neuron network, the part in which the traces of incoming information, reproduced when stimulated, are formed. The rest of the neuron network of the whole organism serves for the transmission of the information to Memory, transforming accordingly the incoming signals from the environment without memorizing them. Since the information transmitted to Memory is of the enormous discretion it should be generalized, or in other words, structured, to make its representation possible. The information generalization is achieved with the hierarchical pyramidal structures which arise in the basis of the neurons and links between them, The foundations of these pyramidal structures are composed of all incoming sensory data and their apices are symbols themselves. Hence, the pyramid of the image consists of the hierarchy of the subimages which can be part of other images, which are connected with it by association. The orientation of the links determines the diffusion of stimulation, i. e. the above mentioned generalization is achieved through directing the stimulation upwards. But the presence of only such links in the pyramids of image is not sufficient and its formation in the form of the remembered structure demands the presence in the pyramids of image of the inverse, i. e. descending links from its top to the tops of the subimages to the foundation of the pyramid.

This claim arose out of the fundamental hypothesis, which states that the reproduction of the memorized image is achieved through the stimulation of the same Memory components, which have provided the image perception, i.e. the components that are in the foundation of the pyramid.

This completely particular statement, which originates, first of all, from the common sense, in its less concrete, i. e. more general aspect which is not connected to pyramidl structures, of images, can already be observed in modern literature as explanatory presence and the role of the inverse links connecting the “upper “ sections of the cerebral cortex with its “lower” sections, responsible for the perception of external information.

In general the above mentioned hypothesis (viewed by the authors as the main hypothesis) indicates the border in the neuron network of the brain between “non-memory” and Memory itself, which in this way begins where inverse descending links end. In this sense brain memory resembles computer system memory, which is the result of outlet to input shorting (through the ansas inside the memory). Stimulation of the outlets in Memory, i. e. the tops of the pyramids, spreads via inverse links and leads to the stimulation of their components including the very foundations. This process, according to our hypothesis, is the phenomenon of recollection as action, the so-called mental look. Thus, the mental look snatches out of the multitude of the limiting memory components their definite combination marked by the top of the pyramid of an image. This recollection determines recognition of the presented image as well - because if Memory contains its whole pyramid (with all upward and downward links), then stimulation, when the image is initially presented, is transmitted through upward links, but as it moves towards the top of the pyramid, it begins to be transmitted through downward links towards the foundation. Thus, the process of image recognition is characterized by dynamic composition and decomposition, which somewhat falls behind the phase. Consequently, as this process unfolds, all the components of the image, except its very top, get stimulated twice and it is the splash of the second stimulation that means the recognition of the image. If the image is not remembered in Memory, then its presentation causes only the process of composition as its formation in Memory through direct (upward) links (the brain can see, not only the eye), but the process of decomposition will not take place because of the absence of the inverse (downward) links – consequently the component of the recurring stimulation will not receive the image, i. e. the image in the whole will be perceived as new, unknown (but exactly – in the whole).

Transition of such image into a remembered one will demand its recurring presentation, in order to create inverse links through advancement of genetically inherited links (which, as it is known, was discovered in neurophysiology). Memory is formed as a result of such training. The view which was presented above by the authors of the work is in essence the interpretation of the facts already known in neurophysiology and psychology. Nevertheless, this view includes some speculative, i. e. hypothetical conclusions, connecting these facts through the cause and effect relations, that makes them convincing. At the same time, we can name some indirect arguments which prove that the views we have mentioned above are true, but cannot be verified experimentally.

Thus, for example, the active sleep stage during which dreams occur, is caused by the rapid eye movements, which appear and are directed by the special center (quadrigeminal bodies), situated on the thalamus, which receives the primary visual information as well as inverse links from the cortex, where it is processed. That is, thalamus contains the lowest layer of memory, stimulation of components of which takes place during the active sleep stage and is accompanied by REM, just like when one is awake. Dreams are appearing images; consequently their appearance proves existence of primary components of the remembered images in the lower level of memory, as well as their reproduction in the form of constituent parts in diverse combinations which represent dreams. The indicated "existence" and "reproduction" are the key aspects of the initial hypothesis.

But the process of reproduction itself (the exact way the stimulation takes place) is not subject to the given hypothesis and, as a matter of fact, is related to the production of the splash of stimulation of the image components, which leads to its reproduction.

The hypothetical assumption that the splash takes place as a result of image decomposition was proved during the analogue–computation of the process of perception of the remembered image, which was conducted using real neurophysiological data. The analogue-computation discovered the oscillatory process, which occurs only when inverse decomposition links are present [2]. But we have direct, factual confirmation of the hypothesis about the image recognition, as a result of the splash.

This confirmation is revealed when studying the phenomena which occur in polygraph (the so-called lie-detector). The tested person is asked questions about the events supposingly known to him and his physiological state is controlled with precise instruments. If the event is known to the tested it means that it is structurally stamped on his Memory. The question about this event, i. e. the introduction of the image of the event into his Memory, will provoke automatic reaction of its recognition. This reaction will manifest itself in decompositional splash. If the event is unknown, i. e. its image is absent from Memory, the splash will not occur. The event which has happened can be stamped only in subconsciousness. The presence of the splash is registered by the instruments (in particular, it becomes apparent in the changes of the encephalogram). Thus, independently from admition or denial by the tested that the supposed event took place, the truth will reveal itself in certain physiological reaction. This reaction indirectly indicates the presence of the splash during the recognition of the once again presented image which is already stamped on Memory. This serves as a proof of the true nature of the hypothesis concerning this process.

Recognition of images as the most elementary function of Memory is inseparable from this concept, and is peculiar to any organism that possesses it. But Memory is that enclosed space which contains different structured knowledge and where all thinking takes place. Thinking is understood in broad sense as the work at knowledge including its perception, analysis, accumulation, synthesis etc. It should be particularly emphasized, that the defying factor of the nature of thinking is knowledge presentation in two forms – the form of structured images and the language form. The language form is developed considerably better in Humans than in lower organisms that possess memory. It is quite possible that it is this difference of transition from quantity to quality that makes humans Human. We should emphasize that these two forms of knowledge presentation are localized in different divisions of Memory space, which are named accordingly "Sensorium" and "Language system" [3], and also in the common highest associative system which unites them. From this point we will turn directly to the subject of the report (as it is indicated in the title). First of all, it should be mentioned that the processes of thinking are divided into conscious and unconscious. Let us consider the following: what is conscious thinking, as the process of work with knowledge in Memory which occurs on the conscious level? According to the subject of the report the answer to this question should proceed, first of all, from the use of that side of the versatile notion of consciousness which refers to its structural embodiment in the brain memory, exactly of the human brain (this notion without specifications refers exactly to Humans).This structural embodiment must ensure the possibility of social communication of a Human, as his being in the world of people. The means of such information communication is Language. All the knowledge with which conscious thinking deals is expressed in Language. The Language system of Human brain contains this knowledge (and only such knowledge is covered by the encyclopaedic notion of "Knowledge"). But, as it was mentioned above all the object images remembered in the sensory memory (Sensorium) also belong to Knowledge (although not obvious knowledge). In the thinking processes the interchange of information between the sensory and language systems of the brain takes place through the transition of stimulation. Thus, we come to the following conclusion (which is an exhaustive answer to the raised question), that the process of conscious thinking takes place due to the stimulation of dynamic structures which consist of the components of both Sensorium and Language system. This process is exactly

consistent because the stimulation in the Language system leads to the pronunciation of the words. And though, this pronunciation is mental it is caused by all the necessary commands of the nervous system (which has been experimentally ascertained in neurophysiology).

This process on the conscious level, in fact, corresponds to the encyclopaedic notion of "thinking". But thinking in the broad sense of this term, as information process of work with knowledge in Memory (no matter what form they are presented in) is characterized by the processes in Sensorium which are independent of the processes of conscious thinking which necessarily involves Language system. The thinking process in sensorium is not connected with consistent pronunciation of words and can be in general deeply disparallelized or, to be more precise, distributed. In unconscious thinking (i. e. thinking without participation of the language system) this process can be of a spontaneous character and be initiated by unconscious desires.

In conscious thinking the process in sensorium gains certain purposefulness because of the functional correlation between logical and figurative thinking which prevail accordingly in Language system and Sensorium. In this case the preservation of the possibility of the disparallelized processing of knowledge is not accidental. As psychological research demonstrated, unconscious component in human thinking is approximately ten to fifteen times more prevalent than conscious. Consequently, subconsciousness on the level of work with knowledge and without exceeding the boundaries of consciousness, though it is in essense its basis, is of considerably more significance for Human thinking process. The above mentioned is especially clearly demonstrated by the phenomenon of targetl thinking as the process of solving the problem, which is set by the initial and target situations which are reflected in Human Memory as models on the conscious level. The essence of thinking according to its classical interpretation lies exactly in the solving of Problems, which are not necessarily defined specifically, as in the case above, but the problems that just mean the achievement of something desirable starting from the point of real. The considered case is the most interesting both for the cognition of human thinking and for the technical application of this knowledge, i. e. bionic approach towards the development of artificial intellect. That is why the hypothetical conceptional model of solving a Problem in the brain memory, which was first stated in 1979 [4] and then developed in further publications of Z.L. Rabinovich, including those written in cooperation with G. S. Voronkov [5] and Y. A. Belov [6], was of great interest among neurophysiologists, psychologists and cyberneticists. Let us briefly review its main theses necessary for further presentation of the report.

The initial and target situations of the Problem which is being solved (the term "model" is only implied) are expressed in Memory as stimulated structures that, as it is known from neurophysiology, due to the so-called emotion of interest (and in general satisfaction of a need) causes the process of aim (target situations) achievement. What kind of process is that? To make the answer to this question clear the notion of "Problem Generator"(PG) was introduced in the above mentioned article. PG was meant as model which unites initial and target situations which are, accordingly, its poles. Thus, PG acts as support for the Problems, which is being solved, of the potential difference on its poles (which symbolizes the above mentioned interest. The solving of the problem is achieved through the formation of the circuit, which closes initial and target situations in cause-and-effect connections. In other words the initial situation is transformed into target – in general through the creation of a line of intermediate situations.

How does this circuit appear? Initial and target situations as it was mentioned are conscious. And though they are reflected in Sensorium, they are also set in the language. This leads to logical reasoning as consecutive creation of links of circuit shorting. The shorting of the circuit possibly takes place from its both poles. Each step of such construction contains an element of conjecture as the search of the necessary information in Memory, or even as a creation of new knowledge in it That is each step like this contains a conscious question and the answer which is received as a result of conjecture. Conjecture is the manifestation of intuition.

If as a result of the whole process of solving the Problem New Knowledge is acquired and registered in the Language, this process is in fact conscious and creative.

The main role in intuitive conjecture is, apparently, played by the processes in sensorium, i. e. on the subconscious level. But after this conjecture appears it penetrates into the conscious level, making up a deficiency in PG circuit, which is perceived as striking when the conjecture has great importance. This striking will be sudden (as a remarkable psychological phenomenon, if at the moment when it occurs, PG is not stimulated, i. e. the process of solving the Problem is absent from the language system, which means lack of current conscious work at knowledge. But for this realization to occur again (i. e. for the stimulation of the PG to take place) – the stimulation of the structures in Sensorium (i. e. on subconscious level), during striking should be

so strong, so that due to the connection of Sensorium with language system it should resume stimulation of PG, inserting in it its missing link. Thus, thinking while solving the Problem involves work at knowledge the form of interaction of Sensorium and language system, i. e. on conscious and unconscious levels. This work is realized through basic operations with vertical bilateral transformation of stimulations in pyramid structures of Memory, which includes through operations connections between the structures of Memory, both vertical and horizontal the result of which is the formation of the circuit of structures which close the PG poles and which means the solving of the Problem.

This circuit – external in relation to PG, creates in this process cyclic rout of stimulation in Memory (simultaneously with stimulation of PG structures for its "support"). These cyclic routs were apparently observed in neurophysiological research as the ones that accompany thinking. Besides such indirect, in fact, experimental confirmation of the described hypothetical conceptual-modeling view on structural implementation in Memory of the process of purposeful thinking, authenticity (or let us say with caution, verisimilitude) of this view is conclusively supported with various neurophysiological phenomena. For instance, the bigger is the difference between initial and a target situation of the Problem, the more mental effort is necessary (applied to its decision through the above mentioned operations). Other examples are the phenomenon of receiving the probable solution of the Problem as sudden striking, the influence of emotional factor on the process of Problem solving (which will be mentioned later) etc.

In general, all thinking can be represented with similar models, only in more broadly defined forms, which express "desirable and real". According to the very notion of "thinking", which manifests itself as interaction of conscious and unconscious thinking the first of which is implemented in consciousness and involves the information processing in the whole Memory, and the second – in subconsciousness and is the prerogative of Sensorium and is not expressed in Language system.

Thus, consciousness is inseparable from subconsciousness, where all the sensory images are represented and subconsciousness can be interpreted as an independent subject of thinking, but working independently of it (under the influence of environment or spontaneously).

Consequently, the inner world of a Human contains both consciousness (which stands to reason) and subconsciousness, which can be also called Human soul according to the colloquial and encyclopaedic meaning of this term and also according to its materialistic interpretation (i. e. the structural organization in the brain), which is the basis of the described conception.

But the inner world of a Human, i. e. his soul contains not only consciousness and subconsciousness, as the semantic categories, but also manifestations of emotions, which appear as the result of the influence of emotiogenious organs which affect directly the physiological characteristics of Memory components with chemical mediators (the excitation thresholds of basic elements, the "conductivity" of links between them). Besides, the chemical mediators are secreted, mainly, in the lower layers of Memory (in its so-called "organic field", which is characteristic of animals, possessing the central nervous system, but in the form of its upper layers). And such state of Memory, which is localized, i. e. connected with the stimulation of its certain structures (already with the help of electric mediators) generates correspondent mood as a feeling. For example – interest, inspiration, astonishment, alarm, etc. Thus, Memory seems to have two classes of outputs – the principal semantic and emotional in which the subconscious plays the main role. On the inputs of Memory the receptor signals have the so-called "emotional colouring" which depends on the type of the influence under which they were formed and the mediator complex by which they were expressed.

Thus, thinking on the conscious and subconscious levels is not only implemented in the sphere of Memory, but undergoes the influence of the changeability of this sphere, in its turn influencing the changeability. That is the dialectics of the process.

In this meaning, the already mentioned process of purposeful thinking is the characteristic example. The emotion of interest as a moving force of the human thinking, in general has a great influence on this process this emotion leads to the increase of the "difference of potentials" of PG and "conductivity" of links between the stimulated PG structures. In such way it promotes the appearance of inspiration, as the highest emotional wholesome influence on this process.

Thus, the Human soul is the concentration of three interconnected components – consciousness, subconcsiousness and the system of emotions, each of which is highly individualized in every human, which, however, determines the importance of determining common conceptual paradigms of the inner world of a Human, both for the purpose of their cognition and for the applied use in the various fields of human activities.

According to the subject of the report, we consider the cognition of Soul from the perspective of the KDS problems. Thus, let us concentrate on the first problem of Soul which correlates with knowledge, at which both Human consciousness and subconsciousness work. In this regard, somewhat digressing from the literal expounding of the subject of the report, let us produce a sensational and interesting piece of information, which additionally proves correctness of assuming as the basis the correlation between the notions of knowledge, language, their representations and emotions.

Among all the natural organisms, only Humans possess such wealth of language and knowledge which is expressed with its help. And by definition, i. e. by agreement, this refers to the notion of Human soul. Maybe it is not right? Maybe it is a defective agreement and the human soul is something else – something independent from knowledge and language notion, which is characteristic to some extent of other organisms, which do not posses human qualities, dealing with Knowledge and the Language of their representation. So, the experiments, which were recently carried on, in Lviv revealed ("Facts" newspaper ("Fukty"), the 17th of March, year 2006), that the luminescence around the human head (aura) lasts for seventy-two hours after his death. This fact proved the results of research conducted by Professor K. Korotkov, of Saint Petersburg State Technical University of information technology, precise mechanics and optics. But this research confirmed the discovery made by K. Korotkov (which has not yet obtained general recognition of the official science), moreover they revealed that such luminescence is completely absent around the dead animals, i. e it disappears immediately after the animal's death. What does it indicate? It indicates the fact that the developed conscious thinking (as a processing of knowledge represented in language) characteristic only to Humans is supplemented with the second, characteristic only to Humans feature – aura inertia. The question naturally arises: can this aura be the material field form of the information representation of a Human soul (that, generally speaking, complies with the theory of noosphere by V. I. Vernadskiy, the phenomena of telepathy and telekineticism etc.)? But if it is so, the conducted interpretation of the notion of soul which refers only to humans, receives additional confirmation (which at a time is not really necessary for the applied technical aspects covered by KDS, but which is interesting for purposes of knowledge

Now, let us move on to the possibility of the bionical approach towards the representation and processing of knowledge in artificial intellect.

As it was mentioned above, the knowledge in Memory is represented by the complex of networks of pyramid structures which consist of two forms of neuron networks – sensory images (Sensorium) and verbal notions (Language system) which have common spheres on their higher levels. These networks are created by the basic link operations between the stimulated elements of these structures, through which the stimulation spreads. Besides, these links are vertical bilateral within each pyramid and horizontal, connecting the other pyramids. That is the neuron networks begin from the entry of information into Memory, but they do not end as separate protruding complex heads of pyramids, instead they form "web" of their interlacements. And it is in this "web" that the images light up dynamically. These images are either recognized, if the "web" contains them, or realized as new, not known yet. The declarative knowledge is concentrated mainly in Sensorium, Language system contains both declarative and operative knowledge.

The bionical approach towards knowledge representation and work on it in artificial intellect mean, first of all, the conceptual analogue-computation of Memory contents, or to be more precise – of our ideas about it. This can be achieved in two ways – through structural and programme realization. The first on because of insuperable technical and information-technological difficulties is practically possible only for limited specialized application. The second one, which can prove not to be so effective in work as the first one, is, in fact universal, restricted only by quantitative parameters. It is already determined by the capacity of application of calculating devices (computer engineering techniques).

Thus, it is the second universal realization which can be practically and effectively applied for solving different tasks that deal with artificial intellect (the most important of which probably are the tasks of problem solving). But it is implemented as analogue-computation of structures and processes only in the Language system of Memory – because all the information processed in the programme must have verbal representation.

Certainly, in this case the mathematical apparatus is indispensable for the representation of the mentioned network of pyramid structures and operations in it. The above mentioned growing pyramidal networks (GPN) can effectively serve as a basis for such apparatus. Indeed, in these networks, the images, as the combination of signs are presented by correspondent pyramids, the connections between which are implemented by common associative elements and through the transmission of stimulation. The mathematical apparatus of GPN has

already been successfully applied for solving a number of practical tasks – the synthesis of chemical compounds, classification, recognition by analogy and a number of others, which was stated in the known publications by V. P.Gladun and the members of his scientific school. This apparatus was approved both for the analogue-computation of some brain processes (namely, the model of aforementioned process of image recognition), and (with appropriate development) in the creation of the methods of designing highly productive computers on the basis of bionical approach [7, 8].

Lately, not without the influence of the demands that arise because of the conditions of the development of this approach, the GPN apparatus was broadened due to the introduction in it of the procedures of decomposition (also known as deductive and divergent), and of the procedures of composition (also known as deductive and divergent),anc of the procedures of composition(also known as inductive and convergent).. The creation or disparallelized variants of solving the problems of artificial intellect is also desirable. These variants should include the procedures based on bionic approach, which are oriented at the realization of these variants based on the super-highly productive multimicroprocessing computers (with cluster architecture), in which, as we can surmise with a certain degree of assurance [9], the sphere of thinking should be effectively reflected. Such machines, like the first models of the Ukrainian line of cluster supercomputers have already been created[10] the development of this line continues [8,9].

Let further development in the direction be perceived as the epitaph to the main designer of these machines, an outstanding scientist, the participant of previous KDS conferences, Professor Valeriy N. Koval who passed away so suddenly.

Literature

[1] Rabinovich Z.L.: Conceptual representation about pattern recognition and problem solving in human memory and possibility of its using in artificial intelligence.// KDS-2005 Proceedings. - Volume I. - XI Conf. - June 20-30, (2005) Varna, Bulgaria. - p. 1-8(in Russian).

