

N.A. Bazhenov

Combining Probabilistic Tagging with Rule-based Multilevel Parsing for Requirements Elicitation

The concerned paper deals with the problem of natural language processing and analyzes with regard to requirements elicitation and analyze, which are realized by requirements and software engineers in cooperation with stakeholders and end-users. To bridge gap between raw requirement texts generated by system's end-users and the conceptual models which are needed for system designers and architects the hybrid NLP approach is proposed. The developed approach is based on POS probabilistic tagging combined with rule-based chunking (shallow parsing) enriched with some additional heuristic procedures of different linguistic characteristics identification. There are the following processing steps established the developed methodology: (1) Tokenization and correction of wrong split tokens; (2) POS tagging provided by a probabilistic trigram-based tagger with possibility of ad-hoc additional lexicons plug-in; (3) The mapping engine for splitting up and reinterpreting the standard tagset into ontology-oriented one; (4) The pre-chunking procedure including compound noun identification, extraction and generation of inflectional word forms, verb transitivity disambiguation, passive voice identification, nouns classification and other morphological information extraction; (5) multilevel chunking procedure including the standard chunking rules engine, multi-words and idiomatic expressions identification and complex phrase sequence identification (such as prepositional post-modifier, parenthesis, homogeneous parts, relative clause, infinitive group, subordinate sentence). This approach focuses on the enriching standard tags optimizing the automatic elicitation of concepts, relations and attributes for requirements engineering purposes. Free text processing in this way can only be successful, if certain NLP standard techniques like probabilistic tagging are combined with special procedures like pre-chunking and chunking.

Based on theoretical foundations of proposed approach the software application for requirements documents processing was developed. The implemented tool was successfully adopted and integrated into the project NIBA (Natürlichsprachliche Information Bedarf Analyse, founded at Alpen-Adria Universität Klagenfurt, Austria), whose aim is natural language processing to support requirements engineering and conceptual modeling.

Al-Ammouri Ali, O.P. Shkurko

Techno-economic Self-control of Informative-control Systems Functioning Efficiency

Nowadays the economy is closely connected with information technology and automation of production processes. Any enterprise achieves great economic success in case the production is not expensive and qualitative.

To attain two such important attributes of the production as low cost price and high quality it is possible only with method of manufacturing processes quality automation. Electronics is the basis of automation, and on its modern development level is microprocessor technology.

Interface is the most pressing issue of the day while applying microprocessor technology in the technical processes of production including, foremost, efficient hardware input data base to microprocessor for the very limited time.

This article deals with matrix structures of computing means connections, intended for sequential information processing in real time, arriving randomly at a Poisson law on several fronts (lines) and processed consistently by phases (matrix columns).

The selection method of optimal matrix structure of computing facilities is offered when data is transferred through random flow by means of every channel (matrix line). Such structures can be attributed to multichannel and multiphase queuing systems.

A.S. Voronov, A.A. Yegoshina

Tools for Integration Domain Ontologies for Knowledge Bases of Intelligent Tutoring Systems

Intelligent educational systems based on systematized knowledge are intensively developing currently. Ontology is an effective way of knowledge presenting and organizing. But the standard ontologies can't be used in specific and dynamic fields of knowledge that makes their unification an actual task. The article deals with the basic problem of a phased creation of ontologies in the knowledge basis of intelligent educational systems. Also it contains the analysis of existing ontology integration tools, such as mapping tools, ontology alignment and unification.

It is shown that the source of complicated ontology is a multiagent search system for ontological information in the subject areas. Searches in this system are formed on the thesaurus basis of subject area and on the expert local knowledge basis objectives. Relevant components of the ontology from external repositories of educational resources and other semantic Web search system, that are loaded in the working database of integrated association of ontologies are found.

The authors created the algorithm of ontology associations, including: ontology order alignment, resolution of names conflict, classes, and their properties and relations association. Based on the proposed algorithm software for integrating ontologies is created, that is an important part of information resources integration tool, which can be used for creation of the ontological knowledge bases.

