Language Technology to Support the Fight against Fraud

Language Technology is used at the Joint Research Centre’s Institute for the Protection and Security of the Citizen to help fight fraud against the Community budget.

The main efforts of the ongoing activity are to purchase or develop software which (a) selects and retrieves textual information from internal file systems or open sources on the internet, (b) analyses the contents of documents and extracts information from them, and (c) visualises the main contents of individual documents in document profiles, or of whole document collections in document maps. A major focus is on multilingual and cross-lingual applications.

Document Profile

Once different information aspects have been identified in the text, they can be displayed in an intuitive manner so that users can get a quick overview of the main contents.

Where possible, the documents are linked to multilingual thesauri and nomenclatures such as EURONOC and the Combined Nomenclature so that the contents can be displayed in a language different from the language of the text.

Visualisation of Geographical References Found in Text

The map shows geographical references extracted automatically from 1500 texts written in English, German, French and Spanish.

Document Map

In this document map, each of the 258 documents is represented by one dot, and documents which are similar to each other are placed near to each other. Groups of documents covering similar subjects form ‘islands’ which are annotated with keywords. By clicking on a specific area, more keywords and a list of the documents in this area are displayed. This map was produced using ThemeScape by Cartia Inc.

Contact: Ralf Steinberger
Ralf.Steinberger@jrc.it
Tel: +39 - 0332 786271