[2] Yuriy A. Belov, Sergiy V. Tkachuk, Roman V. Iamborak. Mathematical and Computer Modelling and Research of Cognitive Processes in Human Brain. Part II. Applying of Computer Toolbox to Modelling of Perception and Recognition of Mental Pattern by the Example of Odor Information Processing, XI-th International Conference KDS-2005 Proceedings V.1, FOI-Commerce, Sophia, 2005, pp. 32-36.

[3] Voronkov G.S., Rabinovich Z.L.: Sensor and lingual systems - two form of knowledge representation. // News of artificial intelligence. (1993) p. 116-124 (in Russian).

[4] Rabinovich Z.L.: A particular bionic approach to structural modelling of purposeful thinking. // Cybernetics (1979) No 2. - p. 114-118 (in Russian).

[5] Rabinovich Z.L., Voronkov G.S.: // Knowledge representation and processing in interaction of sensor and language human neuro-system. // Cybernetics and system analysis (1998) No 2. - p. 3-11 (in Russian).

[6] Zinoviy L. Rabinovich, Yuriy A. Belov. Conceptual Idea of Natural Mechanism of Recognition, Purposeful Thinking and Potential of Its Technical Application. Mechanisms, Symbols and Models. First Intern Work-Conference on the Interplay , Spain, June, 2005, Proceedings, Part I.

[7] Zinoviy L. Rabinovich, V. A. Yaschenko. One approach to process of thinking modelling based on growing neuronal networks // Cybernetics and system analysis (1996), No 5, p. 3-20 (in Russian).

[8] A. A. Morozov, V. A. Yaschenko. PC intellectualisation based on new class of growing neuronal networks // NAN Ukraine 1997, 125 (in Russian).

[9] Zinoviy L. Rabinovich.: On concept of machine Intelligence and its development. // Cybernetics and system analysis (1995) No 2. - p. 163-173 (in Russian).

[10] Valeriy Koval, Sergey Ryabchenko, Volodimyr Savjak, Ivan Sergienko, Anatoliy Yakuba, XI-th International Conference KDS-2005 Proceedings V.1, FOI-Commerce, Sophia, 2005, pp. 98-103.

[11] Yuriy A. Belov, V. P. Didenko, O. E. Citritskiy and others. Mathematical support of complicated experiment, Kyiv: «Naukova dumka», 1982-1990, vol.1-5 (in Russian).

Authors' Information

Zinoviy Rabynovych – professor, doctor of technical sciences, Academician Glushkov Institute of Cybernetics, Academician Glushkov avenue 40, Kyiv-187, 03680 Ukraine, e-mail: : eco@public.icyb.kiev.ua

Yuriy Belov – professor, doctor of physical-mathematical sciences, Taras Shevchenko National University in Kyiv, Ukraine, 03680, Kyiv - 680, Academician Glushkov Avenue 2, building 6, e-mail: belov@ukrnet.net

DECISION–MAKING SUPPORT SYSTEMS AS PERSONAL INTELLECTUAL DEVICE OF A DECISION–MAKER

Alexey Voloshin

Abstract. The problems of the cognitive development of subject "perception" are discussed in the thesis: from the object being studied and means of action till the single system "subject – modus operandi of subject – object". Problems of increasing adequacy of models of "live" nature are analyzed. The concept of developing decision-making support systems as expert systems to decision-making support systems as personal device of a decision-maker is discussed. The experience of the development of qualitative prediction on the basis of polyvalent dependences, represented by a decision tree, which realizes the concept of "plural subjective determinism", is analyzed. The examples of applied systems prediction of ecological-economic and social processes are given. The ways of their development are discussed.

Keywords: practical knowledge; modus operandi; artificial intelligence; decision-making support systems; a decision tree; expert systems; decision-maker.

Pre-introduction

The author agrees with N. Bohr's thought, that "it's difficult to predict something, especially future". Why? It happens so, as representatives of the most theories, teachings and faiths reply that everything depends on non-controlled forces of outer world. It doesn't play a particular role how these forces are "shown up".

The representatives of "stochastic uncertainty" consider, that cause influence on consequence is determined as "objective probability" which comes to a person in the frequency of events occurrence. All this has the only sense under unlimited number of trials. A. Einstein (and we with him) objects to them: "I'll never believe that God himself plays dice".

Deterministes, on the contrary confirm that consequence is defined by the cause unambiguously, "written on Heavens", though "God's ways are inscrutable".

Christian outlook admits "local" freedom of deliberate choice that means "does everything you should".

Atheistic doctrine suggests total dependency of future on human actions in the past and present. The main thing is to act in accordance with "objective laws", which "are opened" to the man in the practical activity process.

Many eastern esoteric teachings confirm that future is formed by human's wishes and will, which determine in their turn whether to accept or not the decisions, on the basis of which the situations with a person may appear. Though, contrary to the previous, "practical knowledge" is not the knowledge of activity management, but "knowledge of managing personal will".

Introduction

What is the role of "scientists" in the structure of modern mankind, studying the world and his place in this world? In particular, those who deal with the problems of "artificial intelligence"? While analyzing this question, first of all it's necessary to define the notion. According to George Luger [Luger, 2003, p.781] artificial intelligence (A.I.) to a less degree represents the theory of regularities, which is in the basis of logical behavior extending empirical methodology of creation and researching of models, which this theory is guided by. "The researchers gradually understood that intellectual programs should be inserted in topic field, but not cherished in the laboratory" [Luger, 2003, p.802]. The aim of scientists is to create instruments and "practical knowledge" thus this "knowledge is neither object nor subject" but "modus operandi of a subject" [Luger, 2003, p.806].

Narrowing the notion "object", "subject" and "modus operandi of a subject" to their narrow natural understanding, one may confirm, that the development of science of "new time" (from Decart, Newton, Leibnitz) resulted out of

isolated object studying (in the first turn on the basis of mechanics modules) and subject (psychology, sociology) and to some extend their interdependency (mathematical biology, mathematical economic theory) till their interaction (the way of acting subject under studying object, in particular, the object itself). In author's opinion, humanity (at any rate its "intellectual representatives") it's necessary to make one more step which is in transition to the creation of single system "subject – modus operandi of subject – object".

The peculiarities of "live nature" models

While modeling "abiocoen" processes the principles both determinism and stochastic uncertainty fully proved themselves. Thus, one may "with high degree of accuracy" (with "great probability") on the basis of mechanics laws and taking into account fortuitous perturbations to define rocket coordinates for "hours – days – years". The adequacy degree of social-economic processes modules, where the subject is "active multiplier" is incomparably lower. They have a very high degree of result uncertainty; moreover, this result often is senseless (i.e. in principal uncheckable). Thus, it refers both to normative models (which can answer the question "what should be done to achieve desirable?"), and positive ("what it will be?"). In [Voloshin, 2005] the analysis of decision-making problems is given and the means of their achievement are discussed. In particular, while creating mathematical and information models of socio-economical processes 2 main problems are emphasized – "subjectivization of objectivity" (direct consideration of subject influence on the decision-making process) and "objectivization of subjectivity" (compensation of influence of cognitive subject individual characteristics – taking into account his objective peculiarities of cognition reality).

In the given paper the emphasis is namely shifted to the "subject modus operandi" under the process of object cognition by the subject. At the same time the author tries to illustrate all approaches of "fighting with uncertainty", mentioned in "Pre-introduction".

Decision-making support system as expert system

In papers [Voloshin, Pikhotnik, 1999], [Voloshin, Panchenko, 2001], [Voloshin, Panchenko, 2002], [Voloshin, Panchenko, 2003], [Voloshin, 2005], [Voloshin, Golovnya, 2005], most of which were represented at KDS conferences, the concept of "qualitative prediction on the basis of polyvalent dependences, represented by a tree of decisions" is developed [Voloshin, Panchenko, 2002]. The basic idea of this concept is "plural subjective determinism". It's considered that corollary is determined by a set of interdependent causes, the degree of influence of which on corollary is defined "subjectively" (with expert measuring). The more parameters which "form" the cause, the better it is (for model adequacy), though this leads to complications in model analysis. Thus, appears "oath dimension", which is necessary to fight with [Voloshin, Panchenko, 2002], in particular with artificial intelligence methods.

In papers [Voloshin, Panchenko, 2003] and [Voloshin, Golovnya, 2005] the device system is described ("the aim of scientists is working out instruments", mentioned above) creating applied decision-making support systems (DMSS) in different spheres. The construction of applied DMSS is reduced to emphasizing problems and sub-problems (of tree knots) and connections between them (tree arches) by experts. Measures (probabilities) of transitions between knots are determined by experts. Fuzzy estimates of experts are assumed with the help of logical variables, described by assignment function values (by real numbers vectors from 0 till 1). Each expert assigns 3 estimates – optimistic, real and pessimistic, scalarization of which is performed with taking into account psychological expert type. This type is defined on the basis of psychological tests, inserted into system. The coefficients of "credibility", "independency" and "discretion" etc. are defined too.

The tree is composed on the base of expert's collective estimation with application of pair comparisons method. For constructing the resulting tree algebraic methods of processing expert information are used. As distance between rankings Hemming metrics and discrepancy function of objects ranks is applied. The resulting tree is defined as Cemeni-Snell's median:

$$Arg \min_{A} \sum_{i=1}^{n} d(A, A^i) \text{ or as a compromise: } Arg \min_{A} \max_{i} \sum_{i=1}^{n} d(A, A^i),$$

where A^i - matrix, assigned by the i[st] expert, in which element a_{ij} =1 then and only then when i – knot is more preferable for j – expert, a_{ji} = –1, for equal objects a_{ij} =0, a_{ji} =0.

In the case of assigned advantage in unclear form matrix elements are given by assignment functions.

For determining optimal paths in the variants of tree algorithms sequential analysis are suggested [Voloshin, Panchenko, 2002] letting to process trees with hundreds knots.

The decision tree is assigned by tables. Each table is a separate tree level; each line of a table is a separate knot at this level. Each element of line is probability with which transition from given knot to low level knot is possible. These probabilities are assigned by assignment functions, representing vectors of real numbers from 0 till 1 of any length. The table is filled with results of expert polling. The existing functions permit to add columns, lines, to assign vocabulary (which allows expert verbal values to be put in accordance with probability by assigning of definite levels), save tables in file, read tables from files.

Matrices are assigned by expert way. It's result of knot variants comparison, which may be inserted into tree. Knots are defined on the basis of matrices analysis, which are inserted into tree, and probabilities, with the help of which the transition from top level knots is possible. If the decision tree is decomposed into several sub-trees, which have similar leaves, at first probabilities of these leaves, are calculated in each of them, then probabilities for the whole tree in general are found out.

A number of applied systems is created – prediction of currency course, account of overall national product, diagnostics of cardiovascular diseases, prediction of index inflation etc. [Voloshin, Pikhotnik, 1999], [Voloshin, Panchenko, 2001], [Voloshin, Panchenko, 2002], [Voloshin, Panchenko, 2003], [Voloshin, 2005], [Voloshin, Golovnya, 2005].

At the KDS–1999 conference the result of calculation national currency course (hryvna) to 01.01.2001 was given, which was analyzed at KDS–2001. Prediction exactness turned out to be $\pm$ 2% at the time, when prediction of absolute majority of foreign and home, state and private organizations exceeded 50% (on both sides). The prediction of inflation index in Ukraine for 2005 – performed in June 2005 in diploma thesis of Satir V.V. equals 12,8%. In the budget of Ukraine was put 9,8%, official statistics on results of 2005 gives 10,5%, international organizations – 12,5-13%. High prediction accurateness, in our opinion, is stipulated by "objective" reason – accounting of a great number of heterogeneous interrelated reasons, influencing upon the result. While constructing the decision tree predicting course of exchange currency, economical, financial, political (changes in legislation, possible government retirements etc.) and other parameters were considered. These parameters characterize social-economic "state", both Ukraine and near abroad countries and the whole world system. The second reason, in author's opinion, is "subjective" – this is highly specialized expert estimation. The expert very often even doesn't aware what finally he predicts.

Developing models of qualitative prediction on the basis of decision tree have one more interesting peculiarity, which was marked so far by Popper in 1959 [Luger, 2003] – "scientific theories have to make mistakes". In other words, there should be conditions, under which the model can't approximate phenomenon successfully. It happens so, because for confirmation of model correctness any finite number of approved experiments is not sufficient. Mistakes in existing models have to stimulate further research. The considering model corresponds to this in full measure. If predicted value doesn't correspond to reality, it means that either some of factors are not taken into account (which appeared probably on the prediction interval) or the extents of parameters influences are estimated wrong (which could be done correctly in the future). So, the principle of "reason influence on the corollary" is to be rejected.

DCSS as personal intellectual device of a decision-maker

While using developed device, described in the previous chapter for developing disease diagnostics applied systems [Voloshin, Golovnya, 2005], the restrictions in adaptability of used DC as expert systems, which average experts knowledge and experience, were finally defined. Thus, while diagnosing the most difficult for psychic

diseases, “experts” from Moscow and Saint-Petersburg schools (both are world – recognized) often estimate the parameters interference degree by an alternative method. In that way we get an “average hospital temperature”.

As a result we came to the following conclusion of DMSS development – DMSS system shouldn’t be an expert system, but an intellectual booster of a decision-maker; in other words “personal intellectual device” DM. Communication with doctors-diagnosticians (particularly with our collaborator in [Voloshin, Golovnya, 2005]) convinced us, that all of them consciously or subconsciously follow “the theory of inserted and executed action” [Luger, 2003], i.e. “experience of the action” (mentioned above). The following two considerations are appropriate here. Firstly, according Worldwide Health Organization data the death caused by the wrong diagnosis placed fifth in the world. Second, from Avicenna time, who claimed that “doctor is a person, who treats from the disease and he knows very little about it; he treats with the help of drug, about which he knows even less a person, whom he doesn’t know at all”, a few things changed. What is the solution? “To model” the DM action (in this case a doctor), but not construct the model based on his knowledge. Thus, the experienced doctor considers tens and hundreds of parameters, whose correlation and interference he could not estimate in principle. Therefore, he has nothing to do but to make out “main” parameters, rejecting the “secondary” ones. But this could lead to the unpredictable results. Communication with the experienced doctor, who has observed hundreds of patients, confirmed our conception about principle of “inserted” action during diagnosing. The diagnosis is determined exactly on the basis of construction (conscious or subconscious) of the decision tree. Therefore, the task is to provide DM with a device for representation and processing of “his” decision tree. It is important to note here, that the information about the decision, tree while determining diagnosis, may be of confidential character. If obtained computer diagnosis doesn’t coincide with the intuitive idea of DM (or with the “true” diagnosis, established by pathologist), then it’s necessary to provide feedback means realization for tree correction; in particular, to elaborate efficient algorithmic procedures of tree analysis on sensitiveness. It is our first important task while developing the described approach.

Conclusions

Not rejecting the necessity of creating decision-making support systems as expert systems, the author is sure in expanding of usage sphere of decision-making support systems as “personal intellectual device systems”, “which are aimed” at concrete user. In the first turn, it’s concerned with “creative” spheres of human activity (as example medical diagnosis can be). Here the analogy with the history of appearance and development of personal computers is essential.

Bibliography

[Люгер, 2003] Люгер Ф.Дж. Искусственный интеллект. Стратегии и методы решения сложных проблем. // Москва: «Вильямс»,2003.-264с.

[Волошин, 1999] Волошин А., Пихотник Е. Экспертная система прогнозирования курса гривны // "Искусственный интеллект", 1999, №2. – С.354-359 (укр.).

[Волошин, 2001] Волошин А.Ф., Панченко М.В. Прогнозирование нестабильных процессов с помощью дерева решений на основе метода парных сравнений для анализа экспертной информации // Труды конференции «KDS-2001»,Том 1, Санкт-Петербург, 2001.-С.50-53 (англ.).

[Волошин, 2002] Волошин А.Ф., Панченко М.В. Экспертная система качественного оценивания на основе многопараметрических зависимостей // "Проблемы математических машин и систем", 2002, №2. – С.83-89 (укр.).

[Волошин, 2003] Voloshin O.F., Panchenko M.V. The System of Quality Prediction on the Basis of a Fuzzy Data and Psychografphy of the Experts // “Information Theories & Applications”, 2003, Vol.10, №3.-P. 261-265.

[Волошин, 2005] Волошин А.Ф. О проблемах принятия решений в социально-экономических системах // Труды конференции «KDS-2005»,Том 1, София, 2005.-С.205-212.

[Волошин, Головня, 2005] Волошин А.Ф., Головня В.М. Система качественного прогнозирования на основе нечетких данных и психографии экспертов // Труды конференции «KDS-2005»,Том 1, София, 2005.-С.237-243.

Authors' Information

Voloshin Alexey Fedorovitch – Kyiv Taras Shevchenko National University, the faculty of cybernetics, professor. Kyiv, Ukraine, e–mail: ovoloshin@unicyb.kiev.ua

VERIFICATION AND APPLICATION OF CONCEPTUAL MODEL AND SECURITY REQUIREMENTS ON PRACTICAL DRM SYSTEMS IN E-LEARNING

Maria Nickolova, Eugene Nickolov

Abstract: The paper represents a verification of a previously developed conceptual model of security related processes in DRM implementation. The applicability of established security requirements in practice is checked as well by comparing these requirements to four real DRM implementations (Microsoft Media DRM, Apple's iTunes, SunnComm Technologies's MediaMax DRM and First4Internet's XCP DRM). The exploited weaknesses of these systems resulting from the violation of specific security requirements are explained and the possibilities to avoid the attacks by implementing the requirements in designing step are discussed.

Keywords: Security, DRM protection, e-learning.

ACM Classification Keywords: D.4.6 Security and Protection

Introduction

"E-learning" - instructional content or learning experiences through electronic technology - is expanding rapidly and is transforming how and where students learn. E-learners can be taught in very large numbers, but also in very small classes, or even as individuals, anytime and anywhere. As a result, e-learning becomes a highly cost-effective and adaptable medium for small education and training institutions and small businesses with limited resources for large overheads. E-learning offers potentially universal access to "the best" content, regardless of location, and it can transform education and training from a passive consumption experience to a more flexible and learner-cantered one. As a consequence of these changes, a new digital educational content market is born with a new commercial approach: from the distribution and sale of tangible goods to the distribution and licensing of intangible products. This new approach has a very strong impact also on the management of the rights linked to educational content (copyright and licensing, etc.).

There are many reasons for wanting to manage the rights associated with the e-learning content. Authors and artists wish to control what is done with their creations, scholars want to ensure that they receive proper attribution, commercial enterprises wish to support business models that involve licenses and fees, and e-learners want an environment free of juridical problems and unexpected costs. Although rights themselves are not technological in nature (they are defined by laws, beliefs and practices), technology can be used to transmit, verify, interpret and enforce rights as they apply to digital content and services – this is the so called digital rights management (DRM). DRM systems are designed to provide a solution for a security problem and that is why their own security is a very important factor in that process.

Although DRM systems use various security techniques (cryptography, watermarking, fingerprinting, etc.), to evaluate the own security of an entire DRM system we need an all-encompassing security model, because the security of the individual components of a system does not guarantee the security of the system as a whole. In [1] we described the security related motives, responsibilities and goals of all participators involved with a DRM system. Based on them, we built a generic conceptual model of security-related processes of DRM implementation used in e-learning and established the core security requirements for this system. The next logical step is to verify the applicability of these security requirements in practice and to analyze their fulfillment in some real DRM implementations.

Microsoft Windows Media DRM

Windows Media DRM (WMDRM) is designed by Microsoft to provide secure delivery of audio and/or video content for the Windows Media platform. The principal scheme of WMDRM (the role of the DRM system is to deliver protected content to the user which can only be accessed with a legitimate license) matches the generic model described in [1] and includes the following components:

- Windows Media Rights Manager SDK for packaging content and issuing licenses

- Windows Media Format SDK (WMF SDK) for building Windows applications which support DRM and the Windows Media format
- Windows Media DRM for Portable Devices (WMDRM-PD) for supporting offline playback on portable devices
- Windows Media DRM for Network Devices (WMDRM-ND) for streaming protected content to devices attached to a home network

Media Player stores the DRM key in separate DLL files in a Windows directory. The encryption is placed in the blackbox.dll data file. After personalizing the first license in the data file IndivBox.key (also a DLL), a specific version of the blackbox.dll for the individual PC is written. This special version also includes the hardware ID of the PC.

There have been two well-known successful attacks on Microsoft's DRM system [2]. Both have worked in the same way. Rather than break the encryption itself, which is practically incomputable, they hook or interfere with the "black box" component as it runs to dump out the content keys or the unencrypted content from the memory.

