Singling out Legal Knowledge from World Knowledge.
An NLP-based Approach

FRANCESCA BONIN, FELICE DELL’ORLETTA
GIULIA VENTURI, SIMONETTA MONTEMAGNI*


1. INTRODUCTION

Scholars committed to modeling legal domain knowledge have widely acknowledged with the need for domain-specific knowledge organization, i.e. legal ontologies, where domain knowledge (legal knowledge) and knowledge of domains of interest to be regulated (referred to as world knowledge) are not mixed. However, as pointed out in Breuker and Hoekstra1, the indiscriminate mixture of the two types of knowledge is a common attitude in constructing legal ontologies. In particular, Breuker and colleagues speak of epistemological promiscuity, putting the emphasis on how this is a serious problem in core ontology development. They point out that many legal ontologies collapse together epistemological and ontological perspectives. Starting from the well-known assumption that “by its very nature, law deals with behaviour in the world”, they discuss how domain independent concepts of law are tained with common-sense notions which refer to social activities. Interestingly, they claim that “the domain ontologies [they] developed in the various project contained almost ninety-nine percent terms that belonged to the category ‘world knowledge’, i.e. the world the legal domain is about”. On the contrary, a core ontology should exclusively include “typical

* The Authors are Researchers at the Institute of Computational Linguistics “Antonio Zampolli” of the CNR - Pisa. The research reported in the paper has been partly supported by the Italian FIRB project “Piattaforma di servizi integrati per l’Accesso semantico e plurilingue ai contenuti culturali italiani nel web”. The Authors would like to thank Angela D’Angelo of the Scuola Superiore Sant’Anna of Pisa and Paolo Plini of the Institute of Atmospheric Pollution, Environmental Terminology Unit (CNR, Rome) who contributed as domain experts to the evaluation process.


XXXVI annata – Seconda Serie - Edizioni Scientifiche Italiane, Napoli
legal concepts, like norm, responsibility, person (agent), action, etc.”. Moreover, the most serious consequence envisaged is that “ontologies mixed with epistemological frameworks have a far more limited re-use and may pose more interoperability problems than clean ontologies.” In fact, the level of generality adopted in constructing a domain ontology is closely related to the reusability issue. According to the state of the art in ontology design criteria reported in Casellas\(^2\), several levels can be established ranging from the more abstract top or upper-level ontologies, which include general concepts not domain-specific, and core ontologies, which provide top-level domain-specific (i.e. legal) concepts, to domain-specific ontologies, which organize world knowledge, providing a description of a specific domain of interest to be regulated.

Building on these emergent issues, Francesconi\(^3\) has recently proposed an approach to legal knowledge modeling based on the separation of legal and world knowledge and oriented to interoperability and reusability. According to the knowledge model suggested, two levels of conceptualization are envisaged: a Domain Independent Legal Knowledge (DILK) level, which provides a model for legal rules independently from the domain they apply to, and a Domain Knowledge (DK) level, which offers information and relationships among entities specific for a given regulated domain. This approach follows Biagioli\(^4\), who claims that a law simultaneously describes the occurring events and regulates them.

In this paper, we face the epistemological promiscuity problem at the level of the acquisition of terminological knowledge from legal texts. Instead of starting from ready-made epistemological and ontological concepts, which are defined a priori on the basis of domain-theoretical assumptions, we propose a term extraction approach overtly aimed at automatically discriminating legal terms from regulated-domain terms.


\(^3\) E. Francesconi, Legal Rules Learning Based on a Semantic Model for Legislation, in “Proceedings of the Workshop on Semantic Processing of Legal Texts” (SPLeT-2010), held in conjunction with the 7th Conference on Language Resources & Evaluation (LREC 2010), La Valletta, Malta, 23rd May.