Developed tools can be used to speed the creation of intelligent education environment for students' training in the direction of Computer Science. Further task is the realization of other integration system components, formalizing integration process and development of a universal algorithm on ontologies.

M.I. Gorbijchuk, S.T. Samulyak, I.V. Shchupak

Classification Functions Synthesis Based on Genetic Algorithms

The task of building the classifiers based on the partition functions synthesis can be found in many technical applications, for example in image recognition systems, in a problem of division of the diagnosis signs space into classes during technical condition evaluation process etc. In most cases it is assumed that the structure of the partition function is given and there is only a need to find its parameters. However the structure of the function is unknown beforehand and the problem of its optimal structure definition is the topic of this study.

The authors used the genetic algorithm in order to solve this problem. The essence of the algorithm is that the structure of the partition function is chosen to be linear. The chromosome associated with that structure consists of binary genes. "Null" value is assigned

when the parameter value in function structure is equal to zero, otherwise the value is counted as “one”. The problem of the optimal partition surface synthesis is solved the following way – one chromosome is found from an initial pool of relatives, so that the fitness function acquires its minimal value.

The novelty in this study is the classification function synthesis method development based on the genetic algorithms. The practical value of the work is that the authors created software for solving the partition functions synthesis problem and the particular example is presented showing its efficiency.

G.V. Dorokhina, V.Y. Trunov, E.V. Shilova

Russian Words Morphological Analysis without Dictionary Module

The article is devoted creation of software for Russian words morphological analysis without a dictionary (RwdMA_IAI). Such system is needed when the word is to be processed by a module of morphological analysis is out of dictionary. Also there was a plan to use this module in word’s paradigms adding automation task.

The database of RwdMA_IAI is formed by using the Russian words declarative morphological analysis module (RDMA_IAI). It contains table of pseudo-inflection paradigms. Each pseudo-inflection paradigm is an ordered set of pairs: information about grammatical form, pseudo-inflection. Two pseudo-inflection paradigms are equal if their sets of pairs are same.

The filling of RwdMA_IAI database includes following stages: successive paradigms sampling, pseudo-inflection paradigm forming for each paradigm, pseudo-inflection paradigm finding in RwdMA_IAI database, rang increasing for found one and adding of that pseudo-inflection paradigm which is absent.

As a result RwdMA_IAI is filled with paradigms of verbs, nouns, adjective from RDMA_IAI database. It is able to process the morphological analysis of unknown words and to synthesize their potential paradigms. The database of RwdMA_IAI contains about 19 thousand pseudo-inflection paradigms that includes (table of 680 thousand rows, 32.5 thousand different pseudo-inflections).

Module using while Forum “Methods of automatic analysis of text estimation: morphological parser for Russian” (<http://ru-eval.ru/>) shows that 4.5 % of processed words is unknown and is to be processed by RwdMA_IAI module. An average time of word morphological analysis by RwdMA_IAI is over 0.01 second.

Module using for synthesizing the potential paradigms of unknown words shows that there are too many hypotheses of paradigms for one word to automate words’ paradigms adding process. The correct paradigm could be selected by analyzing the set of potential paradigms of large text’s unknown words.

Vladimir S. Mayorov, Sergey V. Mayorov, Sergei M. Komarov

Program System for Laser Cutting

Laser cutting is widely used in modern industry for manufacturing of different parts from metal, plastic, wood etc. Depending on the type of material and laser machine itself, different parameters of processing are required. A source drawing for cutting can also include different graphical objects designed by means of different software applications, such as AutoCAD, Corel Draw, etc. Usually outlines are composed of arcs, lines, polylines, Bezier curves, ellipses or NURBS.