The first attack consisted of producing an unprotected copy of the protected WMA file. For the attack to work the user must already have acquired a license for the original WMA file and Windows Media Player must be able to play it. The cracking software plays the file which usually would lead to a wave output of the file's contents to the speaker system of the PC. However, the cracking software captures this sound output, converts it back into an unprotected WMA file and saves it on the disk. Using this mechanism, the loss of quality is minimal; the only loss of information during the process occurs during the audio encoding back to WMA which is not a lossless format. Apart from that, all information is preserved, including stereo information. The resulting WMA files are not protected in any way.

This sort of attack could have been prevented if the following security requirements, we set respectively for the content creator and content publisher were met by the Microsoft DRM implementation:

- (authentication) No component of the DRM scheme sends the protected content to another component, unless the receiving component is authenticated as an official component (i.e. created by the DRM developer) and allowed to receive the content from the sending component
- (secrecy) The content is only accessible by the end user specified in the license and all information and remains secret until it has been converted into analogue form.

Further, all attacks of this type target the same part of the DRM system each and every time. This means that the attack is possible on all installed systems, which violates the requirement about damages limitation we set for the content creator (Prevent the possibility for "break once, run always").

After the first successful attack Microsoft released a new version of Windows DRM that was not vulnerable. Of course content packaged under the new system was not accessible with the old system. But soon a tool called "FreeMe" was created that attacks content protected by the new DRM version which is able to determine the keys used to encrypt the content. This is a violation of the following secrecy requirement: No secret information necessary for the operation of the components or pertaining to the content (e.g. cryptographic keys, content) can be discerned from the content creator /licensed user pair, or from the communication channel between them.

The manner in which the key could be retrieved is the same for each installed system, which again violates the requirement about damages limitation.

In early 2005 Microsoft has released a patch for their DRM system which implements measures that support the requirement about damages limitation and uses another method to hide the used keys. Patched systems are no longer vulnerable to this particular attack.

But this newest DRM technology is used by hackers in another way: to install spyware, adware, diallers and computer viruses on PCs. In 2005 two Trojans, Trj/WmvDownloader.A and Trj/WmvDownloader.B arrived on a compromised computer as a license-protected multimedia files that are either Windows Media Audio (.wma) or Windows Media Video (.wmv) formats. Usually, when someone tries to start a license-protected file but doesn't own the necessary license, WMP searches for a license on the Internet. The two Trojans pretend to download the corresponding license, but instead they redirect to a Web site (serve.alcena.com or protectedmedia.com) that loads a large quantity of modem diallers and other viruses and installs a list of more than a dozen spyware and adware programs onto the user's PC.

Fortunately, Windows Media DRM is designed to be renewable, that is on the assumption that it will be cracked and must be constantly updated by Microsoft. The result is that while the DRM scheme has been cracked several times, it usually does not remain cracked for long.

Apple iTunes DRM

iTunes is sold in the Advanced Audio Coding format (AAC) [3]. Copy protection for the iTunes Music Store's media is provided by Apple Digital Rights Management system, also known as FairPlay. The manner in which iTunes operates also conforms to our model in [1]. Content is delivered to a license creator, who binds it to a license. The user can then acquire the bound content together with an appropriate license.

Fairplay limits the usage of the media to a number of computers and iPods. The default license allows protected files to be played on up to five different computers, and to burn them to CDs. Another limitation is that the only operating systems that are accessible to iTunes are Mac OS and Microsoft Windows, which leaves some widespread systems such as Linux uncovered. It is clear that the constraints of iTunes DRM protection violate the requirement for usability and universality:

- The DRM scheme must not be too restrictive for consumers who legitimately paid for content and want to share it on several devices
- The DRM must work in disconnected environments, multiplatform and multi business models

This lack of universality has logically led to the “hacking” of Apple's FairPlay system although the audio data inside the AAC file is protected by strong encryption.

The first attack was realized by the Norwegian Jon Lech Johansen. Because the iTunes DRM scheme is not compliant to the requirement of self-protection (Besides the basic requirements valid for secure software the DRM implementation must be reverse-engineering-proof and tamper-resistant), he reverse engineered successfully Apple FairPlay and released in November 2003 QTFairUse (written in C) - an open source program which dumps the raw output of a QuickTime AAC stream to a file before it is converted into analogue form. FairPlay works by first decoding the file using a user key from iPod or the Windows system and then creating new mp3 files with the metadata intact without the DRM code. Although the resulting raw AAC files were unplayable by many media players, their raw format required only some trivial additions to convert them to an MP4 files that could be played on any computer. This attack is very similar to the first attack upon Microsoft's Windows DRM system, and the remarks we made above may apply in this case as well (this attack exploits the lack of compliance to the same requirements for authentication and secrecy).

The second attack on iTunes shows also a close resemblance to the second attack on Windows DRM. The program iTunes uses HTTP XML messages to communicate with the iTunes Music Store. The keys used by iTunes (to protect the main key for the encrypted content) are stored encrypted with a system key. The communication is encrypted using an AES128 CBC algorithm. The public disclosure of a list of common XML commands for the iTunes Music Store on the Internet quickly resulted in an exploit, enabling an attacker to compromise the encryption of protected content. The encryption key was determined to be “8a 9d ad 39 9f b0 14 c1 31 be 61 18 20 d7 88 95”. It is known how to reconstruct this key for the Windows platform and for Apple's portable mp3 player. This means that it became possible to remove the encryption from the protected content for these platforms. The considerations about the second attack on MS DRM apply here as well (this attack exploits the lack of compliance to requirements about secrecy).

2005 saw another exploit working on Apple's DRM [4] that dealt with the communication protocol between Apple iTunes and Apple iTunes Music Store. The attacking program interfaced directly with the Apple iTunes Music Store online service, bypassing the security verification within the Apple iTunes program. When content was purchased, the attacking program was able to download an unprotected file from Apple’s servers because the addition of DRM was performed in the Apple iTunes program. This attack exploits the non-compliance to the requirements about independence and authentication:

- The implementation must not rely on a trusted software component or on the user’s computer/device to perform integrity checking, decrypt the content or enforce the usage rights associated with the e-learning content.

- No component of the DRM scheme sends the protected content to another component, unless the receiving component is authenticated as an official component (i.e. created by the DRM developer) and allowed to receive the content from the sending component.

Recently, a third attack on the system was crafted by Johanson [5]. Working with Travis Watkins and Cody Brocious, he went a step further and in March 2005, introduced PyMusique, an open source iTunes music store client that exploited another hole in Apple's DRM, file encoding. Music files transferred from the iTunes music store do not get encrypted by the server, but by the user's iTunes software. PyMusique mimics the iTunes application and connects to the iTunes music store server to purchase (even to register) using real iTunes user identification. As a result, during the purchase of a song, PyMusique does not encrypt the file. Another user advantage of PyMusique is that unlike iTunes, the user can re-download the song just in case he unintentionally deleted the original file. PyMusique also performs iTunes functionalities like browsing through and previewing available songs.

Because this attack consists of replacing a system component with an attacker-supplied component it is clear that the requirement about authentication is violated (no component should send the content to another component, unless the other component identifies itself as a legitimate component).

iTunes version 4.6 was especially vulnerable to PyMusique, but Apple quickly upgraded it to iTunes 4.7. As a counter-attack, Johanson and his partners released PyMusique version 0.4, the last version for Windows and PyMusique's descendent SharpMusique (a C# port of PyMusique) [5].

Another security problem in Apple's DRM scheme is that iTunes up to version 4.7 contains a vulnerability caused by a bug in the way the software handles "m3u" and "pls" playlist files. The problem specifically exists when parsing playlist files that contain long URL file entries. Malicious playlist files can come with either the .m3u or .pls extension. Though their formats are different, the vulnerability in each is the same.

For example a malicious .pls file containing a long URL looks like this:

File1=http://[A x 3245]1234

And a malicious .m3u file with a long URL looks like this:

http://[A x 3245]1234

In both cases '[A x 3245]' represents any string of 3,245 bytes in length. Opening either a malicious playlist file on the Microsoft Windows platform will cause iTunes to crash with an access violation when attempting to execute instruction 0x34333231 (the ASCII representation of 1234). An attacker can exploit this vulnerability to redirect the flow of control and eventually execute some arbitrary code. This attack is conditioned by the violation of the requirement about safety (The specific DRM solution must be carefully checked up from the end user's security point of view).

SunnComm Technologies's MediaMax DRM

The MediaMax technology is based on the active protection method relying on software on the end-user's computer that actively intervenes to block access to the e-learning content by programs other than the DRM vendor's own software. To install the DRM MediaMax uses Windows autorun, which (when enabled) automatically loads and runs software from the disk inserted into the computer's drive. Once the DRM software is installed, every time a new CD is inserted the software runs a recognition algorithm to determine whether the disk is associated with the DRM scheme. If it is, the active protection software won't allow access to the disk, except when it originates from the vendor's own content player application. The proprietary player application, which is shipped with the disk, gives the user limited access to the content.

A security problem arose when in March 2005 MediaMax version 5.0.21 was released. Its installer sets file permissions that allow any process, user, or network client, including low rights accounts, to read, modify, delete or replace the executable code with malicious one. Granting all users "Full Control" rights to executables that will be automatically run by high rights users creates simple but serious security vulnerability. After MediaMax software installation the "SunnComm Common" directory contains the executable MMX.EXE, which inherits the weak security protections. The next time a user plays a MediaMax-protected CD, the possible attack code will be executed with almost complete control over the system. This security hole results from the violation of the safety requirement (The specific DRM solution must be carefully checked up from the end user's security point of view).

Usually, this problem could be fixed by manually correcting the wrong permissions. However, aiming to enforce the DRM protection MediaMax aggressively updates the installed player code each time the software on a protected disc autoruns or is launched manually. As part of this update, the permissions on the installation directory are reset to the insecure state. This principle violates the priority requirement (In case of contradiction between the requirements for illegitimate access prevention and these for end user security, the last must prevail).

The worst is that a malicious attack is possible even if the end user has never consented to the installation of MediaMax, which could be triggered immediately whenever the user inserts a MediaMax CD. The attacker could place a hostile code in the DllMain procedure of a code file called MediaMax.dll, which MediaMax installs even before displaying the user license agreement. The next time a MediaMax CD is inserted, the installer autoruns and immediately attempts to check the version of the installed MediaMax.dll file. To do this, the installer calls the Windows LoadLibrary function on the DLL file, which causes the DllMain procedure to execute, along with any attack code placed there. This security problem results from the violation of one of the principal security goals of the content creator (The site structure (or physical media) should not use any potentially harmful methods of content protection (active protection measures as ActiveX controls, backdoors capabilities, etc.) which could put at risk the security of end users' systems).

Fixing this problem permanently without losing the use of protected disks required installing a patch from SunnComm. Unfortunately, the patch released initially was capable of triggering precisely the kind of attack it was supposed to prevent. In the process of updating MediaMax, the patch checked the version of MediaMax.dll just like the MediaMax installer does. If this file was already modified by an attacker, the process of applying the security patch would execute the attack code.

But the worst is that the MediaMax uninstaller which is supposed to recover the end user security protection introduces another security hole. It uses a proprietary ActiveX control, *AxWebRemove.ocx*, which accepts an arbitrary *validate URL* parameter and does not check that the DLL supplied by the server at that URL is authentic. The ActiveX control itself is not removed during the uninstall process, so its methods can be invoked later by any web page without further browser security warnings. An attacker could create a web page that invokes the installation method and provides a *validate burl* pointing to a page under the attacker's control and the possibility to execute an arbitrary attacker code on the user's machine.

First4Internet's XCP DRM

This DRM technology uses an active protection method similar to the one used by MediaMax and causes similar security problems due to the violation of the same security requirements. The active protection software is installed together with a second program - a rootkit - that conceals any file, process, or registry key whose name begins with the prefix *sys*. The result is that XCP's main installation directory, and most of its registry keys, files, and processes, which begin with that prefix, become invisible to normal programs and administration tools. The rootkit is a kernel-level driver named *sysaries* that is set to automatically load early in the boot process. When the rootkit starts, it hooks several Windows system API functions: NtCreateFile, NtEnumerateKey, NtOpenKey, NtQueryDirectoryFile NtQuerySystemInformation by modifying the system service dispatch table (the kernel's *KeServiceDescriptorTable* structure) and changes them to point to functions within the rootkit, so that calls are handled by the rootkit rather than the original kernel function. The rootkit calls the real kernel function with the same parameters, filters and results before returning them to the application.

But besides the security hole it opens, the rootkit has at least two harmful side effects. First, its design, based on well-known malware methods, tends to trigger security alarms. Second, it facilitates another privilege escalation attack which allows an unprivileged application to crash the system. If a malware utility makes repeated system calls with randomly generated invalid parameters, the original Windows kernel functions handle invalid inputs correctly and the system remains stable, but with the XCP rootkit installed, certain invalid inputs result in a system crash. Users experiencing system instability due to these rootkit bugs would have great difficulty to diagnose the problem, since they likely would be unaware of the rootkit's presence.

Lessons Learned

As we have seen above, although the main goal of the DRM technology is to provide security, its practical implementations often lead to serious security problems due partly to the fact that DRM technologies are based

on secrecy and work with lack of transparency. Besides by improving strategy and technology, DRM security could be increased by decreasing attacker's motivation. Currently most DRM implementations prevent end users from playing their purchased music on any portable digital music players other than the manufacturer's ones. As a result, some end users with technical know-how are tempted to exploit any potential weakness and to break the DRM protection. Another problem is the interoperability between different DRM technologies. If any portable player could play a content encoded in any DRM format this would eliminate the need for paying customers to circumvent DRM technologies in order to play their music anyway. But the principal issue in DRM security is that although nowadays there are many practical DRM implementations there are no established security standards for protecting sensitive rights information, authenticating entities in transactions and providing data integrity. Maybe it will not be possible to create one standard, but several encompassing ones could be accomplished in the near future if the existing DRM vendors conduct a dialogue about the direction of DRM evolution and reach some agreement about the principal problems in today's DRM technologies: interoperability, use of DRM integrated hardware instead of active protection and smart design. Of course, considering security requirements during the process of establishing DRM standards is important and will ensure that end systems will be:

- specified with security in mind not as an afterthought
- trusted by the two parties – those that are interested in the protection of the content and those who use it
- relatively easy to implement
- efficient to operate and maintain
- easy to use from the end user's perspective throughout the lifetime of the protected content
- scalable
- interoperable
- adaptable to permit different security paradigms
- renewable from a security perspective

Conclusions

Like most computer systems and software, DRM technologies have their own inherent security requirements. Defining, understanding and applying these requirements are a fundamental part of establishing effective, interoperable DRM standards, technologies that implement these standards and systems that incorporate these technologies. In this paper we described and discussed the applicability of our security requirements by comparing them to four existing DRM systems. We explained the exploited weaknesses of Microsoft Media DRM, Apple's iTunes, SunnComm Technologies's MediaMax DRM and First4Internet's XCP DRM through the violation of specific security requirements and showed that they could have been identified in an early state and avoided. The generic nature of the requirements implies that they are applicable to most DRM technologies and therefore should be met by any matching systems. Further developments can lead to establishing a theoretical basis that simulates DRM functioning and evaluates how well it conforms to the principal security requirements.

References

[1] M. Nickolova, E. Nickolov, Conceptual Model and Security Requirements for DRM Techniques Used for E-Learning Objects Protection. International Journal «Information Theories and Applications», to be published

[2] T. Hauser and C. Wenz, DRM under attack: weaknesses in existing systems. Digital Rights Management: Technological, Economic, Legal and Political Aspects, volume 2770, pages 206 – 223, Springer-Verlag GmbH, November 2003.

[3] International Organization for Standardization (ISO). ISO/IEC TR 13818-5:1997/Amd 1:1999 - Advanced Audio Coding (AAC).

[4] http://www.computing.co.uk/computeractive/news/2012404/fairplay-crackers-core-itunes

[5] http://nanocrew.net/about/

Authors' Information

Maria Nickolova – National Laboratory of Computer Virology; BAS, Acad.G.Bonthev St., bl.8, Sofia 1113, Bulgaria; e-mail: maria@nlcv.bas.bg

Eugene Nickolov – National Laboratory of Computer Virology; BAS, Acad.G.Bonthev St., bl.8, Sofia 1113, Bulgaria; e-mail: eugene@nlcv.bas.bg

A MULTI-AGENT FRAMEWORK FOR DISTRIBUTED DECISION-MAKING SYSTEMS

Vira Lyubchenko

Abstract: An approach of building distributed decision support systems is proposed. There is defined a framework of a distributed DSS and examined questions of problem formulation and solving using artificial intellectual agents in system core.

Keywords: decision support system, distributed decision making.

ACM Classification Keywords: I.2.11 Distributed Artificial Intelligence

Introduction

Many definitions for decision support system (DSS) are given in research or review papers. Typical for these definitions is that they all require the involvement of computers to produce information to the decision maker. But often the complexity of a decision problem is a hindrance to the rapid development of safe and effective software for DSS. Many tasks are simply too large for one DSS and require the efforts of many DSSs for distributed decision making. Following Brehmer [Brehmer, 1991] a problem requiring distributed decision making is defined as a problem, which requires

- cooperation from a number of decision makers;
- each decision maker owns part of the resources needed to solve the problem;
- no decision maker has a complete overview of the problem as a whole, and therefore the decision makers must communicate to achieve a shared "situational awareness" with respect to the state of the task.

Framework Description

Let's define a framework of a distributed DSS, which is constructed by some domains specializing in concrete problem field. One of the ways to express the result of such knowledge structuring is a task-oriented stance [Cuena, 1999]. A task is an abstract description of how the world needs to be transformed in order to achieve to desired behavior or functionality. Problem solving methods are used to cope with the task. They indicate how a task is achieved, by describing the different steps by which its inputs are transformed into its outputs. The complexity of the decision problem required compound problem-solving methods that decompose the task into subtask. These subtasks may again be decomposed by some methods, giving rise to a task-methods-subtasks tree, whose leave are given by specialized domain. Each of them has own specialized knowledge base (KB). Each domain has a set of the agents, which are capable to simulate behavior of problem field's objects and are used only for solving tasks of this concrete problem field.

Idea of agent-based structuring integrates a collection of functionalities, achieved by the interplay two kind of knowledge: about certain problem types and about the environment in which the agent operates. By this, the agent can react to the environment situation and can interact with other agents to look for solution to its problems. The notion of agents allows a design of modules that balance two aspects:

- specialty level: it is possible to model a detailed functional decomposition by designing agents that specialized in basic functions;
- autonomy level: it is possible to integrate in an agent a significant set of the functions required for the whole application but limited in scope.

The key concept of our approach to the decision of some problem in distributed environment is analogy with group decision making in human community. For this we define an agent in the system as analogue of the man-expert specializing in the solving of the certain class of tasks. The agent can have wide, but superficial knowledge and skills within the bounds of specialization. But skills of the specialized task decision are not obligatory for it. Such type of the agents is useful, because besides participation in the problem solving it can index highly tailored agents and knowledge. Due to this opportunity such agent can delegate authorities of the problem solving to more “qualified” agents or interrogate the agents for realization of an optimal simulation step.

The set of the specialized agents and knowledge, which are stored in the knowledge base, form some area of specialization that further will be called the domain. Domain can be arranged on the separate physical devices that provide for possibility of high-speed interaction between the agents and high-speed access to data and knowledge. There are no restrictions on number of such domains in distributed DSS. The domains form common distributed system with using of Internet infrastructure. The price of data transfer in the Internet is high. But despite of it, it is necessary to realize interaction between domains of system, because the decision problems, which system needs to solve, are seldom highly tailored and exceed the limits of the domain.

The special search mechanism is determined on a network of the specialized domains. This procedure assists to the user during problem's formalization, and searches the information and agents, which are capable to represent objects of problem field and to simulate their behavior [Choi, 1999]. The idea of search consists in broadcasting search keywords of a problem to all domains. The agents of the domain having such information, form the answer from knowledge base, and send an information package with the identifier of the agent and domain, which has given this information. The user separates packages, which are interest, and thus defines the agents, with which he continues interaction in a process of decision making.

When the problem formulation is terminated, the user has some set of the agents, which he intends to use as the actors for the problem solving. The user determines the characteristics and rules of environment, in which agents will participate, sets the goals of each agent, and defines its strategies.

