TEXT PROCESSING
AND
COGNITIVE TECHNOLOGIES

 # 1

1997

Moscow - Pushchino

Summary

The collection is the first issue in the series "Text Processing and Cognitive Technologies". It comprises six papers centred around the comparatively new branch of computer linguistics: text processing based on cognitive technologies, or knowledge-based text processing. The collection presented may beof interest for researchers and students specialising in the field of text processing.

ISBN: 5-201-14342-3
Text Processing and Cognitive Technologies. Paper Collection. N 1. Edited by A.G. Dyachko, - Ì oscow, Pushchino, ONTI PNTS RAN, 1997, 116 p.

Contents

page

TEXT PROCESSING
AND
COGNITIVE TECHNOLOGIES

The collection is the first issue in the series "Text Processing and Cognitive Technologies". The title mirrors the intention of compilers to select scientific works centredaround the comparatively new branch of computer linguistics: text processing based on cognitive technologies, or knowledge-based text processing. The necessity of such an approach has long since been declared. Yet the development of new means of computer communications in the form of Internet-like global networks dramatised the problem of information explosion making it only too evident that being confronted with a sea of available information we are quite helpless. That's why it is our opinion that all efforts to solve this problem must be encouraged and supported.

Presented are six papers reflecting brightly enough the wide range if investigations in this field of computer linguistics.

Replicating the life cycle of scientific ideas the results are represented by theoretical works (Shalyapina, Polyak), demonstrative prototypes of computer systems (Koch, Vazov, Polyakov) and commercial developments (Pominov).

The range of models of language proposed by authors extend from a generation model developed after Chomsky's grammar (Koch) to a model following in the wake ofMelchuk's Meaning <=> Text theory (Polyakov).

The papers describe wide range of phenomenological aspects of natural language: morphology (Pominov), syntax relations (Koch, Polyakov, Shalyapina), aspectual-tense values (Vazov), text structure (Polyak).

The authors tailor their systems to suit varying applications: from computer dictionaries (Pominov) to bibliography systems (Koch).

We believe that such a wide scope of investigations confirms the complexity of the problem of text processing. However, "complex" is not "unsuperable".

We are grateful to our sponsors, without whom the publication of this collection would not be possible. We thank the authors of the papers and invite new authors working in this field.

Academic of I.A.I., professor Anatoly Diachko

SUMMARIES

SEMANTIC ANALYSIS OF A SCIENTIFIC ABSTRACT USING A RIGORISTIC APPROACH

Short paper

Gregers Koch

DIKU, Department of Computer Science,
University of Copenhagen
Universitetsparken 1, DK-2100 Copenhagen, Denmark.
e-mail: gregers@diku.dk

May 14, 1996

For application in the design and implementation of natural language user interfaces to knowledge systems, we have developed some meta systems for inductive semantic analysis, building on dataflow analyses and hence implementing a kind of program synthesis. In this context, we present a semantic analysis of a particular text. A scientific abstract by S. Fong [3] is the selected text, and we construct a logical representation of the semantic content of the abstract, both manually and automatically using one of our inductive meta systems.

Key words: semantic analysis, logical representation, scientific abstract, dataflow analysis, inductive analysis, DCG

ÉVÉNEMENT OU ÉTAT, RÉCIT OU DESCRIPTION: LE RÔLE DU CONTEXTE DANS L'IDENTIFICATION DES VALEURS ASPECTUO-TEMPORELLES EN FRANÇAIS

Nikolay Vazov

LIMSI-CNRS, B.P. 133, 91403 Orsay - ISHA, Université Paris IV - Sorbonne
adresse électronique: vazov@limsi.fr

Nous proposons une procédure pour identifier la valeur aspectuo-temporelle d'une phrase ou d'un paragraphe contenant des formes du Passé Composé. L'analyse de ces valeurs s'effectue par un ensemble de regles d'exploration contextuelle qui s'appliquent selon le type d'unité textuelle a la phrase ou au paragraphe. Les regles parcourent le texte considéré en cherchant a détecter les indices qui permettent a décider sa valeur aspectuo- temporelle. Celle-ci est représentée par les primitives de la topologie temporelle: intervalles fermés ou ouverts.

Mots-cléfs: événement, état résultant, récit, description, regles d'exploration contextuelle

ARGUMENTATIVE TEXT: VERBAL EMBODIMENT OF THE COGNITIVE PROCESS OF ARGUMENTATION

Olga Polyak

Pushchino State University
polyak@adm.pgu.serpukhov.su

Presented is a concept of the argumentative text as a reflex of the cognitive process of argumentation composed of specialized sequences of peculiar cognitive operations. Verbally realized cognitive operations assume the status of predominantly assertion-type illocutions. The attempt to distinguish these illocutions boils down to the analysis of the semantic structure of the underlying operations. Meanwhile isolating illocutionary macrostructure text units amounts to elucidating the canonical text macrostructure which is based on the semantic structure of the governing cognitive event as a whole.

The governing cognitive event consists in justifying decisions and can be represented as a frame with a procedural attachment. Its slot names denote the contents of the mental states of the subject - information fragments actualized as decision-making progresses - and the procedures are argumentative operations carrying decision-making forward (see Fig.1).

