Datasheet of the IKTA4/042 project

I. *Detection of statistical structure of large databases and development of a human interface*

Project start: January 1, 2002, duration: 23 months.

Amount of support: KHUF 44 200, total project cost: KHUF 88 780.

Project leader: ***Nagy Ákos Dr.***
BME Irányítástechnika és Informatika Tanszék
H-1117 Budapest, Pázmány Péter sétány 1/d
<***http://akos@iit.bme.hu***>, phone: +36 (1) 463-2699

Project URL: <>

II. Consortium members (number of members = **2**, the first member is the project co-ordinator)

no	*name*	*support*	*total cost*
1.	Budapest University of Technology and Economics, Dept. of Control Engineering and Information Technology	KHUF 7 600	KHUF 9 180
2.	DSS Consulting Ltd.	KHUF 36 600	KHUF 79 600

III. Public presentations

No presentation is available.

IV. Goals of the project

Data coming from a particular structure (e.g. physical, biological, social) of the real world are tipically arranged in large databases. There are several evident and less evident relationships in the data. Data generated by events not described by strict laws are good candidates for being governed by statistical rules that are not evident by intuition. It is very general that statistical connections among data build up a network we call probability net.

The goal of the project is the development of a methodology and software that makes possible of discovering probability nets in large databases and the generalization of this knowledge to new samples according to the goals of the user.

The goals of the methodology part is to prepare a document, that contains the steps of working up a large database: description, categorization and hierarchies of variables and of missing variables, steps of working with experts, questionnaires and schemas of interviews. Further it contains a list of the typical questions addressed to the database, categories of questions, mathematical solutions for these categories, possibilities of reasoning and optimalization. This document sums up the steps of exploration of probability nets and networks and the possible applications of this knowledge base related to the question categories.

The goal of software development is to implement the mathematical algorithms underlying the answers of the typical questions. The software can be divided into two components: the component designed for detecting statistical relation systems (structural component) and the component that can apply this knowledge to new cases (generalization component). The structural component is capable of generating subsystems and detecting hidden variables. The task of the generalization components is the application of the knowledge, cumulated by the structural component, to new cases; i.e. to make predictions of unknown data based on given variables. The generalization component is complemented with classification and optimization subcomponents.

Mathematical representation of the knowledge base is chosen to be a class of graphical models called probability or Bayes nets. These models represent every variable in a common model that makes possible reasoning without limits, working and handling with and of missing data, discovering hidden variables and integration of expert knowledge.

V. Project results (in case of finished projects)

The project is not finished.

VI. Data on consortium members (number of members = **2**)

1. ***Budapest University of Technology and Economics, Dept. of Control Engineering and Information Technology*** (co-ordinator)

URL: <>

Support for the co-ordinator: KHUF 7 600, and its total cost: KHUF 9 180.
Contract number: .

Team leader: ***Nagy Ákos Dr.***
BME Irányítástechnika és Informatika Tanszék
H-1117 Budapest, Magyar tudósok körút 2.
<***http://akos@iit.bme.hu***>, phone: +36 (1) 463-2699

2. ***DSS Consulting Ltd.***

URL: <***http://www.dss.hu***>

Support for the consortium member: KHUF 36 600, and its total cost: KHUF 79 600.
Contract number: .

Team leader: ***Aszalós Péter***
DSS Consulting Kft.
H-1025 Budapest, Józsefhegyi út 28-30.
<***http://aszalos@dss.hu***>, phone: +36 (1) 345-0900