The characteristics, rules, the initial states of environment and termination condition are transferred to domains, whose agents take part in simulation, so the agents have access to the environment. The agents make the plan of actions according to the goal, given them, and/or make query to the knowledge base for a behavioral model, which was defined by the trainer or was made by other agents and was marked as successful. The agent analyzes the information received by sensors about a status of environment and behavior of the agents-participants, makes a decision, and prepares influence on environment. This influence is reflected on own copy of environment. The change of environment is broadcasted to the agents of other domains through communications channels. The change of environment is distributed to the agents of other domains through lines of the communications. After having received changes of environment from all agents participating in modeling, the agent passes on to the following iteration of reception of the information by sensors. Procedure reiterates. Such approach reduces amount of information transmitted between domains.

During the process of decision making the agent can consult with other agents of the domain. The agent keeps a history of environment states for improved decision making, taking into consideration features of agents' behavior. It can also transfer history to other agents, which take part in problem solving. When agent deviates from the chosen plan, decision about changing the strategy of behavior can be made.

The agents are realize with applications of neural networks and evolutionary technologies, that allows to train them for solution of some class of the tasks, and allows them to store experience and evolutionary in the process of problem solving. Besides this the agents index the information of their domain KB and other agents having the necessary information.

The agent architecture for distributed DSS in such case is built around three major components

- a perception subsystem allows the agent to be situated in the environment by data acquisition and in the society by perceiving agent messages;
- an intelligent subsystem manages the different aspects of information processing as well as full or partial decision making;
- an action subsystem enact the decisions produced by intelligent subsystem, displaying messages to the control personnel or sending messages to other agents.

The agents' dynamic beliefs about the world itself and the others are stored in the KB. We can distinguish two types of information in this KB:

- problem-solving information refers to inputs, outputs and intermediate results of tasks;
- control information specifying in an agenda what is intended to be done.

The task solution is carried out in the domains' network without separation of the agent from the domain, in the virtual environments, which is unique for each separate task. The agents interact due to the special communication protocol that makes illusion of working in common space. The stopping moment for the task solution is the moment of achievement by the environment of some state, which was defined by user as final. After that the analysis of behavior of the agents, interesting for user, is carried out.

Conclusion

This paper has outlined the potential of multiagent framework for decision support. From an abstract point of view, the concept of an agent has been used as modularization principle for the DSSs' software and knowledge. The results of such modularization are specialized domains. The presented framework is flexible and easily scalable, because domains are independence.

Bibliography

[Brehmer, 1991] Brehmer, B.: Time scales, distributed decision making and modern information technology. In (Rasmussen, J.; Brehmer, B.; Leplat, J. Eds): Distributed decision making: Cognitive models of cooperative work. Wiley, New York, 1991.

[Choi, 1999] Choi, Y.S.; Yoo, S.I.: Multi-Agent Learning Approach to WWW Information Retrieval Using Neural Network. Intelligent User Interfaces, 1999; pp. 23-30.

[Cuena, 1999] Cuena, J.; Ossowski S.: Distributed Models for Decision Support. In (Weiss, G. Ed): Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT Press, London, 1999.

Authors' Information

Vira Lyubchenko - Odessa National Polytechnic University, Shevchenko av., 1, Odessa 65044, Ukraine; e-mail: vira.lyubchenko@gmail.com

CONNECTION OF NETWORK SENSORS TO DISTRIBUTED INFORMATION MEASUREMENT AND CONTROL SYSTEM FOR EDUCATION AND RESEARCH

Sergey Kiprushkin, Sergey Kurskov, Eugene Sukharev

Abstract: The development of the distributed information measurement and control system for optical spectral research of particle beam and plasma objects and the execution of laboratory works on Physics and Engineering Department of Petrozavodsk State University are described. At the hardware level the system is represented by a complex of the automated workplaces joined into computer network. The key element of the system is the communication server, which supports the multi-user mode and distributes resources among clients, monitors the system and provides secure access. Other system components are formed by equipment servers (CAMAC and GPIB servers, a server for the access to microcontrollers MCS-196 and others) and the client programs that carry out data acquisition, accumulation and processing and management of the course of the experiment as well. In this work the designed by the authors network interface is discussed. The interface provides the connection of measuring and executive devices to the distributed information measurement and control system via Ethernet. This interface allows controlling of experimental parameters by use of digital devices, monitoring of experiment parameters by polling of analog and digital sensors. The device firmware is written in assembler language and includes libraries for Ethernet-, IP-, TCP- u UDP-packets forming.

Keywords: distributed information measurement and control system, network sensors, Ethernet Interface, client-server technology, distance education.

ACM Classification Keywords: H.3.4 Systems and Software: Distributed systems

Introduction

Up-to-date systems of experiment automation are recently built on modules of software-controlled devices or digital measurement hardware, connected to interface bus. In all cases, hardware is connected to computer with interface device.

Integration of distributed system with remote sensors is more efficient, when the network interface used. It can be built on network chip and microcontroller. By using Ethernet interface, it is possible to connect different digital and analog devices, and the connection with servers will be based on TCP/IP networks.

The goal of this work is to develop a network interface for connecting remote sensors and execution units to distributed information measurement and control system for physical experiments.

The Distributed Information Measurement and Control System

The distributed information measurement system (Figure 1) is based on client/server technology and works in the nets on the basis of TCP/IP protocol stack [Gavrilov et al, 2003] – [Kiprushkin et al, 2005].

The system provides the remote access to information and hardware resources of automation equipped working places. The access to physical equipment is provided by the equipment servers (CAMAC server, GPIB server, the server of access to the MCS-196 microcontrollers and others). The communication server integrates the whole distributed systems. Its functions are: communication with user, system monitoring, security, and proper distribution of resources in multiuser mode.

The experiment process is determined and conducted by client software running on a client computer. It is necessary to emphasize, that the managing experiment software are operated not on the remote computer (as when using Web technologies) [Barrie et al, 1996], [Зимин и др., 2002], but on the user one, connected to the system via global network.

The communication server, the equipment servers, and the client software are implemented as Java applications. The data exchange between them is based on TCP stream sockets provided by java.net package, which is included into Java API standard package. The methods of using the input-output ports for the access to the interface controllers are written in C programming language.

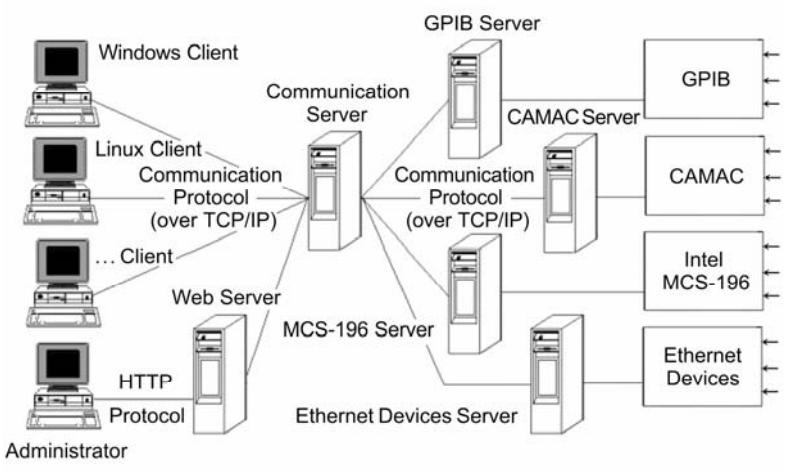


Figure 1. The scheme of the distributed information measurement and control system

Ethernet Interface

There are many specialized processors (network chips) designed for communication over networks. But it is necessary to create a central command unit, which will communicate with devices and control the network chip. This unit can be built on microcontroller.

Choosing network chip, it is necessary to take into account the physical environment and the required transmission rate of data. For communicating over 10 Mbps network, based on twisted pair, Realtek RTL8019AS processor was selected. This chip is compatible with ISA personal computer interface by timings, data and address signals.

By emulating ISA bus with microcontroller, it is possible to gain proper network chip functioning. Atmel AVR microcontrollers are good choice to implement this idea. They perform each instruction per one clock period, so their performance is 16 MIPS for 16 MHz clock rate. This performance is enough for ISA emulation.

Atmel ATMega8535 was used in the described device. It has 8-channel 10-bit ADC, 8 Kb Flash ROM, 512 b EEPROM, 512 b RAM, pulse-width modulator and analog comparator.

The main logic of the device functioning is described below. When the device is turned on, microcontroller firmware program is initiated. By sending the RESETDRV signal, the network chip is resetting. Then microcontroller configures the network chip. Configuration can be made in accordance to the desired aim of operation: e.g. reading data from measurement device and sending these data to specified network address.

The operation modes are:

- Control of the experiment execution through digital or analog devices (relays, step motors, gas injectors, etc.)
- Control of some parameters by polling analog and digital sensors (pressure, temperature, and optical sensors, atom beam sensor, etc.).
- Notify of parameter value, registering by measurement device.

Software

It is possible to present information flow as follows.

Analog value is converted into binary code by ADC. This very value must be received "on the other end of wire" for placing into database. The result is put into TCP packet. TCP protocol provides the reliable transmission of the messages between remote application processes. Then the IP datagram is formed from TCP packet (the level of the internetworking) and is sent to the bottom level – a network interface level.

Protocols of this level must provide the integration into global network: TCP/IP network must have a facility of the integration into any other networks, which doesn't depend on internal technology of data communication in these networks. Hence, this level is impossible to define once and for all. An interface facility must be designed for every communication technology. IP-frames to Ethernet encapsulation protocol pertains to such interface facility. Encapsulation of IP-package into Ethernet-frames is described in RFC1042. Then Ethernet-frame is sent via communication media.

The other side receives the frame and performs the re-conversion by correspondent server software. Processing of the frames encapsulation doesn't take much processor time in personal computer. But microcontroller has smaller speed and less memory. That's why it is very important to solve this task by means of optimized algorithm and assembler language.

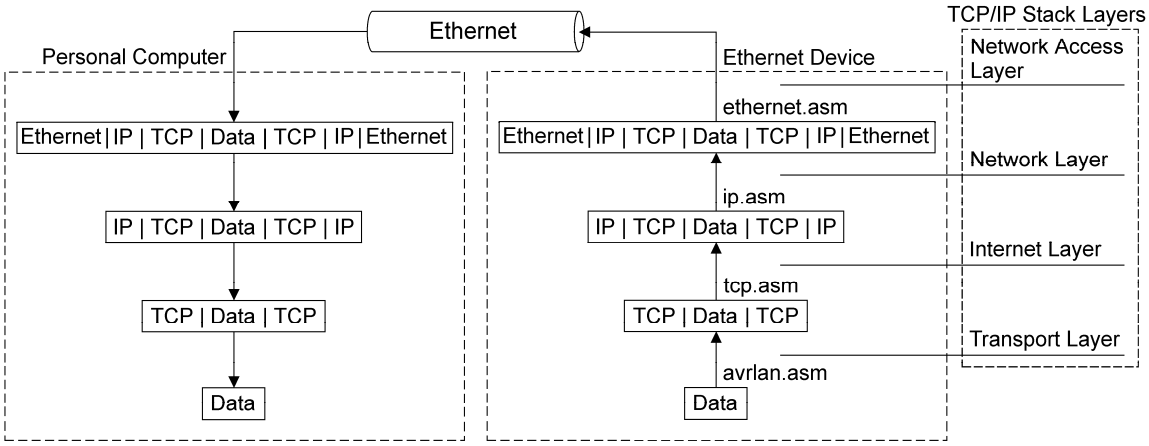


Figure 2. Set of the program modules and sequence of the transmission of the frame, received from remote device by PC

Set of the program modules and the sequence of the transmission of the frame, received from remote device by personal computer, are shown on Figure 2.

It is possible to use complete 3rd-party libraries and functions for TCP/IP implementation, because the software requires standard interactions only. These libraries are presented in all up-to-date programming environments (Java, .NET, Visual C++, Delphi, LabVIEW, etc.).

There is a question that needs to be answered by a developer of software: what level of TCP/IP stack must be implemented. This choice depends on software, used in personal computer, dataflow and processing speed of microcontroller, as well as requirements of reliability of information delivery.

The more preferred way is to use the lowest possible level of TCP/IP stack, sending data in Ethernet-, or IP-frames.

E.g. measured values of remote temperature sensor can be packed into Ethernet frames directly if qualification of the developer is sufficient to use the Ethernet level. But if the LabVIEW used, then you need to use all modules for package framing (from ethernet.asm to tcp.asm) on microcontroller side. Using the Java language for writing client applications also superimposes the restriction: when TCP-socket is used, you need to use the tcp.asm library in microcontroller. If you use UDP (unreliable delivery protocol) you must encapsulate messages with udp.asm library. This library works at transport level of TCP/IP stack that obliges to use as well as all underlayed protocols. Firmware, designed for the described device, is written on assembler language and includes Ethernet-, IP-, TCP- and UDP-package libraries.

The described device can be used in other networks, based on other protocols, but in this case it is necessary to develop the libraries for generation of correspondent frames.

Internet does not give any warranty for time of package delivery. This reason limits the use of the device in system that imposes hard time restrictions of information delivery. This feature can be eliminated by using a special network, used for undertaking the physical experiment only.

Conclusion

Ethernet interface device and corresponding software were developed and created. It implements access to remote sensor and digital device of the laboratory complex, used for scientific experiments in the field of optical spectroscopy and distant education on Physics and Engineering Department of Petrozavodsk State University.

This interface helps to increase the variety of devices, which can be connected to distributed information measurement and control system without using the computer and software-operated module electronics, as well as different instrument interface like GPIB.

Acknowledgments

We would like to express our gratitude to the laboratories' Head I. P. Shibaev for support of this work as well as students A. V. Mandychev and E. A. Vasilieva.

The research described in this publication was made possible in part by Award No. RUX0-000013-PZ-06 of the U.S. Civilian Research & Development Foundation (CRDF) and of the Ministry of Education and Science of Russian Federation.

Bibliography

[Gavrilov et al, 2003] S.E. Gavrilov, S.A. Kiprushkin, S.Yu. Kurskov. Distributed information system with remote access to physical equipment. In: Proceedings of the International Conference on Computer, Communication and Control Technologies: CCCT '03 and The 9th International Conference on Information Systems Analysis and Synthesis: ISAS'03. Orlando, 2003.

[Kiprushkin et al, 2004] S.A. Kiprushkin, N.A. Korolev, S.Yu. Kurskov. Data security in the distributed information measurement system. In: Proceedings of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics: SCI 2004. Orlando, 2004, Vol. 1, pp. 13-16.

[Kiprushkin et al, 2005] S.A. Kiprushkin, S.Yu. Kurskov, N.G. Nosovich Administration of distributed information measurement system. In: Proceedings of the 9th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2005. Orlando, 2005.

[Kiprushkin et al, 2005] S.A. Kiprushkin, N.A. Korolev, S.Yu. Kurskov. Sharing of Instrument Resources on the Basis of Distributed Information Measurement System. In: Proceedings of the Second IASTED International Multi-Conference on Automation, Control, and Information Technology - Automation, Control, and Applications: ACIT-ACA 2005. Novosibirsk, ACTA Press, 2005, pp. 170-175.

[Kiprushkin et al, 2005] S. Kiprushkin, N. Korolev, S. Kurskov, N. Nosovich. Distributed Information Measurement System for Support of Research and Education in Optical Spectroscopy. In: Proceedings of the Third International Conference "Information Research, Applications and Education": i.TECH 2005. Sofia: FOI-COMMERCE, 2005, pp. 171-179.

[Barrie et al, 1996] J.M. Barrie, D.E. Presti. The World Wide Web as Instructional Tool. Science, 1996, V. 274.

[Зимин и др., 2002] А.М. Зимин, В.А. Аверченко, С.Ю. Лабзов и др. Лабораторный практикум по спектральной диагностике плазмы с удаленным доступом через Интернет. Информационные технологии. 2002, N 3, pp. 37-42.

Authors' Information

Sergey Kiprushkin – e-mail: skipr@dfe3300.karelia.ru

Sergey Kurskov – e-mail: kurskov@psu.karelia.ru

Eugene Sukharev – e-mail: eugene-mobile@yandex.ru

Petrozavodsk State University, Lenin St., 33, Petrozavodsk – 185910, Russia

PROCEDURE OF FORMALIZATION OF THE INDICES OF BANKS' STABLE FUNCTIONING IN COMPARATIVE ESTIMATES OF THEIR DEVELOPMENT

Alexander Kuzemin, Vyacheslav Lyashenko

Abstract: The advisability of analyzing the banks liquidity and profitability as the key factor when building the comparative estimate of their functioning is considered. The procedure of formal description of the bank stable functioning indices is substantiated. Fuzzy interpretation of the bank management efficiency estimation is offered. The possibility to formalize the bank functioning estimates on the basis of the corresponding fuzzy set levels hierarchy is analyzed. The comparative estimate of different bank systems functioning is given.

Keywords: bank system, liquidity profitability, fuzzy numbers, ownership function management.

Introduction

Analysis of the financial flows of both a banking system as a whole, and separate banks specifically, is one of the key components for building of the adequate economical security system for any subject of management acting in the market economy. The reason is that it is precisely due to banks and their activity that movement and redistribution of cash and financial resources are realized. Because of this, a consideration of the problems of functioning and development of the banking system are constantly the focus of attention. This fact makes the given direction the currently central one.

The comparison estimate of different banking systems deserves particular attention. On the whole, this is favorable not only to revealing acceptable approaches to solution of different problems but the possibility of the forestalling estimates in the decision-making referring to the corresponding development of the banking system. At the same time, actually emerging difficulties in different stages of development or transformation of social-economic systems, first and foremost, call for an appropriate support of the banks stable activity basic condition which, in the general case, is expressed through establishment of the acceptable relation of liquidity and profitability. In other words, liquidity and profitability should be considered as the main components of the unified system for estimation of the financial stability and reliability of the whole bank management system. Thus, an adequate description of the liquidity and profitability interrelation is one of the key sub-problems demanding a solution.

Substantiation of the Object of the Research

The foundation for the chosen direction of researches is based as a rule on the statistical derivations or models, whose origin being in the approaches of the theory of games. But in both cases the mathematical basis of researches forms generally the probabilistic methods of data analysis. The works of E. Berglof, G. Roland, G.J. Mailath, L.J. Mester, T. Hellmann, K. C. Murdock, J.E. Stiglitz [1, 2, 3] can exemplify such a consideration of the bank activity. Nevertheless, the main problem emerging when building an adequate system of an economic

process or an object management is associated with that the economic development laws assume the presence of such interaction between different subjects of the market as well as control of the action on this interaction of different environment manifestations not having definite statistical nature in the classical sense. Therefore construction of the system for some economic process or object management requires a particular formalization taking into account not only the available statistical uncertainty but the subjective probability objectively present when making economic decisions. Solution of the given problem, in a way, is obtained through introduction of the considered bank activity indices into examination of different aspects of the information saturation. It is precisely this direction which has been chosen as the main one in the work of M. Malyutina and S. Parilova [4]. Nevertheless, the given direction contributes to solution of the set problem not in full measure as a new problem emerges which is associated with the necessity to consider ranging of different manifestations of one or other banking activity indices information saturation. Hence, the use of fuzzy sets theory approaches making it possible to describe the emerging subjective probabilities in the study of economical processes as a whole and banking activity in particular is a natural one. At the same time despite a considerable body of work in the indicated subject field of investigation [5, 6, 7, 8] the question is still open concerning the choice of the procedures relating to construction and substantiation of the type of belonging functions of fuzzy variables which are used henceforward in the appropriate models. The main reason for the given question openness is associated, first of all, with multivectorial nature of directions of the fuzzy sets theory methods use. The majority of these methods are still in the initial stage of their development. On this basis formalization of the stages of the economical parameters interrelation fuzzy description procedure characterizing parameters of the banking activity liquidity and profitability is considered as one of the goals of the given investigation. The validity of the given investigation goal choice is also associated with that in the general case the schemes and diagrams uniting the data for financial-economical activity indices of any economy subject contain a discrepant information. Therefore, the additional processing of such data should be, first of all, directed to the transformation of the financial data into information which will be useful in the process of decision-making, revealing and interpretation of the hidden tendencies.

When facing such a problem as efficiency of different economical systems comparative estimate analysis, the following should be set off:

- relative generalization of the corresponding macro indices dynamics based on the descriptive statistical data [9, 10] ,
- construction of cluster models making it possible to range the degree of development of the systems being compared. Investigations carried out by A.M. Karmisky, A.A. Peresetsky, S.V. Golovan, A.V. Kopylov [11, 12], V. Snityuk [13] can exemplify this approach.

