Datasheet of the IKTA3/063 project

I. *Linguistic Feedback for OCR, Hand-writing and Speech Recognition Systems*

Project start: October 1, 2000, duration: 17 months.

Amount of support: KHUF 9 600, total project cost: KHUF 19 200.

Project leader: ***Prószéky Gábor Dr.***
MorphoLogic Kft.
H-1118 Budapest, Késmárki u. 8.
<***mailto:proszeky@morphologic.hu***>, phone: +36 (1) 361-4721

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

II. Consortium members (number of members = **1**, the first member is the project co-ordinator)

no	*name*	*support*	*total cost*
1.	MorphoLogic Ltd.	KHUF 9 600	KHUF 19 200

III. Public presentations

Presentation #1 (28-február-2001): HTML (120 KB)

Presentation #2 (21-november-2001): PPT (200 KB)

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

IV. Goals of the project

Digitizing of texts in non-digital formats has become a key issue in all languages of the world. The first solutions were developed for optical character recognition (OCR), but, more recently, handwriting recognition and speech recognition have come into the focus of research and development. In these two areas, however, the system has to cope with a much higher level of input uncertainty. - Applications for the Hungarian language are especially complicated. Whereas, in English, statistical methods allow a quite reliable correction of input uncertainties, in Hungarian they are practically unusable, due to the agglutinative nature and the highly free sentence constructs of the language. - In the case of speech recognition, for example, where word limit recognition is also a difficult issue, each input sentence waveform may lead to millions of possible character strings. Earlier systems were unable to handle such high level of uncertainty: extra machine support had to be developed. - MorphoLogic's solution is based on two ideas. On one hand, it is assumed that the system receives all the uncertainties represented as real alternatives (i.e. as character net of alternative strings) from an underlying module. Secondly, all generated texts (alternatives) undergo a linguistic check, and only those variants are accepted that pass this test. This check is performed by the most efficient tool available, i.e. using a finite state automaton, which can also detect the last character up to which an alternative can be accepted as a correct phrase. The system also handles spaces (word limits), thus even the words, which sound as one in the spoken language can be separated with high precision. - Our solution is language and application independent. Its speed exceeds any everyday demand, as it completes the most effective filtering, the world level check in a fraction of a second on all the alternatives of a sentence. It is an efficient tool for the reconstruction of Hungarian texts. However, taking advantage of the company's linguistic databases, MorphoLogic is also testing the system on texts in other languages. The solution can be equally used in OCR, handwriting and speech recognition applications. The material on the Hungarian language has been complemented with underlying databases to support OCR and phonetic recognition.

The abstract of the project – PDF (16 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 = **1**)

1. ***MorphoLogic Ltd.*** (co-ordinator)

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

   Support for the co-ordinator: KHUF 9 600, and its total cost: KHUF 19 200.
   Contract number: OMFB-02466/2000.

   Team leader: ***Prószéky Gábor Dr.***
   MorphoLogic Kft.
   H-1118 Budapest, Késmárki u. 8.
   <***mailto:proszeky@morphologic.hu***>, phone: +36 (1) 361-4721