

EDC-22071 KOM2002/27938

Groupware for Distributed Content Production

KOM2002

Prototype of English to Swedish Translation Engine

Security (distribution level)	Public
Contractual date of delivery	31.12.2003
Actual date of delivery	30/12/2003
Deliverable number	D 8.5
Deliverable name	Prototype of English to Swedish translation engine
Type	PP
Status & version	Final report
Number of pages	
WP contributing to the deliverable	WP8
WP / Task responsible	SYSTRAN
Other contributors	
Author(s)	Elsa Sklavounou
EC Project Officer	Kimmo Rossi
Keywords	Machine Translation Engine, Prototype, Intuitive Coding
Abstract (for dissemination)	This report aims to present the integration and the prototype of English to Swedish with all the modules required in order to accomplish the KOM2002 purposes.

TABLE OF CONTENT

<u>1.</u>	<u>SUMMARY</u>	3
<u>2.</u>	<u>PRESENTATION</u>	3
<u>3.</u>	<u>CONCLUSION</u>	4

1. Summary

SYSTRAN provided to the Web4health interface with the machine translation services English to Swedish. The developed prototype can consult the external customization resources constructed via the terminological workbench and provide customized translation output.

2. Presentation

The English to Swedish machine translation engine has been developed from scratch with all the specific features needed to satisfy the KOM2002 needs.

The developed prototype is composed by three internal modules

Analysis where all updated modules effectuate the analysis of the English input,

Transfer where there is a typological transfer of encoded linguistic rules produced from the analysis module in order to be exploited by the last internal module the synthesis

Synthesis where all encoded linguistic information is transferred in order to be synthesized into Swedish.

The system resources, that is to say the resources used by the executables were constructed based on coverage measurements of high frequency textual data.

For the customization phase of the medical content of Web4health it was of high importance the construction of the Intuitive Coding modules for English and Swedish. The executable is developed with the option to be able to get forced by an external-to-the-system customization dictionary built via the terminological workbench adapted according to the Web4Health needs. For that purpose guess processes for unknown to the system words and composition of rules have been implemented for the Swedish language.

Moreover the English to Swedish prototype has been developed with the perspective to incorporate easily SYSTRAN's New Generation (NG) architecture in the synthesis and optionally the transfer part of the system. With the integration of the NG technology, linguistic improvements can be made more easily through declarative rules and the engine can benefit from future extensions of the NG transfer and synthesis mechanisms.

3. Conclusion

Web4Health content can be translated into Swedish. The raw machine translation output can be customized or improved thanks to the construction of an external dictionary constructed via the terminological workbench. For those purposes the prototype has been conceived and developed on solid and flexible architecture so as to constitute a marketable translation engine with the perspective of further improvement according to the evolution of SYSTRAN's research.