

LREC 2020 Workshop  
Language Resources and Evaluation Conference  
11–16 May 2020

**1<sup>st</sup> Workshop on Language Technologies for Government and  
Public Administration (LT4Gov)**

# **PROCEEDINGS**

Doaa Samy, David Pérez-Fernández and Jerónimo Arenas-García  
(eds.)

**Proceedings of the LREC 2020**  
**1<sup>st</sup> Workshop on Language Technologies for Government and**  
**Public Administration**  
**(LT4Gov)**

Edited by: Doaa Samy, David Pérez-Fernández and Jerónimo Arenas-García

**ISBN: 979-10-95546-62-7**  
**EAN: 9791095546627**

**For more information:**

European Language Resources Association (ELRA)  
9, rue des Cordelières  
75013 Paris  
France  
<http://www.elra.info>  
Email: [lrec@elda.org](mailto:lrec@elda.org)

© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License

## Introduction

Welcome to the LREC2020 Workshop on "Language Technologies for Government and Public Administration (LT4Gov)" in its first edition.

LT4Gov stems from the experience gained by the organising team within the Spanish Plan for Advancement of Language Technologies. This Plan is a reference model of initiatives from the Governmental bodies to promote the introduction and integration of HLT in the workflow of different public entities.

LT4Gov is grounded on a strong believe that Human Language Technologies (HLT) could highly contribute in the innovation and the digital transformation of the public administration either in general or in domain-specific actions related to subsectors such as Education, Health, Culture, Tourism, etc. Thus, this workshop LT4Gov will invite contributions and results of work carried out by the HLT community in relation with public and governmental aspects in the line of language resources or in the line of solutions/tools destined either to these entities or to the citizen.

The sector of governmental bodies and public administration is an important sector representing a considerable percentage of expenditure of the GDP. Expenditure averages lie between 25%-57% in different countries according to reports from OECD, EU and World Bank. Moreover, this sector handles daily huge amounts of data in different formats.

Processing these huge amounts of information by HLT could provide valuable, data-driven and evidence-based insights that would highly impact the workflow for civil servants, the policy-making and the public services offered to citizens in different domains: Health, Tourism, Justice and Law, Culture.

LT4Gov aims at bringing together initiatives where Human Language Technologies (HLT) are used within the context of governmental bodies and public administration in the different domains (Health, Tourism, Justice, Culture, etc.). Governmental bodies and Public Administration Entities could be providers of data or could be users or beneficiaries of solutions in different sub-sectors. In addition, they could provide solutions to enhance services destined to the citizens.

Within LT4Gov and Public Administration, we consider the following three scenarios:

- PublicData4LRs (Public Data for Language Resources) - Public Administrations as providers of data
- LT4PolicyMaking (Language Technologies for Policy Making) - Public Administrations as users
- LT4Citizens (Language Technologies for Citizens) -Public Administrations as providers of services to the citizens

LT4Gov offer six use cases from different countries across the globe in which authors share their experience on how HLT was used in real scenarios within Public Administration and Governmental context. In its first edition, we hope LT4Gov would be a starting point to capture the attention of both the HLT community and the Public Administration on the benefits of applying LT in use cases that would impact the public services offered to citizens or would ease the daily work of civil servants. Finally, we hope researchers and stakeholders would gain useful insights and could find some inspiration in the presented use-cases.

On behalf of the Organising Committee,

Doaa Samy

**Organizers:**

Doaa Samy, Instituto de Ingeniería del Conocimiento (Spain) and Cairo University (Egypt)  
David Pérez-Fernández, Plan de Tecnologías del Lenguaje, Secretaría de Estado para la Digitalización e Inteligencia Artificial (Spain)

**Program Committee:**

Jerónimo Arenas-García, Universidad Carlos III Madrid (Spain)  
María José Del Olmo-Toribio, Secretaría de Estado para la Digitalización e Inteligencia Artificial (Spain)  
David Griol-Barres, Universidad de Granada (Spain)  
Maite Melero, Barcelona Supercomputing Centre (Spain)  
Antonio Moreno-Sandoval, Universidad Autónoma de Madrid (Spain)  
Pablo Haya-Col, Instituto de Ingeniería del Conocimiento (Spain)  
Luis Fernando D' Haro, Universidad Politécnica de Madrid (Spain)

## Table of Contents

|  |    |
|--|----|
| <i>Development of Natural Language Processing Tools to Support Determination of Federal Disability Benefits in the U.S.</i>  |    |
| Bart Desmet, Julia Porcino, Ayah Zirikly, Denis Newman-Griffis, Guy Divita<br>and Elizabeth Rasch .....  | 1  |
| <i>FRAQUE: a FRAME-based QUEstion-answering system for the Public Administration domain</i>  |    |
| Martina Miliani, Lucia C. Passaro and Alessandro Lenci .....   | 7  |
| <i>Enhancing Job Searches in Mexico City with Language Technologies</i>  |    |
| Gerardo Sierra Martínez, Gemma Bel-Enguix, Helena Gómez-Adorno, Juan Manuel Torres Moreno,<br>Tonatiuh Hernández-García, Julio V Guadarrama-Olvera, Jesús-Germán Ortiz-Barajas, Ángela María<br>Rojas, Tomas Damerou and Soledad Aragón Martínez ..... | 15 |
| <i>Research &amp; Innovation Activities' Impact Assessment: The Data4Impact System</i>   |    |
| Ioanna Grypari, Dimitris Pappas, Natalia Manola and Haris Papageorgiou .....   | 22 |
| <i>The Austrian Language Resource Portal for the Use and Provision of Language Resources in a Language<br/>Variety by Public Administration – a Showcase for Collaboration between Public Administration and a<br/>University</i>                      |    |
| Barbara Heinisch and Vesna Lušicky .....   | 28 |
| <i>Legal-ES: A Set of Large Scale Resources for Spanish Legal Text Processing</i>  |    |
| Doaa Samy, Jerónimo Arenas-García and David Pérez-Fernández .....  | 32 |

## Workshop Program

*Development of Natural Language Processing Tools to Support Determination of Federal Disability Benefits in the U.S.*

Bart Desmet, Julia Porcino, Ayah Zirikly, Denis Newman-Griffis, Guy Divita and Elizabeth Rasch

*FRAQUE: a FRAMe-based QUEstion-answering system for the Public Administration domain*

Martina Miliani, Lucia C. Passaro and Alessandro Lenci

*Enhancing Job Searches in Mexico City with Language Technologies*

Gerardo Sierra Martínez, Gemma Bel-Enguix, Helena Gómez-Adorno, Juan Manuel Torres Moreno, Tonatiuh Hernández-García, Julio V Guadarrama-Olvera, Jesús-Germán Ortiz-Barajas, Ángela María Rojas, Tomas Damerau and Soledad Aragón Martínez

*Research & Innovation Activities' Impact Assessment: The Data4Impact System*

Ioanna Grypari, Dimitris Pappas, Natalia Manola and Haris Papageorgiou

*The Austrian Language Resource Portal for the Use and Provision of Language Resources in a Language Variety by Public Administration – a Showcase for Collaboration between Public Administration and a University*

Barbara Heinisch and Vesna Lušicky

*Legal-ES: A Set of Large Scale Resources for Spanish Legal Text Processing*

Doaa Samy, Jerónimo Arenas-García and David Pérez-Fernández