Nevertheless, the use of the above approaches, first of all, assumes in any case the choice of the definite indices used in further comparative analysis. In this case it is necessary to define meaningful facts, to perform their ranging and only then carry out the comparative analysis, this is rather complicated problem by itself. Thus, in our opinion, it is expedient to use somewhat different approach as a preliminary analysis based on the comparative estimate of a definite index defined in terms of the fuzzy sets theory. Such an interpretation makes it possible not only to eliminate the procedure of the agreed data ranging but to perform a corresponding comparative analysis which can complete classical approaches significantly.

Classical Interpretation of the Estimate of the Bank Management Efficient with Respect to Interaction of its Liquidity and Profitability

In the commonly accepted sense the interrelation between the bank liquidity and profitability can be expressed as an inverse relationship. This fact has a very simple economical explanation. So, with an increase in the degree of the banks assets liquidity the possibility of obtaining higher profits from such assets decreases and, vice versa, less liquid assets of the bank are capable a priori of brining higher profits. Classical example of such a manifestation of liquidity and profitability in the banking activity interconnection shows that more risky credit operations can bring higher profits.

Yet, the liquidity parameters are controlled externally as a whole (from the concrete bank standpoint) and are appointed generally according to the banking system. In this case the bank profitability in many instances is associated with conducting of active-passive bank operations and in general can be characterized with a spread between its credit and deposit rates. On the one hand, the given rates in turn are governed by the classical law of

supply and demand conformity and, on the other hand, they are subject to the concurrent effect on the side of other banks.

Thus, when considering the probabilistic interpretation of banking activity management starting from a definite liquidity level one should take into account the fact that the bank tends to support the liquid assets volume at the level sufficient to ensure meeting previously taken commitments. At the same time the bank defines the probability of the need for loan resources to meet its commitments [14].

In other words, the essences of the bank management efficiency estimate in terms of its liquidity and profitability parameters can be interpreted as a probability of finding in some specified region defined with the corresponding indices of the considered parameters (Fig. 1a).

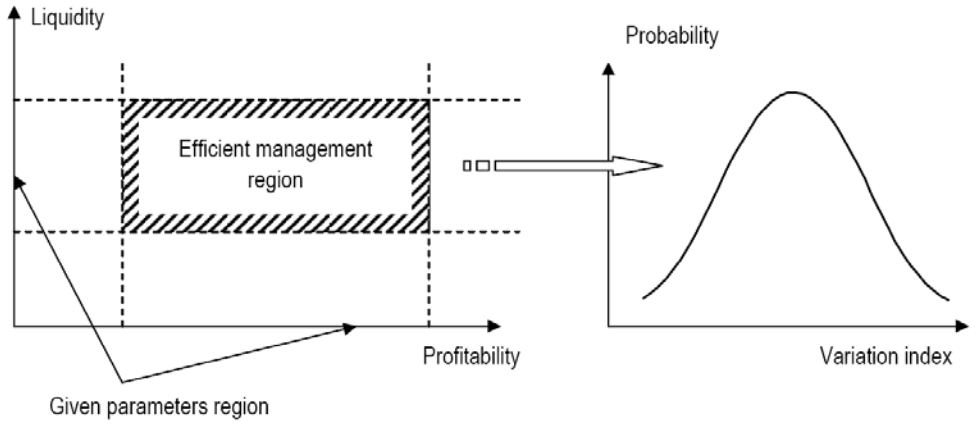


Fig.1. Essence of the bank management efficiency estimate

In this case when specifying different values for variation of the admissible liquidity and profitability values intended parameters it is possible to obtain the curve (Fig. 1b) which characterizes the bank management efficiency in terms of the possible variations of the parameters being analyzed.

At the same time the interpretation of the banking system development based on the liquidity analysis can be considered as a probability for a random two-dimensional value to penetrate into some specified field where acceptable and admissible liquidity and profitability levels parameters manifest themselves as boundaries of such a field. Just this model is applied for further analysis of different banking systems.

Fuzzy Formalization of the Bank Management Efficiency in Terms of its Liquidity and Profitability Interrelation

At the same time the classical interpretation of the bank management efficiency can be considered in terms of the fuzzy sets theory. The given approach becomes possible through introduction into consideration of the ownership function of some set of the bank liquidity and profitability indices corresponding to a subset of efficient managing actions of the given indices.

Then the fuzzy interpretation of the bank management efficiency in the specified phase space is limited to building and estimation of the corresponding ownership functions characterizing the degree of reaching the bank efficient management in the specified variation intervals of the banking activity being analyzed. In this case it is expedient to choose a fuzzy interpretation of the intended parameters variations in the limits of the admissible values of liquidity and profitability indices presented in the probabilistic model by the corresponding probabilistic curve as a formal description of such functions (Fig. 1b). The advisability of such a transition is motivated by that the fuzzy formalization of the corresponding probabilistic curve is possible on the basis of the concept of the fuzzy number of L-R type [15] which in the given case can be regarded as a trapezoidal fuzzy number (Fig. 2a).

Such an interpretation of the ownership function makes it possible not only to describe the processes under investigation formally but to take into account existing economical aspects in their development. Hence, in the given case the edges of trapezoidal representation of the ownership function for estimation of the banking activity efficiency characterize the liquidity and profitability management in terms of their boundary values. At the same time the upper base of the trapezoidal representation of the ownership function can be regarded in terms of such liquidity and profitability values which are in the center of the efficient management field (Fig. 2b).

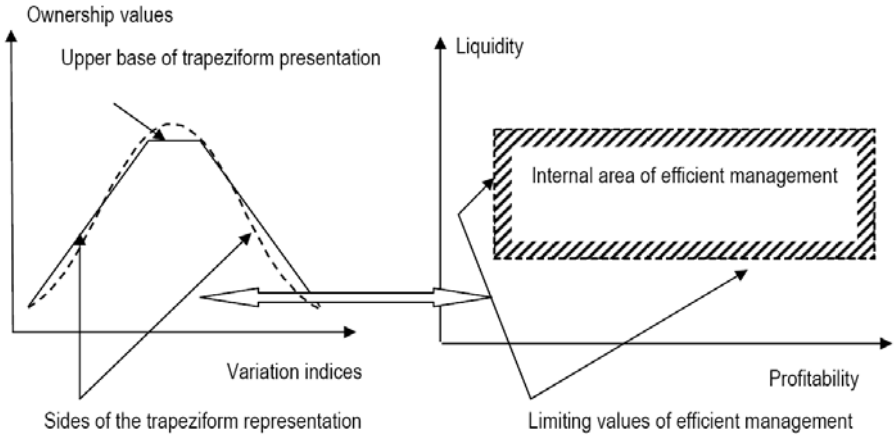


Fig.2. Interrelation between the trapeziform representation of the membership function of the bank activity and bank efficient management area indices

Levels of Bank Efficient Management Fuzzy Representation

The given approach was approved on the analysis of the real data of the banking activity liquidity and profitability indices interrelation for the banking system of Ukraine as a whole during 2003-2005. Due to such analysis different ownership functions characterizing the banking activity efficiency with regard to variations of the liquidity acceptable values for different intervals of the spread between the credit and deposit rates were built (Fig. 3, in this case the current liquidity values were analyzed).

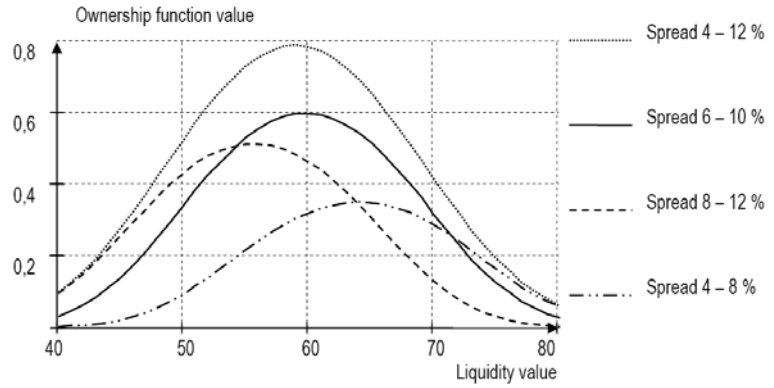


Fig.3. Ownership functions characterizing the degree of reaching the banking system efficient management on the specified intervals of the banking activity indices variations being analyzed

As may be seen from Fig. 3, the offered representation of the bank stable functioning indices formalization makes it possible to analyze informally different combinations of the considered parameters and substantiate the most acceptable ones. In the given case the spread variation between the credit and deposit rates within 4-12% and sufficient level of the current liquidity within 55-65% may be considered the most acceptable in terms of efficient functioning of the banking system of Ukraine as a whole.

At the same time the considered example led to the idea of the necessity to treat different functional representations characterizing the degree of attainment of the bank efficient management for definite values of one of the investigated parameters depending on the interval of variation of the other one. Solution of such a problem is possible on the basis of introduction of the efficient management field fuzzy set levels into consideration. In the given example it is possible to consider functional dependences of the ownership functions of the liquidity values for definite intervals of the spread variation as such levels. Then the bank management efficiency estimate can be defined on the basis of consideration of different conditions of the corresponding level subsets generalization. In this case the essential aspect of such process formalization is defined by the bank

functioning concrete conditions at different temporal stages of its activity; this can be presented as separate operations on fuzzy sets and fuzzy numbers.

Comparative Estimate of the Ukrainian and Russian Banking Systems Functioning

The corresponding similar periods' indices of Ukrainian and Russian banking systems are considered as an example of the use of the banking systems comparison probabilistic estimate based on the account of the liquidity and profitability levels interrelation. In this case the corresponding characteristics of the investigated parameters were analyzed in the assumption of the hypothesis for the presented data normal distribution. So, in the first case the probability of the optimal feasible level of the banking system profitability depending on the possible current liquidity level variation interval (Fig. 4) was considered where the value of spread between the credit and deposit rates act as the profitability. In other words, in the given case the establishment of an acceptable relation between the level of different interest rates is important, which also can be an indicator of the banking system stable development.

As can be seen from Fig. 4, the Russian banking system demonstrates the most stable state of development in terms of the possible admissible current liquidity interval. This is evident both from great values of the corresponding probability, and lesser value of spread between different interest rates level. In this case the fact is taken into account that the profitability over the banking system as a whole is defined not only by the maximum value of the interest rate spread, but the possibility to get more accessible resources and, respectively, the turnover of the value of resources passing through the banking system.

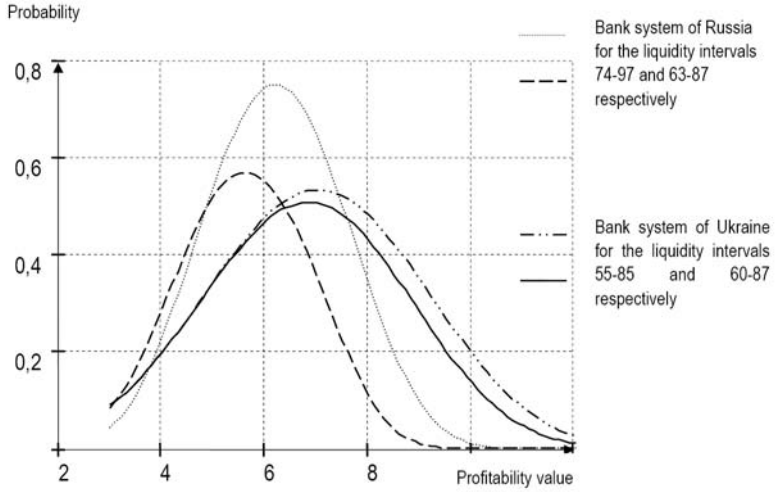


Fig. 4. Probabilistic estimate of profitability of the analyzed banking systems for different possible admissible current liquidity intervals

In the second case, Fig. 5, the probability of advisability to increase the spread between deposit and credit rates taking into account possible admissible variations of the current liquidity intervals is shown.

As can be seen from Fig. 5, the corresponding borderline values of spreads for different banking systems correlate with the data from Fig.4. At the same time the probabilistic estimate of the advisability to increase the spread for the banking system of Ukraine in some instances is greater than for the banking system of Russia. This fact can be interpreted as a great tendency of the banking system of Ukraine to the increase in spread between the credit and deposit rates. In other words, in the given aspect one can speak about less stable development of the banking system which is associated either with the risk of formation or distribution of the corresponding banking activity resource base.

At the same time, the consideration of the probabilistic and fuzzy model of interpretation of such indices of the banks functioning as liquidity and profitability leaves room for some generalization. First of all, this concerns the advisability of considering different probabilistic comparative estimates as levels of some fuzzy set. This is associated with the fact that the estimate of the banking systems functioning is more precise when considering the liquidity and profitability different levels as a single whole, their indices can be modified in the course of time under the action of internal and external factors of such economical systems development.

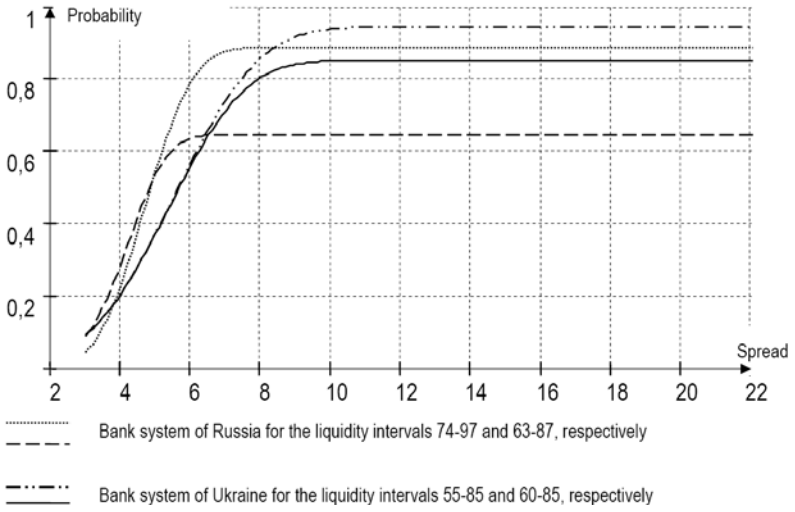


Fig. 5. Probabilistic estimate of the advisability to increase the spread between deposit and credit rates taking into account possible admissible variations of the current liquidity intervals

Conclusion

Thus, the procedure of transition from the probabilistic interpretation of the bank efficient management to its fuzzy model is considered in this work. At the same time the essence characteristics of the problem under consideration is obeyed. This makes it possible to perform the bank management efficiency analysis taking into account possible variation of its different parameters defining the bank functioning stability. At the same time the probabilistic approach of the comparative estimate of the banking system functioning makes it possible to analyze a relative functioning of different banking systems, to reveal definite features of their development, this is rather significant in terms of building an adequate system of economical safety.

Bibliography

1. Berglof , E., Roland, G. “Soft Budget Constraints and Banking in Transition Economies” J. of Comp. Econ., 26, 1998.
2. Mailath G.J., Mester L.J., “A Positive Analysis of Bank Closure”, J. of Financ. Intermed., 3, 1994.
3. Hellmann T., Kevin C. Murdock, Joseph E. Stiglitz ‘Liberalization, “Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?” 1998, Forthcoming in AER.
4. Malyutina M., Parilova S. Banks Behavior under Conditions of Transitional Period Economy: the Causes of Excessive Risks. – M.: EERC, 2001. – 52 p.
5. Rocha-Mier Luis, Villareal Francisco and Sheremetov Leonid. Agent-based Collective Intelligence for Inter-bank Payment Systems // FSSCEF 2004. – Vol. 2. – P. 321–322.
6. Canfora Gerardo, D'Alessandro Vincenzo and Troiano Luigi. Opportunities Within the New Basel Capital Accord for Assessing Banking Risk by Means of Soft Computing // FSSCEF 2004. – Vol. 2. – P. 457–465.
7. Donchenko V.S. Fuzzy Sets: Axiom of Abstraction, Statistical Interpretation, Fuzzy Sets Observations // KDS 2005. – Vol. 1. – P. 218–222.
8. Nedosekin A.O. Application of Fuzzy Models to Bank Finances Management // http://sedok.narod.ru.
9. Medvedeva E.V. Comparative Analysis of Banking Systems of Russia and Czech Republic under Conditions of Transitional Period in Economy // Proc. of Vth International Scientific and Practical Conference “Countries with Transitional Economy under Conditions of Globalization”. – M.: RUDN, 2006.
10. Drobyshevsky S., Kozlovskaya A., Levchenko D.., Ponomarenko S., Trunin P., Chetverikov S. Comparative Analysis of Monetary Policy in Transitional Economies // Proc. of IEPP. – M.: IEPP, 2003. – № 58P.
11. Kaminsky A.M., Peresetsky A.A., Golovan S.V. Models of Russian Banks Ratings. Creation, Dynamics Analysis and Comparison // Preprint #WP 2004 XXP. – M.: RES, 2004. – 56 p.
12. Golovan S.V., Kaminsky A.M., Kopylov A.V., Peresetsky A.A. Models of Russian Banks Default Probability. Preliminary Subdivision of Banks into Clusters // Preprint # 2003 XXX. – M.: RES, 2003. – 49 p.
13. Snityuk V. Evolutionary Clusterization of Complex Objects and Processes // XI-th International Conference “Knowledge-Dialogue-Solution” – Varna, 2005. – Vol. 1. – P. 232–237.

14. Shaffer S. A. Test of Competition in Canadian Banking // Journal of Money, Credit and Banking. – 1993. – Vol. 25, № 1. – P. 37–56.
15. Akhrimenko A.A., Zhelezko B.A., Ksenovich D.V., Morozevich A.N. Methods of Multilevel Aggregated Estimate and Forecasting of Enterprises Financial Position // http://sedok.narod.ru/scgroup.html.

Authors' Information

Kuzemin A.Ya. – Prof. of Information Department, Kharkov National University of Radio Electronics, Head of IMD, Ukraine, e-mail: kuzy@kture.kharkov.ua

Lyashenko V.V. – senior scientific employee Kharkov National University of Radio Electronics, Ukraine

AN ALGORITHM FOR FRESNEL DIFFRACTION COMPUTING BASED ON FRACTIONAL FOURIER TRANSFORM

Georgi Stoilov

Abstract: The fractional Fourier transform (FrFT) is used for the solution of the diffraction integral in optics. A scanning approach is proposed for finding the optimal FrFT order. In this way, the process of diffraction computing is speeded up. The basic algorithm and the intermediate results at each stage are demonstrated.

Key words: Fresnel diffraction, fractional Fourier-transform

ACM Classification Keywords: G.1.2 Fast Fourier transforms (FFT)

Introduction

The analysis of a great number of optical systems and devices requires diffraction computing under various conditions. This task can be solved through the implementation of modern methods of optical and digital image processing. Precise computing of the diffraction pattern obtained by illuminating complex transmitting objects or reflecting surfaces is a problem requiring huge computing resources. Thus, the necessity becomes obvious of introducing fast computing algorithms and of reducing the computational volume by simplifying the solution of the wave equation [1]

$$\nabla^2 \upsilon - \frac{1}{c^2}\frac{\partial^2 \upsilon}{\partial t^2} = -s\,, \qquad (1)$$

where c – speed of light, υ - a scalar quantity, describing the wave in an arbitrary point in space, $s(x,y,z,t)$ – a known function of the irradiating surface.

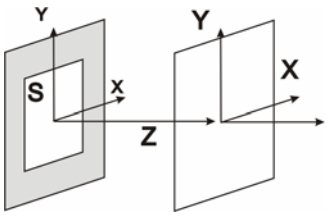


Fig.1. Irradiating and recording surface

In some cases the irradiating and the recording surface (Fig.1) can be presented as parallel planes. Most of the used approximations are based on the solution of Eq.(1) using Kirchhoff's integral [2]:

$$A(X) \approx \int_s \frac{1}{r} P(x) e^{ik(x-X)} \cos\left(\frac{\sqrt{x-X}}{r}\right) dx \qquad (2)$$

$$k = \frac{2\pi}{\lambda}, r = \sqrt{(x-X)^2 + \Delta Z^2} \ ,$$

where *P(x)* – surface irradiation function *S*; *r* - radius-vector; λ - wavelength. At that, the solution is simplified by using different assumptions. For the sake of simplicity, we consider the solution in the one-dimensional case. The most inaccurate approximation of the diffraction integral, used in optics, is replacing it with its Fourier-transform (FT) [2].

$$A(X) \approx \frac{1}{r} \int_s P(x) e^{ik(x-X)} dx \qquad (3)$$

The necessary condition for applying this approach is that the irradiating object aperture and the size of the diffraction pattern be much smaller than the distance between them. What is normally considered in optical systems is light diffraction in vacuum or light transmission through spherical optical elements (lenses) whose implementation introduces a quadratic phase multiplier in Kirchhoff's integral. In order to solve that form of the diffraction integral, FrFT introduced by Victor Namias in 1980 is successfully used [3]. It has different definitions that are proven to be equivalent and are used depending on the field of application. One of the definitions is:

$$f_a(X) \equiv \int_{-\infty}^{+\infty} \sqrt{1 - i\cot(\alpha)}\, e^{i\pi\left(\cot\alpha . X^2 - 2\csc\alpha . x . X + \cot\alpha . x^2\right)} f(x) dx \ , \qquad (4)$$

where *a* is the fractional Fourier-transform order and $\alpha = \frac{a\pi}{2}$.