The goal of this project is to create an intellectual program system for laser cutting, capable of selecting the best parameters for laser cutting of complex outlines. The system presented for your consideration consists of two main modules: geometry processor and converter (GPC) and Decision Support System (DSS). The GPC module imports vector drawings such as .DXF, .AI or .EPS files and converts all outlines into contiguous polylines consisting of lines and arcs. A potential User can save the converted drawing and manually edit it, or immediately turn it into a CNC program. All technological parameters for the CNC are supplied by the DSS module. The DSS module intends to help technicians to choose the best possible conditions for laser processing of materials. This module includes several important blocks and components: physical models and equations, data bases, practical knowledge of experts, experimental results bank, some important criteria, algorithms for analysis and decision. A set of parameters for the description of laser technological processes has been developed. Each process is characterized by the values of such parameters changing within some reasonable diapasons. The DSS module processes technological data and offers the solution using the modern Case-Based Reasoning (CBR) technology (solution of problems by the analysis of precedents). Method "k" of the nearest neighbors (k-NN) brings the best results. The practical tests prove the high efficiency of our program system for laser cutting.

A.S. Minenko

Mathematical Modeling of the Processes Crystallization of Metal

There is researched Stephan's space task regarding to an ingredient and convective movement in a liquid phase, described by Navier-Stokes equations. The method of studying of this problem consisting of the decision expansion for series of a small parameter is offered. In the no stationary case the decision of corresponding boundary-value tasks for definition of the expansion members is formed as fixed points of operators.

Convection influence on the front of crystallization is explored. The method of the conjugation tasks decision arising while Stephan's tasks research in space is also developed.

This paper extend to the rime-dependent case some result obtained by the authors for the steady-state Stefan problem with convection. Referring to the ice-water system, water

is assumed to be incompressible and to obey the Stokes equation $\frac{\partial v}{\partial t} - \nu \Delta v + \nabla_x p = f(u^{(1)})$

(v velocity, ν kinematic viscosity, p pressure, f buoyancy force, $u^{(1)}$ water temperature), while temperature $u^{(1)}$ satisfies the heat conduction-convection equation with temperature-dependent thermophysical properties. The temperature field $u^{(2)}$ in the solid phase is governed by diffusion only. At the ice-water interface, $u^{(1)} = u^{(2)} = 0$ and the Stefan condition holds. The scheme is completed by initial and boundary conditions.

The authors present an existence theorem for the case of two space dimensions. The main difficulty consists in the fact that, to interpret the Stokes equation in a weak sense, some information is needed on the region where the temperature is positive, which is in turn influenced by the velocity field itself. The precise formulation of the problem requires a technical choice on function spaces. Existence of a solution is proved by introducing a temperature dependent penalty term in the fluid flow equation in order to define both the approximating temperature u_ε and the approximation velocity ε in the whole domain. Compactness arguments are used to get a convergent subsequence, whose limit is shown to solve the original problem. The question of uniqueness is left open.

A.S. Minenko

Approximation Analysis of Many Dimensional Mathematical Model of Crystallization of Metal with Convection

Under of the studying of the heat physics processes accompanied by substance phase transitions, described by the mathematical model in which the each phase temperature satisfies the equation of heat transition with its heat physics coefficients, on border of phases division, both temperatures are constant and equal to the temperature of a phase transition and on the set parts of border certain schedule is supported. The phases division surface (“a free border”) is unknown and Stephan’s condition is set additionally for its definition. This condition turns mathematical model into a nonlinear problem of the large difficulty. The Navier-Stokes equations are used to describe speed fields in a liquid phase. The method of a small parameter is offered to solve the task.

Up to now only one paper has been know which contained the proof of the existence of the classical solution to the two-phase multi-dimensional Stefan problem. The paper under review is the second one containing result of no less significance. The author considers the nondegenerate two-phase Stefan problem in a rectangle, assuming that there exists a solution of the corresponding stationary Stefan problem and that the boundary and initial data of the problem are only slightly different from the corresponding characteristics of the stationary solution. The existence of the smooth solution of the problem is proved under not too restrictive assumptions concerning the input data of problem. Moreover, the author indicates a simple condition guaranteeing the existence of the corresponding stationary solution and its stability, the latter being understood in its usual sense. This paper may be strongly recommended to all interested in parabolic free boundary problems in general and in their classical treatment in particular.

