

## Datasheet of the IKTA3/027 project

### I. Development of Part-of-speech Tagger Method for Hungarian by Applying Machine Learning Algorithms

Project start: October 1, 2000, duration: 21 months.

Amount of support: KHUF 36 750, total project cost: KHUF 73 500.

Project leader: **Gyimóthy Tibor**

SZTE Informatikai Tanszékcsoport

H-6701 Szeged, Árpád tér 2.

<<mailto:gyimi@inf.u-szeged.hu>>, phone: +36 (62) 544-139

Project URL: <<http://www.inf.u-szeged.hu/~gyimi>>

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

<i>no</i>	<i>name</i>	<i>support</i>	<i>total cost</i>
1.	University of Szeged, Dept. of Informatics	KHUF 18 250	KHUF 36 500
2.	MorphoLogic Ltd.	KHUF 18 500	KHUF 37 000

### III. Public presentations

Presentation #1 (28-február-2001): HTML (44 KB)

Presentation #2 (21-november-2001): PPT (144 KB)

The presentations are accessible through a special webpage summarizing the project presentations.

### IV. Goals of the project

The goal of the project is to develop a Hungarian part-of-speech tagging method, prototype system and technology, including the following elements: (1) Establishment of a medium-size (1 million word long) manually annotated Hungarian learning corpus that can be used for other natural language purposes beyond the scope of this project. (2) Efficient disambiguation of texts (in different domain). (3) To develop an adaptable system that is able keep track of the changes in the (spoken or written) language. (4) The chosen technology should be applicable to other European natural languages. (5) The system should use internationally accepted part-of-speech (morphological) classes, augmented with the special attributes of Hungarian language. (6) The per-word accuracy of the solution (i.e. the number of well-tagged words divided by the number of all words in the given text) should reach at least 97-98%. The above issues summarize the technological goals. In addition to this, from the scientific point of view, papers dealing with applications, accuracy, combinations, and limits of learning algorithms can be of international interest.

The abstract of the project – PDF (24 KB) – is available via the webpage summarizing the basic project data.

### V. Project results (in case of finished projects)

The project is not finished.

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

#### 1. University of Szeged, Dept. of Informatics (co-ordinator)

URL: <<http://www.inf.u-szeged.hu/~gyimi>>

Support for the co-ordinator: KHUF 18 250, and its total cost: KHUF 36 500.

Contract number: OMFB-02508/2000.

Team leader: **Gyimóthy Tibor**

SZTE Informatikai Tanszékcsoport

H-6701 Szeged, Árpád tér 2

<<mailto:gyimi@inf.u-szeged.hu>>, phone: +36 (62) 544-139

2. *MorphoLogic Ltd.*

URL: <<http://www.morphologic.hu>>

Support for the consortium member: KHUF 18 500, and its total cost: KHUF 37 000.

Contract number: OMFB-02509/2000.

Team leader: ***Prószéky Gábor Dr.***

MorphoLogic Kft.

H-1118 Budapest, Késmárki utca 8

<<mailto:proszeky@morphologic.hu>>, phone: +36 (1) 361-4721