The strict (accurate) proof and the conditions for its use in various optical diffraction problems is elaborated by Ozaktas, Zalevsky and Kutay [4]. The diffraction integral for the optical lens system and the propagation of the wave throgh vacuum is described in the following way:

$$\hat{h}_{lens}(x, X) = \delta(X - x) e^{\frac{-i\pi X^2}{\lambda f}}$$

$$\hat{h}_{space}(x, X) = e^{\frac{i2\pi d}{\lambda}} e^{\frac{-i\pi}{4}} \sqrt{\frac{1}{\lambda d}} e^{\frac{i\pi (X-x)^2}{\lambda d}} \qquad , \qquad (5)$$

$$\hat{g}(X) = \int_s P(x) \hat{h}(x, X) dx$$

where $\hat{g}(X)$ is the complex amplitude of the wave field in the plane of diffraction at a distance *d*, $\hat{h}_{lens}(x, X)$ is the nucleus ised in the case of using thin lenses with a focal distance *f* and $\hat{h}_{space}(x, X)$ is the wave propagation in vacuum.

Like FT, FrFT can be presented in the form of a sum instead of in the form of an integral. The main reason for that transition to discreet functions instead of continuous functions is the possibility of implementing computer processing of the digital images. FrTT can be represented by several sequential operations one of which is FT [4,6]. The natural elaboration of the approach is to seek fast computational algorithms analogous to fast Fourier-transform (FFT). An algorithm and software for fast transform of the authors referred to earlier are used for the computation of FrFT.

Problem

In a number of cases the condition for the implementation of FrFT cannot be fulfilled because of the large aperture of the object compared to the distance at which the diffracted wave is recorded. There is no analytical solution of the integral in that case. It is known that in the far field the diffracted wave function can be described by FT. The function behaviour in its intermediate states in the transition of the function to its Fourier form is represented by FrFT. Under such conditions, the solution of Kirchhoff's integral can be carried out by FrFT and find the FrFT order at which the best approximation is achieved.

Computing algorithm

In the image a line is selected that passes through some complex area, i.e. the line consists of elements of different and, if possible, large amplitude. Thus, the approximation will have a more distinct maximum with the change in the approximation parameter. For this value of the approximation parameter, Kirchhoff's integral is calculated, taking into account Eq.(2) and fast FrFT (FFrFT) (6). The aim is to find the best coincidence of the two solutions by changing the FrFT order. Using the parameter of the selected line found in this way, the FrFT for the whole image is calculated.

The control of the approximation can be carried out by selecting several lines and columns for which the parameter is calculated and its average value is obtained.

Another version of the algorithm is based on calculation of the Kirchhoff's integral followd by reverse transform of the obtained data by means of FrFT. In that case, the reconstructed and the original image can be compared more easily because, normally, there is no complex component, and in the reconstructed image the complex component is present only as a result of inaccurate approximation and calculation.

Results

In order to verify the algorithm, a simple object was chosen – a ring, (Fig. 2) with a constant value of the illuminated areas and a zero background. For the sake of simplicity, only the behaviour of the horizontal component of the diffraction was analysed. In this way, the changes in the image at the periphery of the illuminated zones are seen more clearly. A square aperture sized 102.4 μm was chosen and the discretisation step was 100 nm.

Fig. 2. Original test image

Fig. 3. Diffraction pattern in the Fresnel zone

The calculation of Kirchhoff's integral in the Fresnel zone was carried out at a distance of 10 nm. A wavelength of 533 nm was chosen (Fig.3).

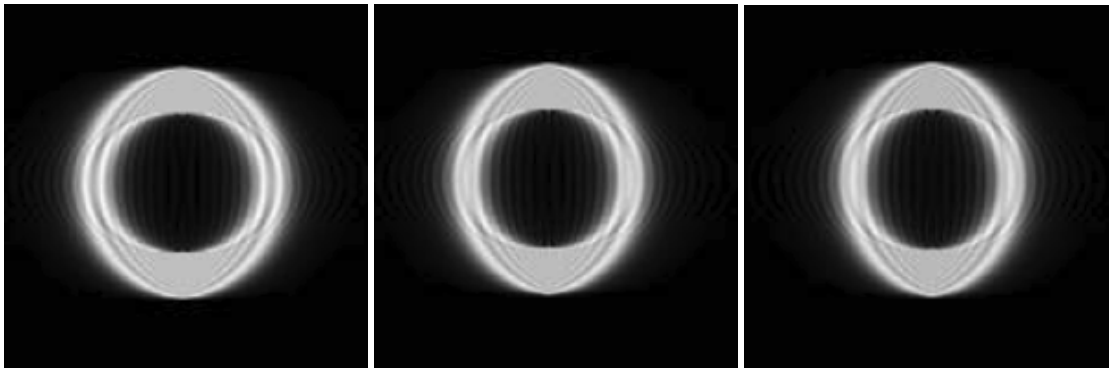


Fig. 4. Reverse transform of the lines -0.2, -0.3 and -0.34 FrFT

The reverse calculation is accomplished by seeking a solution by FrFT. The results at different values of the FrFT order are shown in Fig.4.

When the function is known, the optimisation criterion can be sought as the smallest mean square error of the difference between the original and reconstructed image. If the original is unknown and only the amplitude response is varied and the phase is kept constant, the approach of minimisation of the imaginary component of the reconstructed image can be applied. This is the case of a parallel laser beam passing through an amplitude mask.

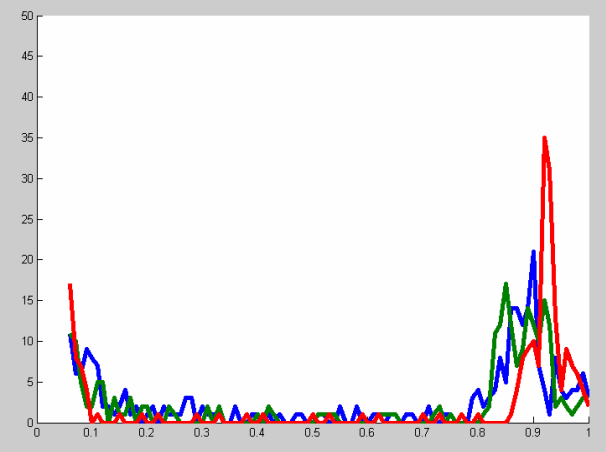


Fig. 5. Histogram of the image values for various FrFT orders

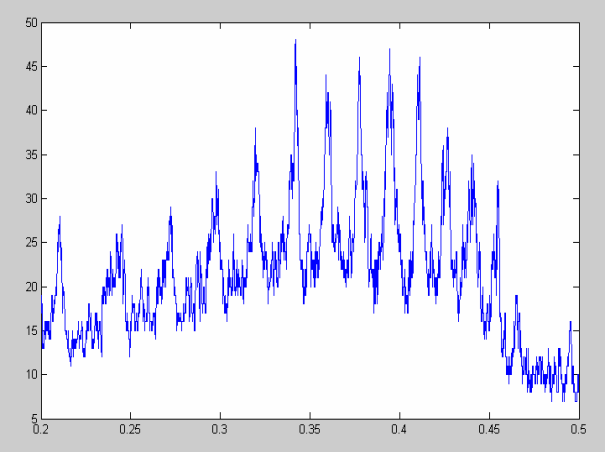


Fig. 6. Maximum in the histogram depending on the FrFT order

In the object that we selected, the amplitude has just two values: 1 and 0. Thus, the clarity of the white part of the image can be used as a criterion for successful reconstruction. Fig.5 displays a histogram of the values for some FrFT orders. In real measurements, the calculation of FrFT and the normalization of the data suppress the real value of the amplitude. If there is a solution close to the target, the values will be grouped in two areas: around 0 and around the amplitude. When normalization is accomplished after FrFT, the amplitude value is slightly below 1. The values close to zero are not shown since we are looking for a maximum close to 1.

The search for a solution of the problem begins by successive changes in the FrFT order in the range from -1 to 0. The solution is an order at which the highest maximum in the histograms in Fig.5 is achieved. Because of the periodicity of the function for searching the global maximum, scanning is possible only in the given range. The calculation of the FrFT order with an accuracy of 0.01 makes possible the implementation of fast converging algorithms, for example, the method of division of the range in two.

Fig. 7. Reconstructed image

It is seen that the maximum appears at a FrFT order of around 0.34 (exact value 0.3419). The reconstructed image at this value is shown in Fig.7. Scaling of the image is not taken into consideration in the calculation process. When the parameter takes different values, the size of the image varies. The effect exhibits itself as an image deformation in horizontal direction, since the calculations are made in that direction.

Software

A software program is developed in two parts. The programming language is Microsoft Visual C for MS WINDOWS. The first part of the program computes Kirchhoff's integral by the rectangular method. Due to the oscillating nature of the curves, the integration error is slightly higher than in the case of trapezoidal formula or other approximations of higher order. The computational process is based on a two-kernel processor Atlon 64 4400+ and operating system Windows XP. Processing of an array of 256x256 pixels takes 16 minutes, processing of an array 512x512 – 8 hours, and 1024x1024 – 120 hours.

The second module of the program takes care of the FrFT and of its order optimization. Computation of FrFT of an image of 1024x1024 pixels takes 1 minute. In the optimization process, FFrFT is used just for one line containing 1024 pixels. In this case, scanning for optimization purposes in the range from -1 to 0 with a step of 0.01 takes three seconds.

Conclusions

An algorithm is proposed for calculation of light diffraction in the Fresnel zone by finding the most suitable value of the FrFT order in one cross-section and its subsequent use for computing the whole image. Results are shown from test image processing for each stage of the algorithm. For the sake of obtaining the best visualization processing is carried out only along one axis.

References

1. Papoulis A., Systems and Transforms with Applications in Optics, McGRAW-HILL Book Company, 1968.
2. John M.Cowley, Diffraction physics, Amsterdam: North Holland, 1975.
3. V. Namias, The fractional order Fourier transform and its applications to quantum mechanics, J. Inst. Math Appl., 25, 241–265 (1980).
4. H.M. Ozaktas, M.A. Kutay, and G. Bozdagi. , Digital computation of the fractional Fourier transform. IEEE Trans. Sig. Proc., 44:2141{2150, 1996.
5. Ozaktas H., Zalevsky Z., Kutay M., The Fractional Fourier Transform with Application in Optics and Signal Processing, John Wiley & Sons, Ltd, 2001.
6. Bultheel A., K.U.Leuven H. Martnez-Sulbarany , Recent developments in the theory of the fractional Fourier and linear canonical transforms, Bulletin of the Belgian mathematical Society- Simon Stevin, 2006.

Author's Information

Georgi Stoilov – CLOSPI BAS, research associate, I degree, Sofia 1113, 101 Acad. Georgi Bontchev Street, P.O. Box 95, e-mail: gstoilov@optics.bas.bg

DEVICE FOR COUNTING OF THE GLASS BOTTLES ON THE CONVEYOR BELT

Ventseslav Draganov, Georgi Toshkov, Dimcho Draganov, Daniela Toshkova

Abstract: In the present paper the results from designing of device, which is a part of the automated information system for counting, reporting and documenting the quantity of produced bottles in a factory for glass processing are presented. The block diagram of the device is given. The introduced system can be applied in other discrete productions for counting of the quantity of bottled production.

Keywords: device for counting, automated information system

ACM Classification Keywords: J.2 Physical Sciences and Engineering

Introduction

In all discrete productions it is needed the ready production to be counted as well as reporting and documenting of the received data. In the present paper a device for counting the quantity of the produced glass bottles, moving on conveyor belt and which is designed by the authors is presented. It is a part of the automated information system for reporting and documenting of the ready production in a factory for glass processing [Draganov, 2006]. The information system has to meet following requirements: collecting data for the ready production, moving in one direction on the conveyor belts; archiving the data for each shift; reporting the quantity of the production for a shift (eight hours).

Different company developments of production counting systems are known [Solid Count, 2006; Fast Counts, 2006; Patent 0050111724, 2005]. One of them is the system SolidCount™ [Solid Count, 2006], which is designed for an automatic collecting of data for the ready mixed (of different kinds) production from a single production line, reporting the quantity of the production and receiving statistical data for the production in real time. The system Fast Count™ [Fast Counts, 2006] serves for: collecting data from several lines; reporting of the quantity of the production in different formats; monitoring of the productivity; archiving of the data; statistics and diagnostics in real time. For counting of the ready production a method and apparatus for counting is suggested in [Patent 0050111724, 2005]. The data for the ready production are received by comparison between the image of the product on programmable zoned arrays of light sources and photo detectors and known images.

The software and hardware products, which are considered, are of general use. They are expensive, very complicated and less reliable. These disadvantages are avoided in the system for counting, reporting and documenting of ready production, moving in unidirectional way on four conveyor belts as well as the entire production of the factory for glass from the four conveyor belts. The system is developed by the authors and it is introduced in a factory for glass.

Structural Diagram of the Automated Information Systems

The structural diagram of the automated information system in the factory for glass is depicted in Fig.1.

Each of the four input conversion devices (ICD) feeds an electric impulse to the device for counting control and indication (DCCI) when a ready production unit passes the conveyer belt in front of the input conversion device.

In DCCI information about the quantity of the impulses, which have come from the four ICD, is gathered. On the basis of this information the necessary signals for control of the indications I1÷I5 are depicted. In the presence of danger of overflow of any of the counters, registering the input impulses, DCCI sends a signal for overflow ($\overline{O}$) to the device for printing control (DPC). The last also receives information for the state of all counters in DCCI (Q). DPC gives a command to the printing device (PD) for printing the results and after that to DCCI – a command to clear the counters (R). The printing with consequent clearing is also accomplished by external signal from an operator through clearing button, lying on the command panel, which is a part from the DCCI, at the end of the shift. In case of power failure DPC saves the current information and after restoring the electricity supply the necessary commands for printing and clearing are passed to PD.

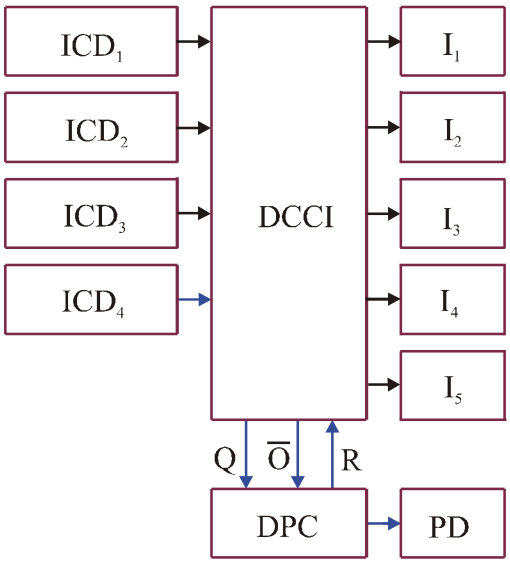


Fig.1 Structural diagram of automated system for counting of bottles on conveyor belt

Scheme Solution of the Device for Counting of Bottles

To receive reliable information for the quantity of the produced glass bottles it is necessary each input converting device from the automated information system for reporting and documenting of the quantity of ready production to be designed. The device has to meet the following requirements: to convert the information for the number of the glass bottles, which move on the conveyor belt separately or in groups in electrical impulses with TTL level in contactless way; the number of the electrical impulses to correspond strictly to the number of the passing glass bottles and errors, caused by bottles, which are contiguous one to another or by the uneven optical density of the glass from which the bottles are made or by vibrations of the conveyor belt have to be expelled; the device to be simple and cheap at most and with high reliability of the scheme solution; the construction to be with high mechanical stability and manifacturability.

The main goal of the work is to design a device, which meets the attached requirements and free of the indicated disadvantages.

Devices for counting of objects, based on electro-contactable, capacitive, inductive and other principles are known. One of the most perspective one is the photo-converting principle, which has following advantages: broad field of application; contactless way of operation; high reliability and long exploitation time; high promptitude; low feeding voltages and small consumption of electrical power; broad temperature range of operation; possibility for miniaturization and integration and etc.

The photo-converting devices frequently operate in a mode of transmission [Bergmann, 1980], in which the counted objects cross and modulate a ray, emitted from light source to a light receiver, situated on the other side of the object. There is a possibility for operating in another mode – mode of reflection [Bergmann, 1980], in which the light source and the light receiver are situated on one and the same side of the moving object, reflecting directly or diffusely part of the light, emitted by the light source to the light receiver. An operation in a mode of autonomous emitting [Bergmann, 1980] at which the object itself is a light source is possible.

The photo-converters may operate with unmodulated and modulated light [Bergmann, 1980]. The schemes of the photo-converters with unmodulated light are simplified but they are adversely influenced by the disturbing light – daylight or artificial, emitted by other sources of light. The photo converters with modulated light are protected from the influence of the disturbing light in a high degree, but their scheme solution is complicated and expensive.

In the designed device the photo converting principle of operation, based on mode of transmission of the unmodulated light is used. Thus a simplified scheme solution is obtained.

The disadvantages of principle of the devices operating with unmodulated light are not substantial in the concrete case as the application of the device to be designed is characterized by a small distance between the light source and the light receiver and lack of parasitic lighting. For the purpose an appropriate construction is developed.

The possible errors, caused by vibrations of the conveyor belt and by the uneven density of the bottles may be avoided by transmission of light ray at the height of the mouth of the bottles. But even in this case the light ray is discontinued repeatedly when a single bottle is passing and the number of the obtained output impulses is arbitrary.

Scheme solutions by which this disadvantage may be avoided – with using of integrator, by their processing with monostable multivibrator are known. The difficulty in using them in the concrete case is caused by their irregular movement of the conveyor belt because of the vibrations, which strongly hampers the specifying of the time constant of the delay circuitry.

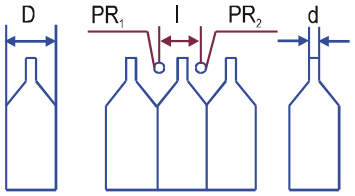


Fig.2 Scheme solution of the device for counting of bottles.

The problem is solved by using of $\overline{R}$ - $\overline{S}$ trigger, to which both inputs impulses are entering from the both photoconverters (Fig.2). Each photoconverter contains emitter and receiver. When the mouth of the bottle passes between the source and the receiver of the first photoconverter PR_1, the light ray is discontinued repeatedly. The obtained output impulses enter the first (for example ”S”) input of the trigger. The first impulse fixes a certain state - in the case logical “1” at its output and the succeeding ones do not change the output state regardless their number. When the second light ray crosses the mouth of a bottle, the obtained output impulses from the second photo receiver PR_2 enter the second (“R“) input of the trigger. The first one of them alters the output state of the trigger into logical “0” and the succeeding ones are not of importance. Thus the obtaining of only one output impulse when a bottle passes is guaranteed.

The chosen scheme solution is characterized by extremely high reliability, high stability, simplicity and lack of necessity of adjustment at producing and in the process of exploitation.

The main problem in designing of the construction is the right choice of the distance l between both photoconverters. In order the impulses not to enter the both inputs of the $\overline{R}$ - $\overline{S}$ trigger simultaneously this

distance has to be as big as possible. But its excessive augmentation would lead to errors from missing of bottles if they do not move closely one to another. From Fig.2 it can be seen that if the ray diameter is small enough following condition has to be fulfilled:

$$d < l < D \quad (1)$$

where d - maximal diameter of the mouth of the bottle; D – minimal diameter of the body of the bottle.

On the basis of the described principle the entire block scheme of the device for counting of glass bottles on the conveyor belt (Fig.3) is developed.