Figure 1.
Macroillocution of Argumentation
(with propositional contents on the left and illocutionary component on the right):

Contents of mental states	Argumentative operations:
(i) initial situation (ii) adopted values (iii) problem situation (iv) set of possible decisions (social or mental actions) (v) set of feasible results of contemplated decisions (vi) optimum result (from the standpoint of values) vii) optimum decision (from the standpoint of result)	A. Actualization of information initial situation and adopted values - illocutions: Information, Description B. Actualization of potential decisions, results and/or means - illocution: Prediction C. Comparison of initial situation, potential results and/or means(and decisions with adopted values - illocution: Evaluation
(viii)set of possible means (ix) optimum means (from the standpoint of result) (x) proposition P representing decision made	D. Choice of problem situation, optimum result and/or means and decision - illocutions: Argumentative Statement, Proposal

The structure of the argumentative text is predetermined by the inherent ambiguity of the Argumentative Statement: the Speaker making an Argumentative Statement and viewing P as true - that is assigning proposition P the value of truth within the limits of his world model - simultaneously admits a possibility of non-P taken as true. This ambiguity calls into playa special mechanism - the mechanism of cognitive necessitation accounting for the development of argumentation.

Key words: decision-making, cognitive argumentative operations, illocutionary discourse structure, speech act.

LANGUAGE ABILITY MODELLING USING COMPUTER SYSTEM "NEDOROSL"

Vladimir Polyakov

Department of Computer Science and Economy of
Moscow State Steel and Alloys Institute (Technological University).
Pushchino Research Centre of RAS, tan@ibpm.serpukhov.su

The paper describes the problem of computer system development for natural language sentence processing using formalisms of Meaning Units Grammar and Manyaspect model of sentence meaning. Formulated are the target and tasks that were tackled when developing the laboratory computer system "Nedorosl". The main components are described: functions, data base, user's interface and concrete design solutions adopted. During the development of system "Nedorosl" the ideas underlying Meaning units grammar and Manyaspect model of sentence meaning were tested. Though "Nedorosl" is a laboratory system developed for exploration aims, hence it can be used as a base for automated working place of linguist and knowledge engineer for creation of system, using semantic text processing (machine translation systems, knowledge base interfaces, text knowledge acquisition systems etc.).

Key words: semantics, linguistic processor, Meaning units grammar, Manyaspect model of sentence meaning, grammar parsing.

SOME QUESTIONS OF COMPUTER DICTIONARIES DEVELOPMENT

Andrey Pominov

pom@rsc.msk.su

An approach to the creation of automatic dictionaries is described featuring the system Multitran implemented under Windows. The model is based on the open systemarchitecture which allows to append and modify linguistic information offering the user a flexible tool of dictionary maintenance. The morphological scheme includes identification of any form for words stored in the system as well as synthesis of all other forms. The translation scheme covers processing of single words and phrases allowing to retrieve relevant sets of translations. Appending procedure allows to update both stem and translation dictionaries using an untuitive user-oriented approach. Additional linguistic features and modes allow to fine-tune the dictionary.

Key words: computer dictionary, morpholical model, text translation, lexicography, opened systems, machine translation

A FORMALISM FOR SEMANTIC DESCRIPTION OF WORDS AND WORD-COMBINATIONS

Z.M. Shalyapina

Institute of Oriental Studies
of the Russian Academy of Sciences, Moscow

A formalism is introduced for representing meanings of linguistic units within an entity-oriented framework of linguistic description. The framework envisages, among other things, that syntagmatic relations SyntR between linguistic entities LE of any level (from morphemes to sentences to supra-sentential unities) are seen as actualizations of these LE's valencies. The formalism is based on the componential view of lexical semantics, extending it to all LE levels. Meanings of LE, semantic interpretations of SyntR between them, markers of paradigmatic and syntagmatic semantic classes are uniformly represented as semantic formulas SF - augmented dependency structures of semantic elements SE and semantic relations SR, the latter determined by reference to SE's valencies and specified as to the "communication modes" in which they occur as these valencies' actualizations. The SR that can link non-elementary SF (and thus serve as interpretations of SyntR between the LE whose meanings are represented by these SF) are defined by reference to the "free" valencies of the SE contained in the SF in question and to the positions in these of the occurrences of the SE with the corresponding valencies. In this way, semantic interpretations of SyntR can, in the general case, be as complex as the componential representations of LE's meanings, which lends this formalism more power for semantic description of syntax than the usual "deep case" mechanisms.

Key words: formalism for representing meanings of linguistic units, entity-oriented framework, componential view, deep case, syntagmatic relation, semantic element, semantic relation, communication mode, valency.

CALL FOR PAPERS

Dear colleagues,

Committee of the paper collection invites authors who works in the fieldof scientific text processing using cognitive technologies.

These are the topics of papers possible:

- scientific text processing(STP) as application of artificial intelligence and computer linguistics;
- scientific text as object of exploiration: kinds, structure and forms of scientific text;
- scientific text as result of scientific investigation;
- scientific text as synthesis of natural-language, graphs, formulas and other kinds of information;
- aims, means and methods of STP;
- scientific text and scientific knowledge;
- scientific text and context;
- cognitive models and cognitive processes connected with STP;
- applications connected with STP: documental interfaces to the bases of scientific data and knowledge; information searching systems, bibliography systems, thesauruses, text processors, etc.
- hardware and software for STP;
- other topics related.

Please, contact us at the e-mail address:

tan@ibpm.serpukhov.su (for Polyakov)
polyak@adm.pgu.serpukhov.su (for Polyak)

or at the postal address:

Vladimir Polyakov
P.O.B. 97
142292 Pushchino, MO
Russia

Yours sincerely,
Olga Polyak,
Vladimir Polyakov.