CESAR

CENTRAL AND SOUTH-EAST EUROPEAN RESOURCES

■ Project partners
● Languages covered

A true cross-European project providing the language resources, tools and services that are inclusive and universal, but at the same time focusing specifically on a number of European languages that are currently underresourced: Bulgarian, Croatian, Hungarian, Polish, Slovak, and Serbian.

Project duration

2011-02-01 – 2013-01-31

www.cesar-project.net

CESAR

Project partners

HASRIL: Research Institute for Linguistics, Hungarian Academy of Sciences

BME-TMIT: Department of Telecommunications and Media Informatics, Budapest University for Technology and Economics

FFZG: University of Zagreb, Faculty of Humanities and Social Sciences

IPIPAN: Institute of Computer Science, Polish Academy of Sciences

ULODZ: University of Łódź

UBG: Faculty of Mathematics, University of Belgrade

IPUP: Institute Mihajlo Pupin, Belgrade

IBL: Institute for Bulgarian Language Prof. Lyubomir Andreychin, Bulgarian Academy of Sciences

LSIL: Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences

Contact

Tamás Váradi
(project coordinator)
Research Institute for Linguistics
Hungarian Academy of Sciences
Benczur u 33
1068 Budapest
Hungary
Tel: +36 1 342 9372 ext. 6010
Fax: +36 1 322 9297
E-mail: varadi@nytud.hu

ICTPSP
ICT POLICY SUPPORT PROGRAMME

META-NET

META-NET

CESAR

CENTRAL AND SOUTH-EAST EUROPEAN RESOURCES

CESAR
CENTRAL AND SOUTH-EAST EUROPEAN RESOURCES

Main goal

CESAR project is oriented towards the stimulation of ICT-based cross-lingual communication, collaboration and participation and thereby contribute to the creation of a pan-European digital single market. ICT research has started in Central- and Southeast-Europe with a lag behind Western European countries and has experienced rather short-term and fortuitous cooperation on the level of national research. Language technology has emerged in the respective participating countries autonomously, i.e., with national support both in the academic and in the private sector. As a result, the resources developed often reflect the isolated circumstances of their cre-ation, and still often lack standardisation.

European dimension

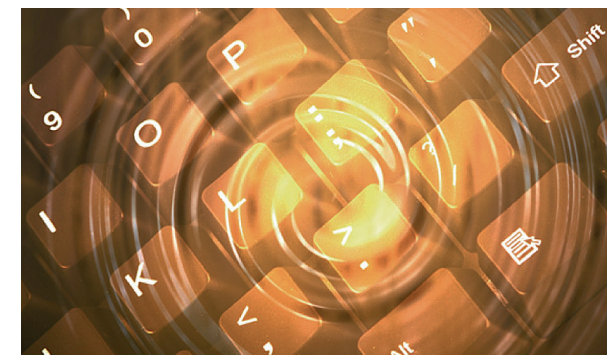The CESAR project is fully committed to providing synergy with relevant strategies and activities at the European, regional and national level. The partners are key players in the language technology scene for their respective languages and have built extensive links with LRT centres across the whole of Europe but especially in the geographical region that the consortium covers. At the same time, they have served and will continue to serve as catalytic force in the language technology scene of their countries. Typically, they have led national consortia involving the other major partners in the countries and have an established track record of their capability to mobilise their national partners.

The European and global dimension of the CESAR project results will be visible through the META-NET Multilingual European Technology Alliance and its sharing facility META-SHARE. CESAR project resources and tools will also be accessible through other types of linguistic infrastructures as well as partners' individual sites.

Target audience

Target user community of language resources and services embraces all stakeholders at the modern digital market: everyday end-users, professional end-users (business, administration, media, education, libraries, etc.) and expertise holders (researchers, industrialists, policy makers, etc). Our concern is a careful investigation of the needs of various types of users – from individual users to large multinational organisations – from the perspective of the current status as well as from the near future prospects.

Languages and technology involved

The project will make available a comprehensive set of language resources and tools covering the following languages:

- Bulgarian
- Croatian
- Hungarian
- Polish
- Serbian
- Slovak.

Resources will include interoperable mono- and multilingual:

- speech databases
- corpora
- dictionaries
- wordnets.

Relevant language technology processing tools will include:

- tokenisers
- lemmatisers
- taggers
- chunkers and parsers.

Depending on local capacities and efficiency considerations, the resources and tools may be either made available as web services through partners' sites or could be contributed through the META-SHARE digital exchange facility.

www.cesar-project.net

CESAR is a part of
www.meta-net.eu

META-SHARE