Two identical channels, each one including light source (LS_1 and LS_2), light receiver (LR_1 and LR_2), source of reference voltage (SRV_1 and SRV_2), comparator (C_1 and C_2) and matching device (MD_1 and MD_2) are used.

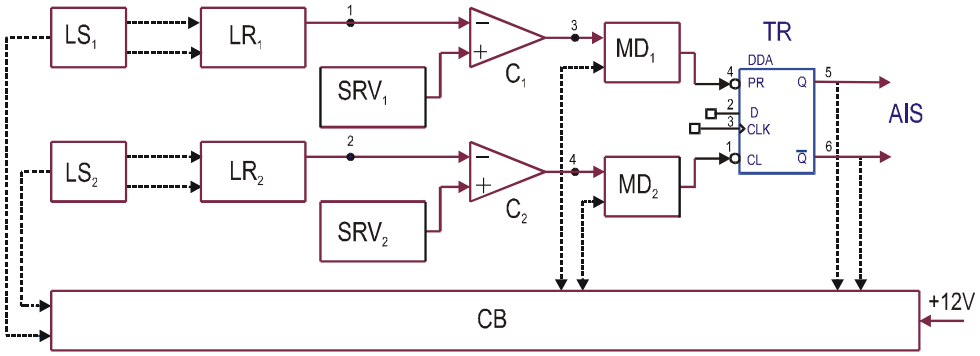


Fig.3. Structural scheme of the device for counting of bottles

The principle of operation is illustrated through the time diagram from Fig. 4. When the light receiver LR is lighted up, the voltage of the inverting input of the comparator C is higher than the reference one (Ur). The corresponding output voltage of the comparator is low. At the output of the amplifier MD a high TTL - level is obtained as the amplifier is an inverting one. When the light receiver LR_1 is shaded by a passing bottle at the output of the comparator C_1 a high level is obtained and at the output of MD_1 – low level. The $\overline{R}$ - $\overline{S}$ trigger TR is established in condition “logical 1”. When the light receiver LR_2 is shaded analogous processes occur and the trigger TR is cleared. The trigger TR eliminates the influence of the winkings.

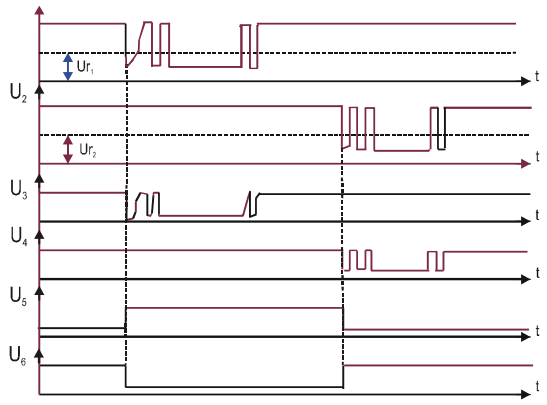


Fig.4. Operation time diagram

The device has a symmetrical output. This enables sharp decreasing of the disturbances, which may penetrate through the line, connecting the output of the device to the input of the Automated information system (AIS) as well as for possible errors, caused by the disconnecting of the connecting wires at their connecting to “ground” etc. For the purpose in the receiving block of AIS a circuitry “sum of modulus two” is connected.

A control block (CB) for diagnostics and control of the normal operation [Marinov, 1980] is provided as a part of the device and through which the good working order of the LS_1 and LS_2; the output signals of the comparators, received from MD_1 and MD_2; the signals, received from the outputs of the trigger; the presence of supply voltage are supervised.

Conclusion

The designed device is a composite part of the automated information system for control, reporting and documenting the quantity of produced glass bottles, which is introduced in the factory for glass processing in town of Elena. The device enables the counting of empty bottles, discolored or of different coloring, of different form and size. It also may be successfully applied for counting of full bottles regardless the content and it level. These qualities of device provide its comparatively wide application in different branches of industry.

Bibliography

[Draganov, 2006] Draganov, V.D. , G.P. Toshkov, D.T. hristov. System for reading and documenting of the ready production quantity, Acta Universitatis Pontica, Volume 6, Number VII, 2006

[Solid Count, 2006] www.cornerstoneautosys.com/solidcount.htm - Solid Count

[Fast Counts, 2006] http://www.accusort.com/applications/counting.html - Fast Counts

[Patent 0050111724, 2005] United States Patent Application: 0050111724, A1, May 26, 2005. Method and apparatus for programmable zoned array coundter

[Bergmann, 1980] Bergmann H. Fotoelektrische Schalter, Radio fernsehen elektronik, 12/1980, pp. 769-770, in German

[Marinov, 1980] Marinov, Ju., E. Rangelova, V. Dimitrov. Technical diagnostics of radio-electronic systems and devices, Tehnika, Sofia, 1980, in Bulgarian

Authors' Information

Ventseslav Draganov – e-mail: v2draganov@abv.bg

Georgi Toshkov – e-mail: g_toshkov2006@abv.bg

Daniela Toshkova – e-mail: daniela_toshkova@abv.bg

Technical University of Varna, 1, Studentska Str, Varna, 9010

Dimcho Draganov – "Technotrade", Varna, 9000.

ABOUT METHODS OF MATHEMATICAL MODELLING IN THE DEVELOPMENT OF INFORMATION SYSTEMS

Maria Eremina

Abstract: This article describes the approach, which allows to develop information systems without taking into consideration details of physical storage of the relational model and type database management system. Described in terms of graph model, this approach allows to construct several algorithms, for example, for verification application domain. This theory was introduced into operation testing as a part of CASE-system METAS.

Keywords: information system, database, metadata, mathematical model, graph.

ACM Classification Keywords: H.2.4 Systems - Relational databases; D.2.2 Design Tools and Techniques - Computer-aided software engineering (CASE).

Introduction

The necessity in development of large information heterogeneous system is essentially increasing now. As is well known, *information system* is a complex of information recourses, technologies of getting and processing data, and keeping it in actual and consistent state. This definition is formulated from the point of user's view, but from the realization point of view information system is a complex combination management and technological solutions, hardware and software, and also information content.

The basic part of any information system is *database*, which complexity and capacity constantly increasing, leading to new difficulties in its operation. In particular, in many situations different nodes are controlling by different database management system. Such information systems are called *heterogeneous* [2].

Present paper is devoted to the analysis of special approach for describing database in large heterogeneous information systems. In particular, the publication demonstrates the method of working with such databases in natural for users terms, giving them possibility not to take into consideration the details of physical storage data in the relational model and type of database management system as well.

In order to provide the formulated above features the new approach is offered. It is found on additional metadata about information system, stored in the base. This approach is formally described in terms of graph model, that allows to formalize some algorithms of data processing, for example of verification problems.

The results of work are used in project of development CASE-system METAS, being a part of it.

METAS technology

CASE-system METAS (METAdata System) is a foundation for creation of system, which is controlled by metadata. This technology is destined for decreasing labour-intensiveness of the development of information systems and increasing their flexibility, scalability and adaptability in exploitation. The key feature of the discussion technology is usage interrelated metadata, which describes information enterprise system [7].

The majority of existing CASE-system generates code on some programming language according to definite specifications, which describe application domain. The main difference of METAS consists in usage of this metadata during its work. So the process of application's creation is reduced to writing necessary metadata.

The usage of metadata gives possibility of flexible application's adjustment and functionality. This creates the necessary prerequisites for development of "intellectual" system, which can automatically adapt to user's needs and to changing service conditions.

All metadata in METAS are divided into separate models (or levels), which of them describes the definite part of information system. Some model can describe the same part of information system, but from the different points of view. All models are interrelated, because one model can be based on another, being higher-level description of information system [8].

This article is devoted to the consideration of *logical model (LM)* of metadata. It directly founds on *physical model (PM)*, which describes all objects of database in information system: tables and their fields, tables relations, indexes and others. The detailed description of PM is presented in work [8].

There are some other models except listed above. Foe example the *presentational model* describes visual user's interface [6], the *reporting model* – queries and reports in information system and *security model* – right of users for execution operations with different objects in information system. This publication doesn't consider these models.

The main aims of logical model's creation

Some difficulties appear in usage of any relational databases. They are connected with characteristics of data keeping in relation tables and necessity of normalization. The structures of data, storing in database, often different from those, that user want to see. Let us consider this situation in more detailed way.

The first difficulty appears on the stage of application domain's analyze during designing database. The main method for this process is building the different entity-relationship diagrams or class diagrams in UML. In all these approaches the specificity of relational data model is necessary to take into consideration. The specialist, who dealing with analyses and designing, must work with physical tables and their relation and also must know the main principles of normalization. At the same time he must deeply understand the application domain of information system. This makes high demands of developer qualification.

Another difficulty appears in further work with database. In the process of SQL-query's construction it is necessary to remember all specificity of information storage in relational tables and to consider it.

For overcoming these difficulties the following approach was used in METAS. On design stage, as before, it is necessary to build diagram, which consists of entities and their relations. However in this case the term "*Entity*" means not one, but several related tables in database. This conception making entity closer to real life, simplifies the process of design.

The essential advantage of this approach is the possibility to work in natural for user terminology, abstracting from physical method of information storage. In particular, this slant provides to METAS independence from database type and support of heterogeneous systems.

For practical realization of discussed above approach the component called "*Logical model*" was developed. It is based on metadata of logical level. All other components that apply LM may work with database in logical level terms that strongly simplifies their development.

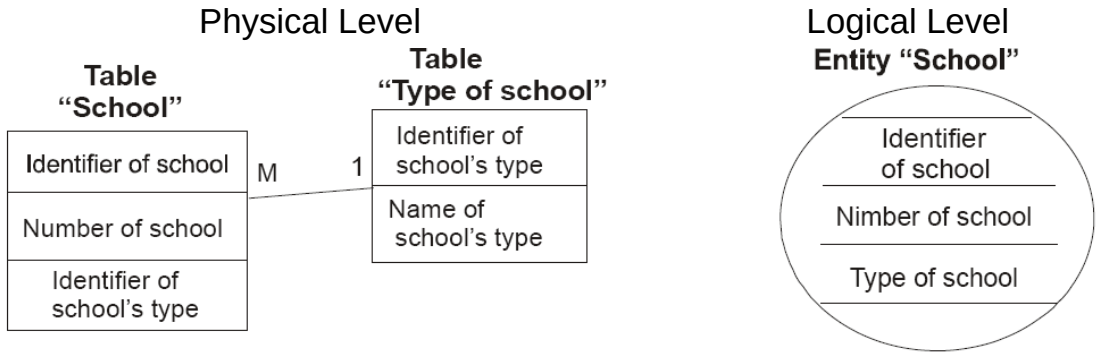
In this publication we will consider only main principles of LM. More detailed description may be found for example in [3, 4].

The main conception of logical model

The really existing objects in LM are presented by set of characteristics (attributes). These attributes can be physically stored in different tables of PM. So let as introduce some generalize construction called entity that is a set of these attributes.

For illustration we shell consider the fragment of database with information about schools: their numbers and types (for example lyceum, gymnasia). Obviously in relational database it is necessary to create two tables for presentation of this information. At that the first tables must contain foreign key to another. This tables form a physical level of metadata.

To organize the logical level for this fragment we may define the entity "*School*" with attributes "*Number*" and "*Type*". In this case "*Type*" becomes the same attribute like other. The only difference is a presence of mark that attribute's value is choosing from dictionary. Entity itself provides all work with this dictionary transparent for users. Thus every entity corresponds to set of physical tables. One of these tables is main, other are related with it as "1:M", in other word, they are *dictionaries* for it.



The main task of entity is to provide users possibility to work in term of application domain, without thinking about physical information storage. In other words, there is no necessity on logical level to take care about distribution of fields between physical tables. Now user may work in terms of entity, simply calling its name in combination with the name of attribute. Entity builds SQL query itself and either executes it itself (operations of inserting, deleting and updating information in database) or returns ready SQL expression.

Sometimes it is necessary to represent the main information about real object of some entity in the short form. For example let two entities are related as "1:M". In this case entity on side "M" has an attribute representing another entity (*foreign attribute*). In the discussing example entity "*Pupil*" has foreign attribute "*School*". It is not necessary to represent in this attribute all information about pupil's school, but only the main, which is enough for its identification. For example it may be school number. Such main information for entity is determined by the special *presentational expression*.

Presentational expression looks like SQL expression and may contain any attributes of corresponding entity. Any operation and function of SQL language are allowed. All attributes must be enclosed in square brackets, for example, for entity "*Pupil*" the following presentational expression is acceptable: *[Last name] + [First name] + [Middle name]*.

It was mentioned before that all entities are a set of attributes. There are some types of attributes listed below.

- *Own attribute*. It is any attribute from the main table of entity.

- *Dictionary attribute*. Any entity's dictionary has corresponding attribute. When we work with it we may choose necessary values from the list or add a new value in dictionary.
- *External attribute*. If the entity relates to another as "M:1", then it must contain additional attribute with information about this relation. For example entity "*Pupil*" has foreign attribute representing entity "*School*". The work with external attributes is similar to work with dictionary attributes, but its values may be chosen but not changed or added.
- *Key attribute*. In fact it is an ordinary own attribute. It contains a link to key field of entity's main table, so it unambiguously determines values of all the rest attributes. This term is introduced only for handy work.

We must not mix terms "presentational attribute" with "entity's attribute". The first one is not an attribute in the ordinary sense, because its value doesn't store in database. This is some value, which is calculated according to presentational expression and represent the main information about this entity. It is similar to attributes, because it is stored in entity together the rest attributes.

Let us introduce the concept of relation between entities. It is analogy with table's relation, but on logical level, in terms of application domain. For example entity "*School*" may be related with entity "*City*", which defines its location.

There are two basic relation types.

- 1:M. This relation corresponds to physical relation "1:M" between entities' main tables.
- M:M. This relation corresponds to physical relation "M:M", which is realized using intermediate table in relational database management system.

For any entity relation we may concretely define the count of entity on both sides. They have minimal and maximal number of entity, which may participate in relation. For example the following subtype for "1:M" are possible: «1 : 2..3», «0..1 : 0..1», «0..1 : 1» and other.

Building mathematical model

For building algorithms for METAS work it is necessary to formalize the description of using metadata for all models. Let us consider the formal definition of metadata systems for physical and logical levels in terms of graph model. As an example of using this model we will describe one of verification algorithm.

Physical level

Let us assume that *Fields* is some abstract set with elements of the same name. Let us also define the set *Tables* as reciprocal overlapping subsets of fields. For all tables the set of its fields we will name $F(t) \subset$ *Fields.*

Let us define the binary relation $R \subset Fields \times Tables$, and name it the set of relation between tables. The following notation will be used for it:

$$r \in R \Leftrightarrow r=(f,t),\ f \in Fields,\ t \in Tables.$$

Thus any relation $r \in R$ is built between some field $f \in Fields$, which belongs to the first table and the second table wholly (i.e. subset of a set *Fields)*. At the same time it is possible that relation connects the field of some table with the same table, i.e. $r \in R$, $r=(f,t)$, $t \in Tables$, $f \in F(t)$.

Logical level

Let us assume that *Attr* is some abstract set with elements of the same name. Let us divide all set *Attr* into groups, named entities. The set of entities called *Ent* must be *disjunctive*. This means from one side that all attributes in a set *Attr* must belong some entity, and from other side entities are not overlapped. If we name the set of attribute belonging entity $e \in Ent$ as $A(e) \subset Attr$, then the following expressions are correct: $\bigcap_{e \in Ent} A(e) = \varnothing$, $\bigcup_{e \in Ent} A(e) = Attr$.

All attributes in any entity we divide into three non-overlapping set. For any $e \in Ent$ we will denote *the set of own entity's attributes* as $O(e) \subset A(e) \subset Attr$, *the set of dictionary attributes* as $S(e) \subset A(e) \subset Attr$ and *the set of external attributes* as $V(e) \subset A(e) \subset Attr$. So conditions of non-overlapping of these sets and mandatory including of any attribute into one of these sets both take place:

$$\forall e \in Ent\ \ O(e) \cap S(e) = \varnothing,\ V(e) \cap S(e) = \varnothing,\ O(e) \cap V(e) = \varnothing,$$

$$\forall e \in Ent \ \ \forall a \in A(e) \Rightarrow a \in O(e) \vee a \in S(e) \vee a \in V(e).$$

The subset of entity's attributes including in presentational attribute we name P(e). It will contain attributes from any set O(e), S(e), V(e). So,

$$\forall e \in Ent \ \ P(e) \subset A(e), P(e) \neq \varnothing.$$

Strict definition of concept entity follows from the above theses. The *entity* is define by five

$$e = (A, O, S, V, P),$$

where all of the last four sets are subsets of the set A with described above properties.

At all pictures below we will use following graphical symbol for introduced concept. We will denote as circle any member of set *Attr*, and any member of set *Ent* as big circle. If attribute belongs some entity then it will be on corresponding circle. The attribute's color of filling means, to which set O(e), S(e) or V(e) it belongs. Corresponding colors for these sets are white, gray and black. The set of attributes including into presentational attribute we will outline dotted line.

There are many different means to define entities, which satisfy the condition of disjointness, if we have fixed the set of attributes. For all entities there are many different ways to define sets O(e), S(e), V(e), P(e). It is not conflict with discussed model, but in practice the model must correspond to real application domain. So all divisions of attributes into entities and divisions attributes by types are unambiguously determined. That's why we can think, that all divisions are defined unambiguously in our model.

Let us name as Y the set of all external attributes of all entities. Then $Y = \bigcup_{e \in Ent} V(e)$. Let us define *W* as single-valued transformation, which any $v \in Y$ associates with some $e \in Ent$ by means of one-for-one relation. Another word

$$w \in W \Leftrightarrow w=(v,e), \ v \in V(e_1), \ \ e, e_1 \in Ent.$$

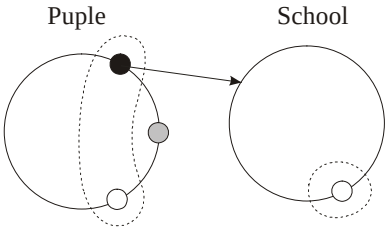
Thus any relation $w \in W$ connects one of external attributes of the first entity with the second entity. At the same time it is possible that relation connects the attribute of some entity with the same entity, i.e $w \in W$, $w=(v,e)$, $e \in Ent$, $v \in V(e)$. Graphically we will denote such relation as directed arc from external attribute to entity.

The second binary relation $M \subset Ent \times Ent$ we will name as set of relations "M:M" ("many-to-many")between entities. Graphically such relation we will denote as undirected arc between entities. Since there are many relation "M:M" between two entities in application domain, our graph will be of *multigraph* type or a *graph with parallel edges* [9]. For edges identification we will mark any such edge by unique name. This name will be used further in realization, for example in building the tree of objects serving for user's navigation inside the system [6]. Hence any relation "M:M" represents triplet $(e_1, e_2, name)$, where e_1, e_2 interrelating entities, and *name* is unique relation name. We will use the following notation:

$$m \in M \Leftrightarrow m=(e_1, e_2, name), \ \ e_1, e_2 \in Ent, \ name \in String.$$

Thus any relation $m \in M$ connects two entities wholly.

Now we will describe the example. Let us suppose that the entity "*School*" has own attribute "*Number*", and the entity "*Pupil*" has attributes "*Name*" (own attribute), "*Sex*" (dictionary attribute) and "*Pupil's school*" (external attribute). These entities are related as "1:M" by means of the attribute "*Pupil's school*".
For "*Pupil*" as presentational attribute we may use expression "*Name*" + " " + "*Pupil's school*", and for "*School*" – its attribute "*Number*."



In this example the set Attr consists of four elements and the set Ent consists of two elements. Let us denote the entity "Pupil" as $e_1 \in Ent$, and the entity "School" as $e_2 \in Ent$. Then $A(e_1)$ consists of three elements and $A(e_2)$ consists of one element. The sets $O(e_1)$, $S(e_1)$, $V(e_1)$ contain one at a time attributes, and set $P(e_1)$ contains two attributes: "Name" and "Pupil's school". For entity "School" sets $S(e_2)$ and $V(e_2)$ are empty, and sets $O(e_1)$ and $P(e_1)$ contain only one attribute "Number".

Triplet L=(*Ent*, *W*, *M*) is called complete graph of application domain. Under the term "complete" we mean, that the model contains all information about application domain. Further we will describe reductive graphs, which are

building from complete graph using some transformation. For example if in a task we don't interested in relation "M:M", then in complete graph we may replace the set M on empty set. This reductive graph will be more suitable for solving this task.

