

The system is a powerful tool for fish resource estimation and prediction. Along with that it provides calculation of admissible catches and allows distributing fishing quotas optimally between regions and states. The system consists of several specialized intelligent subsystems which co-operate in local or global networks. Specialized systems include knowledge bases on food fish, food state, hydrologic assessment etc.

2. Text categorization system "Cluster"

The program system is designed for categorization of Russian text informational messages. The system determines whether a given text in Russian language belongs to one of predefined classes of texts. User-defined hierarchical classifier is employed.

The system operates in two modes:

- system training using learning samples of texts or expert help;
- classification of input texts.

The system extracts terms (both single- and multi-word) using shallow parsing approach. The conventional statistic estimation of term significance is supplemented with analysis of forms of term occurrence in the text.

3. Control system for high-safety space vehicle maneuvering

The system was created in co-operation with Russian State Scientific-Research Experimental Center of astronaut training to study processes of automatic docking and manual piloting of a space vehicle (SV). The system solves three problems in series, one for each item in the complete set of possible safe results of SV maneuvering: rendezvous, hovering, safe passage after a rendezvous. This allows to accomplish comprehensive on-line analysis of situation and to give complete information about safe rendezvous to the operator. A research prototype of the system has been designed and implemented for now. It can be used as a pilot training simulator for astronauts.

4. Automatic detection of newsmaking

The 'Newsmaking' task: extract information about the *events of newsmaking* (not to be confused with *newsmaking events*). Newsmakers are people or organizations that make informational announcements. The system solves the following tasks:

- detection of newsmaking as a fact;
- newsmaker identity (accurate to anaphora resolution);
- newsmaker type identification (person or organization).

The system includes the following modules:

- graphematical and morphological analysis;
- applied syntactic analysis (fragmentation, extraction of noun phrases of different structure);
- applied semantic analysis (identification of newsmaking domain entities);
- pronominal anaphora resolution for newsmaking domain entities.



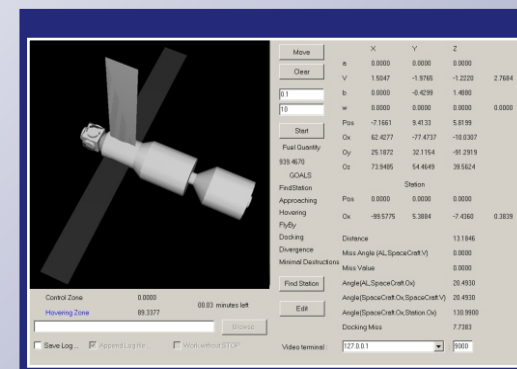
Artificial Intelligence
Research Center

Program Systems Institute
Pereslavl-Zalessky 152020 Russia
Phone, Fax: +7(48535)98065
E-mail: airec@botik.ru
Web-site: <http://www.botik.ru/PSI/AIReC>

Russian Academy of Sciences
Program Systems Institute



Artificial Intelligence
Research Center



Pereslavl-Zalessky
2005

Artificial Intelligence Research Center (AIReC)

is a research division of PSI RAS.

The centre emerged from the Laboratory of Knowledge Representation Problems in 1990. AIReC is one of the leading research groups in artificial intelligence area. The employees of the centre possess vast experience and have shown theoretic results in the field of artificial intelligence. The staff of AIReC includes 6 doctors of science, 3 candidates of science (Ph.D.), and 11 programmers and researchers. Special attention is paid to post-graduate and undergraduate student research.

Director of AIReC E.P. Kurshev, Ph.D.
Deputy Director A.N. Vinogradov, Ph.D

Structure of the center

Laboratory of Intelligent Dynamic Systems

- Knowledge-based dynamic systems: theory and applications
- Dynamic knowledge bases
- Information technologies for diagnostic and treatment support
- Co-operative problem solving in distributed intelligent systems

Laboratory of Intelligent Control

- Adaptive control of complex motions
- Intelligent control of complex dynamic objects
- Image analysis of data, cognitive graphics and methods of representation of dynamic information
- Intelligent video- and telemetric information systems
- Parallel geometry processors for motion control

Laboratory of Intelligent Internet Technologies

- Semantic information retrieval on the Internet: ontology-based approach and multi-agent support
- Information extraction
- Automatic text categorization
- Development and specification of a coherent system of ontologies for information retrieval task
- Cognitive models of personality

Research areas:

1. Knowledge-based dynamic systems.

A class of dynamic systems that use expert and empirical knowledge as control rules is researched. Status variables of such systems can be of quantitative or qualitative (logical or linguistic) nature.

