

# Preface

Language technologies and tools (hereafter called language resources) increasingly require sophisticated infrastructures to share, deploy as services, and combine for supporting research, development, innovation, and collaboration. To meet this need, several infrastructures have been already established over the past few years, such as Language Grid, Language Application Grid, META-SHARE, MLI, and PANACEA. The main theme of the International Workshop on Worldwide Language Service Infrastructure (WLSI) is technological and institutional challenges that are significant for constructing a worldwide interoperable language service infrastructure. The first workshop focused on language service infrastructures in Asian areas, and a related workshop, Language Technology Service Platforms: Synergies, Standards, Sharing (LTSP 2014), was held at the ninth edition of the Language Resources and Evaluation Conference (LREC 2014). The aim of LTSP 2014 was to provide a forum to enhance international cooperation and sustainable collaboration among worldwide initiatives. The second workshop was held during January 22–23, 2015, in Kyoto, Japan.

The workshop featured five prominent invited speakers: Toru Ishida from the Department of Social Informatics, Kyoto University, who introduced intercultural collaboration activities of the Language Grid; Nancy Ide from the Department of Computer Science, Vassar College, who presented the Language Application Grid framework to create custom natural language processing applications; Khalid Choukri from the Evaluations and Language Resources Distribution Agency, who explained the MLI Hub Project that aims at compiling the specification of the next generation of language grids; Núria Bel from the Department of Translation and Language Sciences, University of Pompeu Fabra, who reported characteristics of users in humanities and social sciences in the Spanish CLARIN Center; and Nicoletta Calzolari from the European Language Resource Association, who summarized policy issues related to language service infrastructures. The first four invited speakers are leaders of each ongoing project in Asia, the USA and Europe, and the last one is a representative of the association to promote language resources. The workshop included 11 oral presentations, and four posters. Participation in the workshop was by invitation only, and there were 29 professionals from 10 countries: China, France, Greece, Indonesia, Italy, Japan, Spain, Thailand, the USA, and Vietnam.

This volume includes 14 selected papers presented at the workshop. The papers are categorized into four parts. The first part introduces metadata and annotations to describe what kind of functionalities and annotations language services provide, and how to invoke the language services and convert the output of a language service to the input of another service. In META-SHARE, Piperidis et al. have focused on processing language datasets with appropriate linguistic annotation services such as tokenization, POS tagging, lemmatization, dependency parsing, and so on. On the other hand, in the Language Application Grid, Ide and Verhagen have addressed the language service interoperability to combine various services by defining Web Service Exchange

Vocabulary (WS-EV), which specifies a terminology for a core of linguistic objects exchanged among linguistic annotation services, and LAPPS Interchange Format (LIF), which represents linguistically annotated data including WS-EV for Web service invocations.

The second part provides technologies for service platforms that compose atomic language services across different interfaces, policies, and licenses. Ide et al. have proposed the Language Application Grid platform that enables language service composition using Galaxy workflow engine in workflow layer, LIF in messaging layer, and WS-EV in vocabulary layer. To solve licensing issues, Cieri and DiPersio have proposed the Language Application Grid license schema by establishing two classes of enforcement, requirement and notification. Mai et al. have tackled policy-aware language service composition by modeling the parallel execution policy of atomic language services. Moreover, Otani et al. have introduced Language Mashup to combine different licensed services, commercial language services, and open-sourced language services.

The third part focuses on the development of language resources and services, especially low-resource languages. Aili and Wushouer describe how to build Uyghur language resources such as dependency Treebank and grammatical information dictionary. Martadinata et al. explain how to implement a language identification tool with Wikipedia corpus and Twitter data. This tool is useful for classifying social media posts like Twitter into several regional languages in Indonesia, which can contribute to monolingual corpora creation in those languages.

The fourth part collects reports on language service application. Luong et al. have developed a Vietnamese multimedia agricultural information retrieval service using a Vietnamese agricultural thesaurus. Liu and Gao have proposed an approach to mine the opinion polarity of songs based on song lyrics in a multilingual environment. Gratta et al. have presented the Cooperative Philology WordNet Platform (CoPhiWordNet) that connects different WordNets in both modern and classical languages such as the Ancient Greek WordNet, the Latin WordNet, the Italian WordNet, the Croatian WordNet, and the Arabic WordNet. Sornlertlamvanich and Kruengkrai have applied a semantic relation extraction approach based on simple relation templates to the Thai cultural database for generating knowledge maps and infoboxes.

We hope this book will strongly support and encourage researchers who are willing to utilize various language services worldwide to create customized language applications and multilingual environments. We are grateful to all the participants and those who have supported this workshop.

January 2016

Yohei Murakami  
Donghui Lin

Worldwide Language Service Infrastructure  
Second International Workshop, WLSI 2015, Kyoto,  
Japan, January 22-23, 2015. Revised Selected Papers  
Murakami, Y.; Lin, D. (Eds.)  
2016, X, 201 p. 71 illus. in color., Softcover  
ISBN: 978-3-319-31467-9