Enabling global identity
Protecting digital trust

28th Meeting of the Wiesbaden Group on Business Registers
Call for Papers

The Hague, 2 - 6 October 2023

Global Legal Entity Identification Foundation (GLEIF)

Session No. 6

"New data sources: Opportunity and challenges"

LENU – Using AI for legal form detection

The Legal Entity Identifier (LEI) is a 20-digit alpha-numeric code based on the ISO 17442 standard. It connects to the key reference information that enables clear and unique identification of legal entities. The Global LEI Repository is the transparency island in a cloudy environment – it provides open, free-of-charge, high-quality legal entity data with global coverage.

There are more than 2.3 Mio LEIs in the system. The reference data of these LEIs is collected, verified, and managed by a network of LEI issuing organizations across the globe. The embedded global standards in the data format and the established data quality framework ensure consistency and high-quality data..

Using the LEI data, GLEIF and our partners from Sociovestix Labs, developed a state-of-the-art open-source AI tool. The tool, called LENU (Legal Entity Name Understanding), automatically assigns standardized Entity Legal Form (ELF) codes (ISO 20275) to entities based on legal name and legal jurisdiction only.

As of today, we utilize a wide range of traditional Machine Learning models as well as more advanced Deep Learning (transformer) models that predict the ELF code for any legal name within a given jurisdiction. This enables public and private organizations of any size to start adopt the ISO standard for legal forms by assigning the codes easily and effortlessly.

In our paper, we demonstrate the end-to-end process from analyzing the feasibility of the initial idea, testing of different algorithms from simple string matching to more sophisticated neural networks, to making available the open-source tool for the public. We will also highlight the motivation and benefits of having standardized data as part of an organization’s dataset and the use cases we see for the example of standardized legal forms.