L.P. Mironenko, I.V. Petrenko, I.A. Novikova

The Unique Approach of Obtaining of the Canonical Equations of the Second Order Curves

There is a number of approaches to the description and classification of the second order curves. This question is not new in analytical geometry. From our point of view there is a possibility for more laconic and clear statement for the problem of getting of the 2-nd order canonical equations and their classification.

In the offered theory this problem has been solved on the basis of unique definition for all the second order lines. In the base of this definition the eccentricity notation and the eccentricity distinction for various types of curves were applied. Let’s define the second or-

der line as the set of points of a plane, for which the relation of distance r from the fixed point F named focus of the given curve, to any point M of a curve to distance d to some straight line D named the directress, is a constant e named eccentricity.

Let's apply this definition to the 2-nd order general equation with two variables x and y

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y - a_{33} = 0,$$

where $a_{11}^2 + a_{12}^2 + a_{22}^2 \neq 0$ and $a_{11}, a_{12}, \dots, a_{33}$ are the equation factors, which are real numbers. Let's arrange axis y of Cartesian system perpendicularly to directress D . The focus of the asked line must be on the axis x and differ from

Cartesian system origin. At such conditions the curve must be symmetric with respect to the axis x and factors a_{12} and a_{23} must be zeroes. In fact we'll get the equation

$$a_{11}x^2 + a_{22}y^2 + 2a_{13}x - a_{33} = 0.$$

In this expression it is easy to write down the full square with respect to the variable x . It means there are two different cases, when $a_{11}x^2 + a_{22}y^2 - a_{33} = 0$ and $a_{22}y^2 + 2a_{13}x = 0$. Let's write down the both cases by using Kronecker's symbol δ , which is defined as

$$\delta_{mn} = \begin{cases} 1, & \text{if } m = n \\ 0, & \text{if } m \neq n \end{cases}.$$

$$\text{Finally we'll get } (1-\lambda)x^2 + a_{22}y^2 + \delta_{\lambda,1}2a_{13}x - (1-\lambda)a_{33} = 0,$$

where λ is a new parameter and $a_{11} = 1 - \lambda > 0$. Since a is the big semiaxis and b is the small semiaxis of an ellipse, where $b = a\sqrt{1-\lambda} < a$ at $\lambda < 1$, therefore we'll obtain:

$$\frac{x^2}{a^2} + \frac{y^2}{(1-\lambda)a^2} = 1, \text{ where } b^2 = (1-\lambda)a^2 \rightarrow \lambda = 1 - \frac{b^2}{a^2} = e^2 \rightarrow e_{\text{ellipse}} = \sqrt{1 - \frac{b^2}{a^2}} < 1,$$

where e_{ellipse} is the eccentricity of an ellipse.

If $\lambda > 1$, then it follows $e > 1$ and we'll get the expression for the eccentricity of a hyperbole:

$$b^2 = (e^2 - 1)a^2 \rightarrow e_{\text{hyperbole}} = \sqrt{1 + \frac{b^2}{a^2}} > 1.$$

If $\lambda = 1$, then it follows $e = 1$. If assume $p = -a_{13} / a_{22} > 0$ we'll get the canonical equation of a parabola: $y^2 = 2px$, where $p/2$ is the distance from the focus to the origin of Cartesian co-ordinates.

Now it is possible to write down all the second order lines equation by using the unique expression, which depends from the eccentricity e :

$$(1 - e^2)(x^2 - a^2) + y^2 = \delta_{e,1}2px.$$

At $e = 0$ we'll get the circle equation: $x^2 + y^2 = a^2$.

In the traditional definitions of the ellipse and the hyperbole take place only one focus and one directress, but the symmetry of the last equation with respect to the axis y means that these curves have two focuses and two directresses. Further in the paper has been considered the main geometrical properties of the second order curves.

K.V. Murygin

Detection of Car License Plate on the Basis of the Mixed Cascade of Classifiers

The automation of auto transport traffic control is now actual and important problem in which frameworks systems of automatic recognition of license plates actively are developed. There are many the similar systems focused on national standards of license plates which can essentially differ under many characteristics: the size, the form, color of symbols and a background, quantity and structure (letters and figures) symbols, the alphabet of symbols. Therefore the problem of recognition of license plates of vehicles accepted in Ukraine is actual and important in practice.