As it follows from the above, the complete graph is a *mixed* graph (that is it has both directed and undirected arcs) with parallel edges and loops.

In terms of discussed graph models we may build algorithms for solution different tasks in CASE-system METAS. For example there are algorithms for separation subschema, reengineering and migration of data [1]. In the following part we describe another example using this model.

Using graph model for verification metadata

Metadata in METAS is using for generation of all functionality for information systems. If developer mistakes in building metadata, he may not observe this, because the system of metadata may be very complex. That's way it is necessary to realize algorithms, which analyze structural static features of model and discover some types of mistakes.

As it was mention above, reductive graphs of application domain may be often used for description of some algorithms in terms of graph model. They may be built from complete graph using some transformation, which is specific for every algorithm. For example for some of them we must remove all relation "M:M" from complete graph, in other case we must remove all attributes.

Let us describe the example of algorithm, which checks the application domain on availability of cycles.

Under the term "*cycle*" we mean the following. Let there are two entities and both of them have external attribute to each other. Let both of these attributes be included in presentational attribute of their entity. In this case when we try to ask information about the first entity, we must know the value of presentational attribute of other entity. But it contains external attribute, that's why we must ask information about presentational attribute of the first entity. As a result we have circular reference and the system will not be able to execute this query. Obviously such cycle may include more than two entities and be less evident. That's why we must exclude such cycles on the design stage.

Let consider the formal model of this situation.

In complete graph of application domain we don't interested in such objects as relation "M:M", own and dictionary attributes. That's why the set *Attr* has only foreign attributes in reduced graph. The sets of entities and their presentational attributes are not changed. The set W (relations "1:M") is not changed too. Although condition $P(e)\neq\varnothing$ in reduced graph is broken , but it is not essential for our task.

As a result we get a graph $L_1=(Ent_1,W,\varnothing)$, где $Ent_1=\{e_i\}$, $e_i=(Attr(e_i) \setminus O(e_i) \setminus S(e_i), \varnothing, \varnothing, V(e_i), P(e_i))$, i.e.

$L_1=(Ent_1,W,\varnothing)$, $Ent_1=\{e_i\}$, $e_i=(V(e_i), \varnothing, \varnothing, V(e_i), P(e_i))$.

The formally definition of cycle is introduced below.

Let $RV(e_1,e_2)$ is set of attributes, which satisfies the following expression:

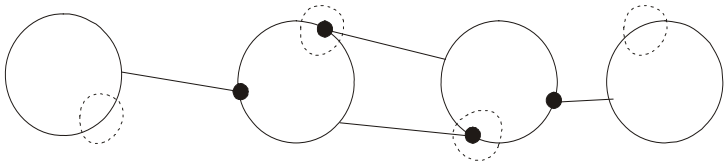
$$\forall a\in RV(e_1,e_2),\ a\in V(e_1),\ a\in P(e_1),\ \exists w_1\in W,\ w=(a,e_2),\ e_1,e_2\in Ent.$$

We suppose that the following condition is correct:

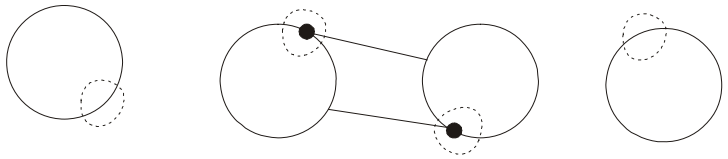
$$\exists a_1\in RV(e_1,e_2),\ \exists a_2\in RV(e_1,e_2),\ \dots,\ \exists a_{n-1}\in RV(e_{n-1},e_n),\ \exists a_n\in RV(e_n,e_1),\ e_i\in Ent,\ i=1..n.$$

In this case we will say that graph of application domain has a cycle.

The example of graph with cycle is demonstrated in the following picture.



Obviously if external attribute is not a part of presentational attribute, then it doesn't affect on cycle's availability. That's why we may reduce our graph more when we remove all such attributes. Also it is necessary to remove all corresponding relation "1:M". We will have the following graph:



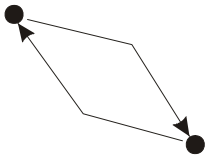
$L_2=(Ent_2,W_2,\varnothing)$, $Ent_2=\{e_i\}$, $W_2=(v_{ij},e)$, $e\in Ent$, $e_i=(\{v_{ij}\}, \varnothing, \varnothing, \{v_{ij}\}, P(e_i))$, $v_{ij}\in V(e_i)$, $v_{ij}\in P(e_i)$.

Further we may reduce this graph to classical directed graph. For this purpose we consider that all external attributes are graph's nodes, and edges from one node to another are contained in graph if only external attribute, corresponding the first node, is related with entity, which contains external attribute, corresponding the second node.

$G=(GV,GE)$, $GV=\{v_i\}$, $v_i\in Attr_2$, $GD=\{(v_i,v_j)\}$, $\forall v_j\in A_2(e)$, $w=(v_i,e)\in W_2$, $e\in Ent_2$

Here $Attr_2$ is a set of all attributes of graph L_2, and $A_2(e)$, $e\in Ent_2$ is its subset for entity e.

Further we may use any standard algorithm of finding cycles in graph.



Conclusion

Publication is devoted to the approach to the description of database, which allows to work in naturally for user terms and don't take into consideration details of physical storage of the relational model and type of database management system. Formally described in terms of graph model, this approach allows to develop several algorithms, for example, for verification of application domain.

The developed theory has a practical realization as a part of CASE-system METAS. Based on the last one the following information systems were developed: "The educational system of Perm region" and "Interdepartmental information system of personified registration of children in Perm region". These systems were successfully introduced into operation testing.

The results of paper permit generalization on distributed database case, when all information may be divided between different network nodes. This provides guidelines for future work [5].

Bibliography

[1] D.A. Borisova. The component of restructurization of CASE-system METAS. In: Mathematics of program systems. Perm: PSU , 2003. (in Russian)

[2] C. J. Date. An introduction to database system. Eighth Edition. Addison-Wesley, 2003.

[3] M. E. Eremina. Generation of SQL-expressions based on metadata. In: Mathematics of program systems. Perm: PSU , 2003. (in Russian)

[4] M. E. Eremina. Using metadata for generation of SQL-queries to database. In: Modern problems of mechanics and applied mathematics. Voronezh: VSU, 2004. (in Russian)

[5] M. E. Eremina. Realization of queries in distributed heterogeneous systems, controlled by metadata. In: Youth. Education. Economics. Part 4. Yaroslavl: Remder, 2004. (in Russian)

[6] E. Yu. Kudelko. Generation and adjustment of screen forms based on metadata. In: Mathematics of program systems. Perm: PSU, 2003. (in Russian)

[7] L. N. Lyadova, S.A. Ryzhkov. CASE-technology METAS. In: Mathematics of program systems. Perm: PSU, 2003. (in Russian)

[8] S.A. Ryzhkov. Conception of metadata in the development of information systems. In: Mathematics of program systems. Perm: PSU, 2002. (in Russian)

[9] M. Svami, K. Thulasiraman. Graphs, networks and algorithms. Mir, 1984. (in Russian)

Authors' Information

Maria Evgenevna Eremina – Perm State University; 614990, Russia, Perm Bukireva st. 15; e-mail: erm6@mail.ru

PARALLEL BETWEEN DEFINITION OF CHESS PLAYING PROGRAM AND DEFINITION OF AI [1]

Dimiter Dobrev

Abstract: In this report we will explain some earlier papers [1, 2] which are about definition of Artificial Intelligence and about perfect AI. The definition of AI is intuitive in [1] and formal in [2]. The perfect AI is a program that satisfies the definition for AI but which is absolutely useless because of the combinatory explosion.

Most people do not understand these papers because they never saw AI and that is why for them the notion of AI is too abstract. In this report we will make parallel between definition of chess playing program and definition of AI. Of course, the definition of chess playing program is useless because people already know what this is. Anyway, we will give you this definition because its construction follows closely the construction of the definition of AI. Also the results are almost the same with the only difference that we can optimise the perfect chess playing program in order to obtain a real chess playing program, but for the moment we cannot optimise the perfect AI in order to obtain a real AI.

In this report we will not speak about AI. The only matter which we will observe will be about chess playing programs. If you understand the construction and the results about chess playing programs then you can read the papers [1, 2] and to see similar results about AI.

Keywords: AI Definition, Artificial Intelligence, Chess Playing Program.

ACM Classification Keywords: I.2.0 Artificial Intelligence - Philosophical foundations

What is a Chess Playing Program?

For us a chess playing program (CPP) will be a step device which inputs the move of the opponent and outputs a correct move as an answer on every step. If this was the only requirement, then the random player would be chess playing program, but this is not true because the random player plays too bad. That is why we will want from the chess playing program to play no worse than a human being.

Here you can ask the question: "Who is the human being?" The answer is that this is not so important because the difference between the average chess player and the world chess champion is not that big. This is because the program which plays better than the average chess player is almost the same as the program which plays better than the world chess champion. The main difference is in the number of moves which are calculated in order the next move to be chosen. Of course, the second program needs more time or faster computer than the first one.

The definition which includes the world human is not formal but it can be formalised.

To make the definition formal first we have to formalise the opponent. We will suppose that the opponent of our chess playing program is the random player. This is the opponent which every time plays a random move. Of course, you can have many different random players which have different possibility for choosing their random moves, but we will assume that our random player chose its random move with equal possibility. That means that if this player has *N* possible correct moves, then the possibility for each of them to be chosen is *1/N*.

The random player is very weak opponent, but, anyway, it is a dangerous partner because sometimes it makes genius moves. Really, the possibility to make a genius move is pretty small. The good side of the random player is that it is simple and well defined. Another positive feature of the random player is that it has no memory. This means that it cannot learn and that the next game will not depend on the previous.

We will make some additional assumptions. We will assume that the rules of chess are fixed and that these rules do not allow infinite game (for example, if 20 moves no pawn is moved and no figure is taken then the game is draw). We will assume that our CPP plays with whites in every game.

[1] This publication is partially supported by the Bulgarian Ministry of Education (contract **БОЕ 4-02/2004 г.**)

We may assume that our CPP is a deterministic program (i.e. that it plays only one strategy). If we assume that CPP is nondeterministic program and that it plays many different strategies, then we have to know what is the possibility of each possible strategy. In the first case the average success of CPP will be the average success of its strategy. In the second case its average success will be the average from the average successes of all strategies (the sum of *AverageSuccesses(i).Possibility(i)* where *i* runs true all strategies of CPP). Notice, that the number of all strategies is finite due to the fact that the tree of the game is finite, which is so because the rules of chess do not allow infinite games.

Now we have to ask the question about the goal of the game. In the case of AI this question was difficult and we had to introduce the notion of meaning of life. Here we have no such problem because the goal is obvious and it is to win the game. Speaking more formally, let us give to CPP one for victory, zero for lost and half for draw game. The goal will be to make the biggest average success.

The Perfect CPP

After this formalisation we can say when one CPP is better than another and even something more, we can show the perfect chess playing program (or the perfect chess playing strategy, which is the same if the program is deterministic). The perfect CPP will calculate the tree of all possible moves before making any move, and that is why this program will be practically useless due to the combinatory explosion. Anyway, it is interesting from the theoretical point of view to have such a program.

The perfect CPP will evaluate all positions in the tree of the game by Max-Sum algorithm (maybe in this case it is better to call this algorithm Max-Average instead of Max-Sum). This algorithm is theoretically possible because the tree of the game is finite. The Max-Sum algorithm evaluates first the leaves of the tree and gives one for these leaves whose position is victory for CPP, zero respectively for victory for the opponent and half for draw positions. After this, Max-Sum algorithm evaluates the rest of the vertexes calculating maximum from the evaluation of the successors when the move is made by CPP and calculating average when the move is made by the opponent. Of course, after this evaluation Max-Sum algorithm plays by selecting this move which leads to the vertex with maximal value. (Of course, if vertexes with maximal value are more than one the Max-Sum algorithm chooses one of them. Because of this choice we can have more than one perfect strategy and also we can have nondeterministic perfect CPP which combines several perfect strategies.)

The perfect CPP is perfect but useless due to the combinatory explosion. It is easy to see that this program will make the best average success against the random player. Anyway, the perfect CPP is a little bit strange. If it is in a winning position, then it will win, which is OK, but if the perfect CPP is in draw position, then it can lose the game. Actually, perfect CPP can move from draw position to losing position, which is strange, but this is because the perfect CPP is perfect against the random player. If the opponent was different then the perfect CPP would be different.

Possible Variants of the Definition

Now we have the perfect CPP and the question is how to define what is a CPP. We can say that this is any program which is as good as the perfect CPP. At least the perfect CPP covers this requirement. This is not a good idea for definition because we need a program which can play in real time. Also, we do not need a player so perfect as the perfect CPP.

That is why we will define CPP in the following way:

First variant of the definition: CPP will be the program which makes average success which is on smaller distance than ε from the perfect.

If this ε is chosen smaller enough, then the chess playing program will play no worse than a human.

Here we have one problem. We do not know how big is the perfect result. If the starting position in chess is winning for the whites, then the perfect result is one (or $1-\delta$ where δ is zero). If the starting position is not winning for the whites, then the perfect result is $1-\delta$, where δ is some very small number bigger than zero. We have an algorithm for calculating the value of δ, but we cannot calculate it because this algorithm will not end in reasonable time (i.e. before the end of the universe). The conclusion is that the previous definition is not convenient because it includes two constants whose value cannot be calculated (δ cannot be calculated due to the combinatory explosion and ε cannot be calculated at all because it depends on the human beings, which

means that this constant has no exact value). Even if we choose a value for ε we will still not know the value of the $\delta+\varepsilon$. This is the reason to define a CPP in a different way.

Second variant of the definition: CPP will be the program which makes average success which is bigger than $1-\alpha$.

Here α is some small positive number. It must be equal to $\delta+\varepsilon$ in order to be the new variant of the definition the same as the previous variant. In any case α must be greater than δ because in the opposite case CPP will not exist.

The second variant of the definition gives us the possibility to check it for a concrete program by the tools of the statistics. There are two problems. First, we cannot say how big should be α for the CPP to exist and in order to play no worse than a human being. Of course, such value exists but we cannot say which it is. The second problem is that the value of α is very small, which makes it extremely difficult to calculate it by the methods of the statistics.

In order to eliminate the second problem we will change the opponent. We need a stronger player as opponent in order to have bigger expectation for $\delta+\varepsilon$. Let us try with human being as an opponent.

Third variant of the definition: CPP will be the program which makes average success against human being bigger than 1/2.

The third variant of the definition is not formal, but this variant is usually used as a definition of CPP. The good side of this definition is that it can be checked by the methods of the statistics. For example, if we make ten games between one program and a human being and if the average success of the program is 60% (0.6), then we can say with big possibility that this program satisfies the definition for CPP. Really ten games are too few and for bigger certainty we may make 100 or even 1000 games. Here our expectation for $\delta+\varepsilon$ is 1/2. Really the game is not symmetric (opponent plays always with the blacks) and that is why we are not sure that human against human will make 1/2 average success. Also, you have to notice that this variant of the definition is not formal and that the results will depend on the humans we use for opponents.

New Fixed Opponent

Anyway, we need a formal definition of CPP and that is the reason why we will use a well-defined opponent. Such opponent is the standard chess playing program which calculates *N* moves in the tree of the game and by Min-Max algorithm selects the best move. The problem is that this opponent is deterministic. We prefer the opponent to be nondeterministic because if both CPP and the opponent are deterministic, then the result of the game between them is also deterministic (actually, there is only one possible game between them). In this case we cannot use the methods of the statistics. Also, it is not a good idea to use deterministic opponent for the definition because all deterministic programs which have average success one will be CPP but maybe some of them accidentally made victory in this only game, which does not prove that these programs are CPP.

For opponent we will use nondeterministic variant of the standard chess playing program. Let our opponent depend on one integer *N* and two functions - *F* and *P*. Let function *F* evaluate positions and on every position *F* return an integer. Let function *P* give the possibility for one move to be chosen. This means that the input of *P* is a list of integers which represent the values of the positions which can be obtained on the next move and the output is a list of the possibilities each of these positions to be obtained. The integer *N* and the functions *F* and *P* should be concrete and fixed because they will be parameters in our definition. If we change one of these parameters, then the definition will also be changed.

How will this fixed opponent work. It will evaluate by the function *F* all positions which are obtainable after *N* moves from the current position. After this, by the Min-Max algorithm, it will calculate the values of the positions which are obtainable on the next move. After that, by *P*, it will calculate the possibility of each of these positions to be obtained and on the end will randomly choose one of the moves, but in such a way that the possibilities of the next positions to be the same as the ones given by *P*.

This fixed opponent is nondeterministic (except the case when *P* gives 1 to one of the numbers and zero to the rest of them).

After the change of the opponent we also have change in the perfect CPP. It will be constructed almost in the same way, but instead of average value it will calculate the sum of *Value(i).Possibility(i)* where *i* runs true all

possible moves of the opponent. Here *Possibility(i)* will be given by the function *P*. (Of course, before applying the function *P* we have to use function *F* and Min-Max for N moves in order to calculate the values of all positions which are obtainable on this move.)

If the number *N* is big enough and if the functions *F* and *P* are reasonable, then our fixed opponent is a good chess player. This means that our expectation for δ and ε will be high. Let us assume that ε is 10%. This means that we suppose that a human being will make average success against the fixed opponent 10% worse than the perfect CPP. Let us assume that the average success of the perfect CPP is 80% (i.e. that δ is 20%). If the starting position is winning for the whites, then δ is zero but our expectation is that the starting position is not wining and not losing. This leads us to the following:

Final variant of the definition: CPP will be the program which makes average success against the fixed opponent which is bigger than 70%.

As we mentioned, this definition depends on the fixed opponent and from the number α, which here is fixed to 30%. If we fix also the number *N* and functions *F* and *P*, then we will have one concrete definition. We can easily check whether one program satisfies the requirement of this definition by the methods of statistics. Really, we cannot check this for sure, but with a great percent of possibility (which percent grows with the number of the test games).

This definition is made with one assumption:

Assumption 1: Here we assume that the average success of the perfect CPP is about 80%. If this conjecture is true, then there exists a program which satisfies the definition (at least the perfect CPP does). If the average success of the perfect CPP is smaller than 70%, then there is no such a program (of course, in such case we can change this parameter and make it smaller than 70%).

Conclusions

The first concussion is that CPP exists and that the perfect CPP satisfies the demands of the definition. Really, this is true if assumption 1 is true, but if we take the first variant of the definition then this is true without assumption 1.

The second conclusion is that the perfect CPP is CPP but it is useless due to the combinatory explosion. In order to obtain a real CPP we have to optimise the perfect CPP and make a program which may be a worse player than the perfect but which can play in real time. That is an easy task in the case of chess, but not so easy in the case of AI.

Bibliography

[1] Dobrev D. D. A Definition of Artificial Intelligence. In: Mathematica Balkanica, New Series, Vol. 19, 2005, Fasc. 1-2, pp.67–74.

[2] Dobrev D. D. Formal Definition of Artificial Intelligence. In: International Journal "Information Theories & Applications", vol.12, Number 3, 2005, pp.277-285.

[3] Dobrev D. D. Formal Definition of AI and an Algorithm which Satisfies this Definition. In: Proceedings of XII-th International Conference KDS 2006, June, 2006 Varna, Bulgaria, pp.230-237.

[4] Dobrev D. D. Testing AI in One Artificial World. In: Proceedings of XI-th International Conference KDS 2005, June, 2005 Varna, Bulgaria, pp.461-464.

[5] Dobrev D. D. AI - What is this. In: PC Magazine - Bulgaria, November'2000, pp.12-13 (in Bulgarian, also in [7] in English).

[6] Dobrev D. D. AI - How does it cope in an arbitrary world. In: PC Magazine - Bulgaria, February'2001, pp.12-13 (in Bulgarian, also in [7] in English).

[7] Dobrev D. D. AI Project, http://www.dobrev.com/AI/

Author's Information

Dimiter Dobrev - Institute of Mathematics and Informatics, BAS, Acad.G.Bonthev St., bl.8, Sofia-1113, Bulgaria; P.O.Box: 1274, Sofia-1000, Bulgaria; e-mail: d@dobrev.com

TABLE OF CONTENTS OF VOLUME 1, NUMBER 2