*Fields of application:

- modeling of purposeful behavior of complex technical systems, robots in particular;
- modeling of ecological systems, e.g. water basins;
- modeling of social systems, e.g. social tension and regional conflict analysis.

*Results achieved:

Behavior of knowledge-based dynamic systems was investigated, as well as sets of attainable states, criteria of attainability, requirements for existence of purposeful behavior plans, rules for choice of behavior goals, algorithms of automatic plan synthesis, etc. Architecture of appropriate program systems was developed. Experimental software tools for development of application systems were implemented.

A number of application systems were developed, in particular, a system for modeling automatic docking of an active spaceship and an orbiting station, and some others.

2. High-precision semantic information retrieval on the Internet.

Project goal: essentially reducing the amount of irrelevant information delivered by existing search engines. Special methods of natural language query semantic analysis are proposed for this purpose. The results of the analysis are used for semantic filtering of documents.

Main features of the method:

- capability to use existing search resources (resources are added in a semi-automatic manner);
- natural language query support;
- domain-independent surface semantic analysis;
- automatic summarization and categorization of documents.

Results achieved:

Appropriate methods of morphological, syntactical and semantic analysis were investigated, and a research prototype of an intelligent meta-search system was developed. The basic principles of this system are described above.

Preliminary experiments have shown essential increase of search relevance, in comparison with existing methods.

3. Information extraction

Objective: extract meaningful information of a pre-specified type from (typically large amounts of) texts for further analytical purposes. The result of information extraction is presented with data structures of a pre-specified format filled scenario templates. Both the type of information sought and the output format are strictly determined by the task specification. The extracted information typically includes events of a certain kind and their attributes: time, location, participants; as well as relations between these entities. The filled templates can then be stored in a database for subsequent retrieval and usage.

Possible application scenarios include (but are not limited to) inference of new information (knowledge discovery); query formulation and answering in human-computer systems; automatic generation of abstracts and summaries; visualization of document content. Information extraction is useful for social applications (mass media monitoring; social situation analysis given its reflection in mass media; conflict detection and prevention) as well as for business intelligence (customer feedback analysis; R&D, marketing tasks).

Information extraction technology allows processing unrestricted real-world text containing large numbers of proper names, irregular punctuation, irregular capitalization, "creative spelling", etc. The technology assumes no in-depth text understanding (only a relatively shallow level of understanding in a predefined domain area) therefore it is possible to process relatively large volumes of text in reasonable time.

Information extraction is based on a variety of natural language processing (NLP) techniques: text pre-processing, morphological analysis, parsing, special kinds of semantic analysis, pattern matching etc.

Results achieved:

NLP methods and algorithms (morphology, parsing, complex sentence segmentation, pronominal anaphora resolution for a fixed domain) were developed and implemented in a research prototype of information extraction software tools. The prototype employs a CPSL-like language for information extraction rule representation. It was developed with respect to portability and has been successfully tested in Microsoft Windows and Linux (cluster) environments.

Accomplished projects

1. Technology for distributed integrated decision-support system development *SIMER+MIR*

SIMER+MIR technology is a highly integrated complex of software tools, which support all stages of intelligent distributed system development and of their adaptation for solving applied tasks.

This technology includes software tools for

- knowledge acquisition;
- data mining;
- situation analysis and decision making (using obtained knowledge and relations).

SIMER+MIR technology supports quick generation of appropriate configuration for distributed decision-support system. The technology may prove useful in such fields as medicine, management, justice, military science, sociology, demography, etc. A tool for automatic reasoning included in the technology solves the tasks of prediction, diagnosis and interpretation of data as well as quality assessment, resource evaluation and localization, hypothesis proof/refutation. The final product is generated quickly and easily: in most cases no special knowledge in software engineering is required. SIMER+MIR technology allows to deliver the final product in term from two weeks to three months (depending on the variety of functions being implemented).

A number of applied intelligent systems are based on SIMER+MIR technology, among them a distributed integrated intelligent decision support system for estimation, prediction and quoting of biological fish resources in Caspian and Black Sea basins.