The first development stage is the decision of a problem of a car license number detection which is greatly defines quality of algorithm complex of license number recognition system.

For detection of car license number on digital images the wide variety of heuristic approaches and their combinations based on prospective differences in properties of images of license plates, and the images which are not concerning them is used. Some of such approaches is based on: maps of vertical borders; spectral characteristics; fuzzy maps of an membership of points to license plates; calculation of number of the borders following on certain distance from each other; morphological transformation for search of lines of the symbols having certain width, etc. Unlike listed approaches, the used method is based on the general principles of search of objects on images. As the initial data for getting of effective algorithm of search the set of grayscale pictures of car numbers, and background images is used. Thus any heuristic knowledge of properties of images of numbers it is not used.

Detection was based on a cascade of classifiers and multiscale image scanning technique. The classifiers of two types are used:

1. MCV-classifiers – at the cascade initial stages;
2. The linear classifiers received by means of algorithm AdaBoost – at the cascade ending stages.

At the expense of use at the initial stages of MCV-classifiers the number of properties used for classification is reduced at classification quality preservation that in a combination to possibility of treelike representation of the MCV-classifiers allows reducing search time essentially.

During training the cascade of classifiers having following characteristics has been received: positive detection rate – 0.941, an error of false detection – $<10^{-8}$. Rate of processing of images in the size of 640*480 pixels on computer Pentium Core2Duo with frequency of the processor 2.33GHz was 12-15 frames per second. The considered algorithm allows using successfully it as the first stage of processing of images in systems of recognition of car license numbers. The further researches in a direction of development of automatic car plate recognition system can be bound up with normalizations of the car plate image, segmentation of symbols and recognition of the symbols presented on the images of license plate.

A.N. Serebrovskiy

About Technology of Knowledge Extraction from the Informative Resources of Expert System Subject Domain

The approach to creation of knowledge extraction technology from text electronic resources which have not preliminary semantic description is offered. Technology is intended for forming and update of knowledge base of expert system and includes the following processes: constructions of ontology and thesaurus of subject domain; the automated semantic markup of text documents; extractions of knowledge from marked texts.

Description of thesaurus construction technology and ontology of subject domain

Thesaurus and ontology of subject domain are built one time at creation of expert system and then repeatedly used for its exploitation.

Technology includes the following stages: forming of text set characterizing a subject domain; compiling of terms dictionary of subject domain; forming of list of notions of subject domain; classification of notions according to criteria: object, process, event, property, value; establishment of base semantic connections between notions. As a result the description of number of base relations between concrete notions of subject domain is formed, and the accordance between ontology notions and their lexical presentations in the test documents is set.

Semantic markup of electronic texts

Semantic markup includes stages: breaking up of electronic documents on fragments (sections, pages, paragraphs); assignment of set of tags to every fragment corresponding to metadata described in a fragment.

Extraction of knowledge from marked texts

Extraction of knowledge includes stages: forming of query for targeted knowledge extraction; search on the formed query in the library of marked texts; ordering of the found text fragments; filtration of found fragments package; presentation of knowledge from every fragment as the set of suggestions in the limited natural language; internal knowledge representation of formalized suggestions and their loading in a knowledge base.

S.A. Subbotin

The Methods of Synthesis and Models of Neuro-fuzzy Networks for Diagnosis and Classification Problem Solving on Features

Currently, there are several types of neuro-fuzzy network synthesis methods and models, which have different performance. Therefore the problem of determining the appropriateness of methods using for the in practical problem solving is actual. This problem is decided in the paper as a problem of evaluation of temporal and spatial complexity of the neuro-fuzzy network synthesis methods.

The complexity of the learning process for the Mamdani-Zadeh and Takagi-Sugeno-Kang neuro-fuzzy networks depends on the complexity of the network structure (number of connections and number of neurons), of the dimension of the training sample, of the complexity of the problem, of the weight initial values in the network, as well as the required accuracy and iteration complexity of the error backpropagation method.

