



Central and Southeast European Resources (CESAR)

European Commission
The Information and Communication Technologies Policy Support Programme
ICT-PSP PB Pilot Type B
Project ID number: 271022
<http://www.cesar-project.net>

List of partners
Nyelvtudományi Intézet, Magyar Tudományos Akadémia (HASRIL), Budapest, Hungary
Budapesti Műszaki és Gazdaságtudományi Egyetem (BME), Budapest, Hungary
University of Zagreb, Faculty of Humanities and Social Sciences (FFZG), Zagreb, Croatia
Instytut Podstaw Informatyki Polskiej Akademii Nauk (IPIPAN), Warsaw, Poland
Uniwersytet Łódzki (ULODZ), Łódź, Poland
Faculty of Mathematics, University of Belgrade (UBG), Belgrade, Serbia
Institut Mihajlo Pupin (IPUP), Belgrade, Serbia
Institute for Bulgarian language Prof Lyubomir Andreychin (IBL), Sofia, Bulgaria
Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (LSIL), Bratislava, Slovakia

Project duration: February 2011 — January 2013

Summary

The main objective of the project – an integral part of META-NET – is to make available a comprehensive set of language resources and tools covering Bulgarian, Croatian, Hungarian, Polish, Serbian and Slovak. Building on a wide range of already existing resources and national or international activities, the project creates, populates and operates a comprehensive language-resource platform enabling and supporting large-scale multi- and cross-lingual products and services. The resources already involved in the project include interoperable mono- and multilingual speech databases, mono- and bilingual corpora, dictionaries, wordnets and relevant language technology processing tools such as tokenisers, lemmatisers, taggers and parsers. The main pillars of CESAR activity are seen as enhancement of resources and tools, adaptation of resources and tools to become compliant with the agreed standards for interoperability, upgrade of resources and tools by combining them with other resources and tools in order to achieve the foreseen level of interoperability and in adapting user-interfaces. A special effort is taken to achieve a common standard of involved resources and tools in order to enhance and facilitate the foreseen interoperability between them, as well as to evaluate their license schemes and IPR issues. A special track of activity is turning the linguistic development environment NooJ (www.nooj4nlp.net) into an open-source package running on most popular platforms. Key resources covered by the CESAR project will be linked and made interoperable using the facilities of the META-SHARE repository, which eventually will become an important component of a language technology marketplace for HLT researchers and developers, policy makers, language professionals (translators, interpreters, content and software localisation experts, etc.), as well as for industrial players, especially SMEs, catering for the full development cycle of HLT, from research through to innovative products and services.