The methods of synthesis of neuro-fuzzy networks with a non-iterative weight evaluation based on a preliminary analysis of the training sample and on the automatic selection of fuzzy terms. These methods do not require a user actions for the number and parameter

of fuzzy terms setting, automatically form a partition of feature space, allow to take into account the information about the feature and fuzzy term informativeness in the structure and parameters of the network, allow to ensure a high level of interpretability of constructed models.

For these methods the analytical estimations of temporal and spatial complexity are presented in the table. To simplify the resulting estimates put the assumption of the model dependencies and values of the parameters affecting the complexity. This would facilitate the assessment of complexity and makes them comparable to the different methods and models.

These complexity estimations show that the most expedient method is the method of synthesis of three-layer neuro-fuzzy network with binary output, the most economical use of memory has the three-layer network with discrete output synthesis method, the most rapid and least demanding of memory is a model of three-layer network with a binary output.

The analytical evaluation of temporal and spatial complexity can be used to automate the selection of the synthesis method and model of neuro-fuzzy networks for solving specific practical problems.

He.S. Tsibulnik, V.N. Pigus

Intellectual Control Unit of Access

However, if as a reading device to put the microcontroller having onboard АЦП and with big enough speed (25 MIPS and more) it is possible to improve security of the monitoring system of access considerably.

It is reached as follows. The internal power supply system of key DS1990A has no voltage stabilizer and the target field transistor of a key has final internal resistance. It means that level "0" at the key answer depends on pressure of an external food, capacity of the condenser and internal resistance of the target transistor of a key. If to measure values of pressure in the beginning of a phase of the answer of a key - presence pulse, in middle this phase and in the end of a phase, to write down in memory of the monitoring system of access these values together with a key code this data can serve as additional parameters of identification of a key.

For enough high-speed microcontroller it is possible to write the program which not simply sets parameters of time intervals (time slots) at key data reading, but also changes them at key data recording in access system. The data which correspond to the minimum duration time slots thus register.

Then key data reading can be spent to three stages: at the first reading standard time parameters are used, the additional data for this concrete key is taken, repeated data reading with the minimum values of time is made. The key is completely identified, if repeated reading is spent successfully and its data coincides with the first reading.

V.E. Tsyganash

The Analysis and Resolution of the Preferred Initiative Task during Electroheat Installation Control

In article the basic problems of management are formulated and analysed by the three-phase arc furnace which control system represents set of three automatic regulators and one adjustable object - furnaces into which energy is entered by three electric arches.

The situation of carrying out of fusion is considered at the increased share of a breakage in batch when regular systems of automatic control become ineffective, and the person-operator because of the possibilities limitation is compelled to make a decision in the conditions of the raised levels of uncertainty.

In such conditions traditional construction technologies using of control systems by powerful electrothermal installations, for example, on the basis of parametrical optimization, does not give desirable results. Therefore for decrease in level of uncertainty at decision-making the new method – a method of motions divisions on slow and fast is chosen, allowing substantially to reduce the level of fast movements, at approach of a working point of the unit to an optimum trajectory. It is carried out by introduction of absolute, portable and relative systems of readout for the variables characterizing the process. Thus results of the power supply interaction with the energy consumer are represented in relative system of readout. For better physical possibilities of coordination the person that making decision (operator) with variables of process, as a defining variable power gets out. Such choice is caused by range expansion of the analyzed signals representation, as towards low frequencies, and aside higher frequencies and decrease in regularity of signals space because of transition to a uniform defining variable, notably, to the actual capacity arriving in loading from the power supply.

Simplicity and efficiency of the offered method are confirmed by results of introduction on 12-ton of the arc steel-smelting furnace of Joint-Stock Company NKMZ. Its introduction has allowed to reduce considerably quantity of breakages of electrodes and to reduce the expense of the electric power on 20 – 30 thousand kilowatt-hours monthly.
