

Poetry Lab

An Open Source Toolkit for the Analysis of Spanish Poetry Corpora

Elena González-Blanco | Salvador Ros Muñoz | Javier de la Rosa | Álvaro Pérez Pozo | Laura Hernández | Mirella de Sisto | Aitor Díaz | Omar Khalil | José Luis Rodríguez | Leire Leguina
(POSTADATA PROJECT ERC, UNED, España)

MAIN GOAL

Tools for automatic poetry analysis, largely based on Natural Language Processing.
Carry out the detection of literary phenomena relied on linguistic characteristics

POSTDATA Project

- ERC-funded POSTDATA Project network of ontologies
- Interchange of data using Linked Open Data
- Varied research interests result in different corpora formats

PoetryLab leverages NLP to foster the completeness of the interchanged corpora

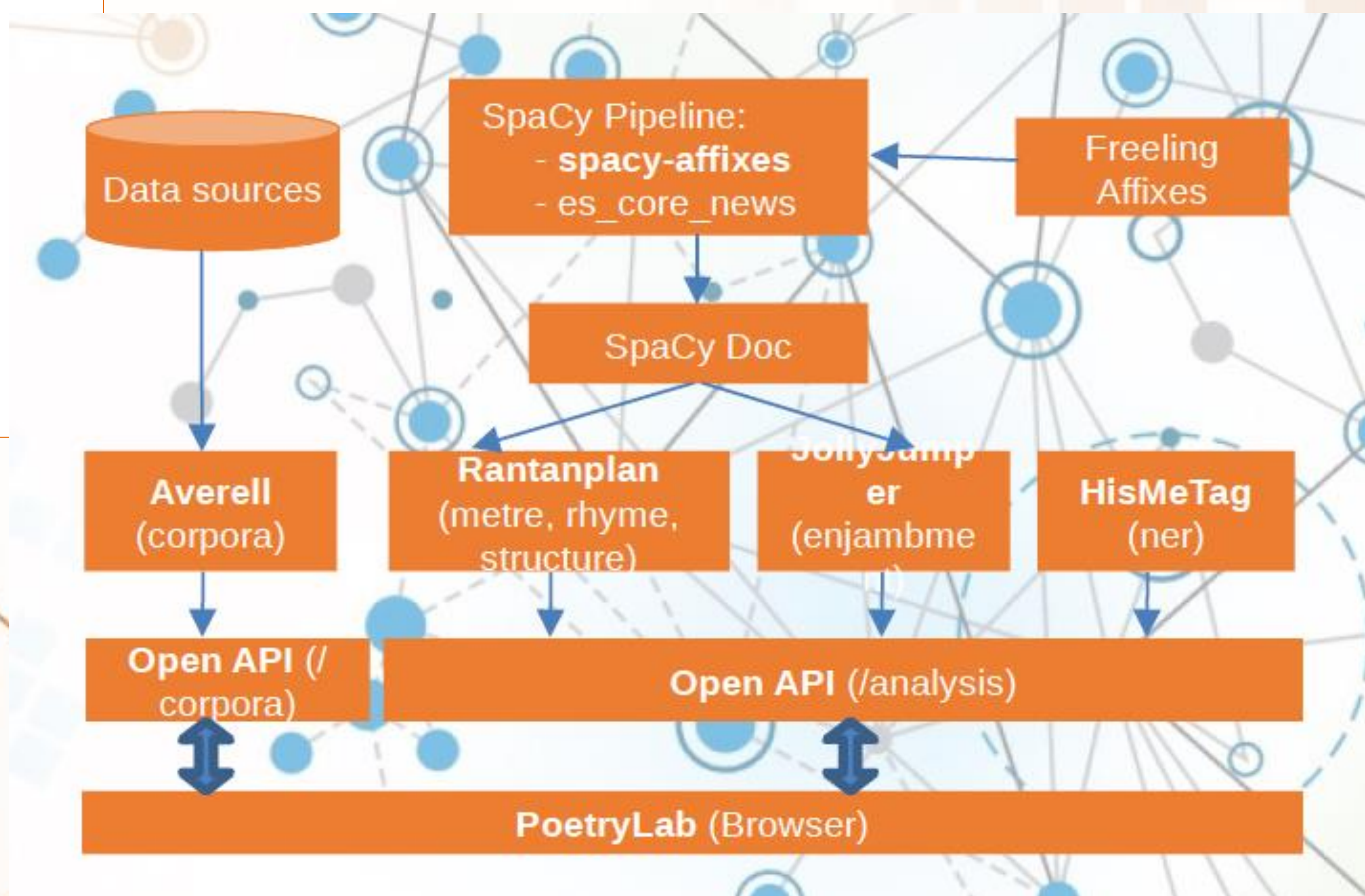
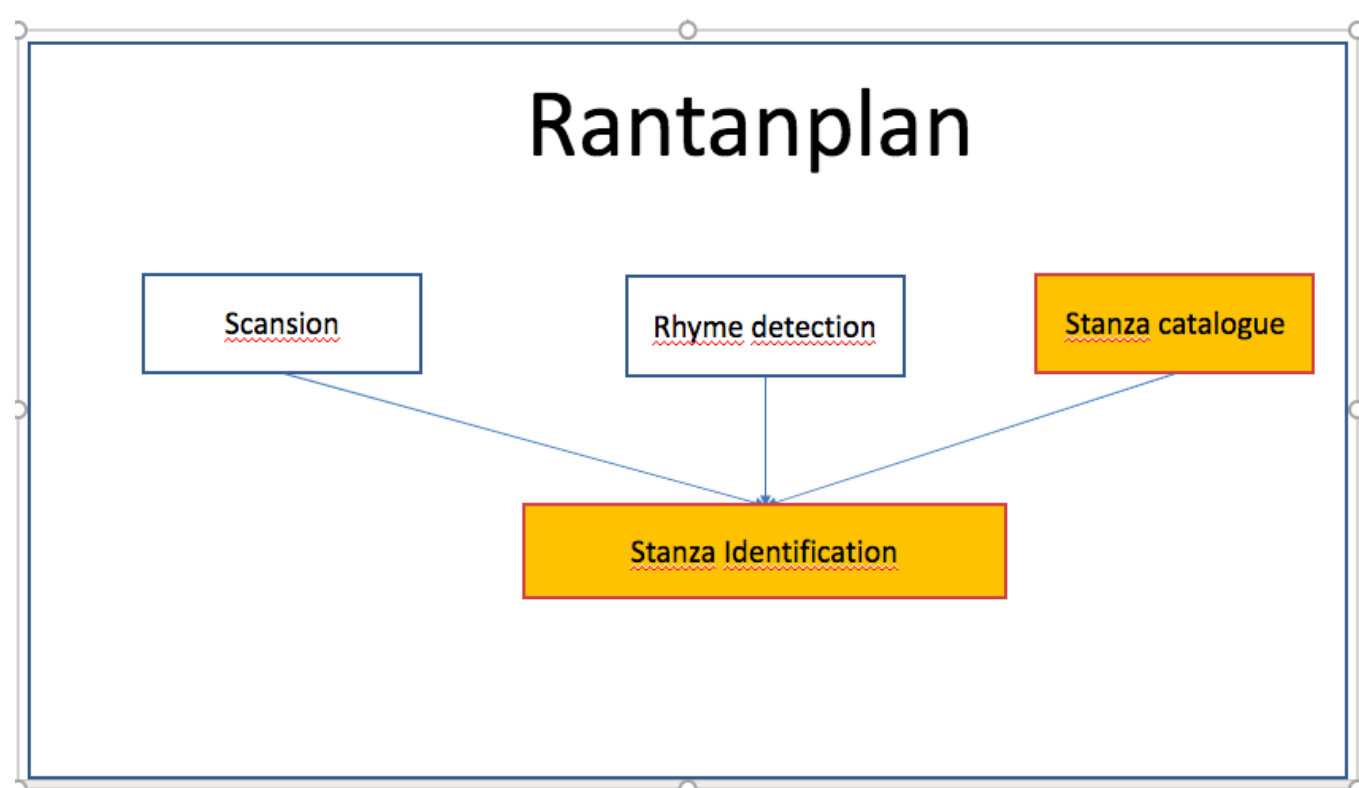
RANTANPLAN

Current implementation

- **Scansion:**
 - Rhythm pattern: '+ - - + - - + -'
 - Length: 8
- **Rhyme detection:**
 - Rhyme pattern: 'abbabccb'
 - Rhyme type: 'consonant'

Stanza catalogue:

CONSONANT_RHYME, ← Rhyme type
"quinteto", ← Name of the stanza
r"(ababa|abaab|abbab|aabab|aabba)", ← Regular expression for the rhyme pattern
lambda lengths: all(14 > length > 9 for length in lengths) ← Lengths condition

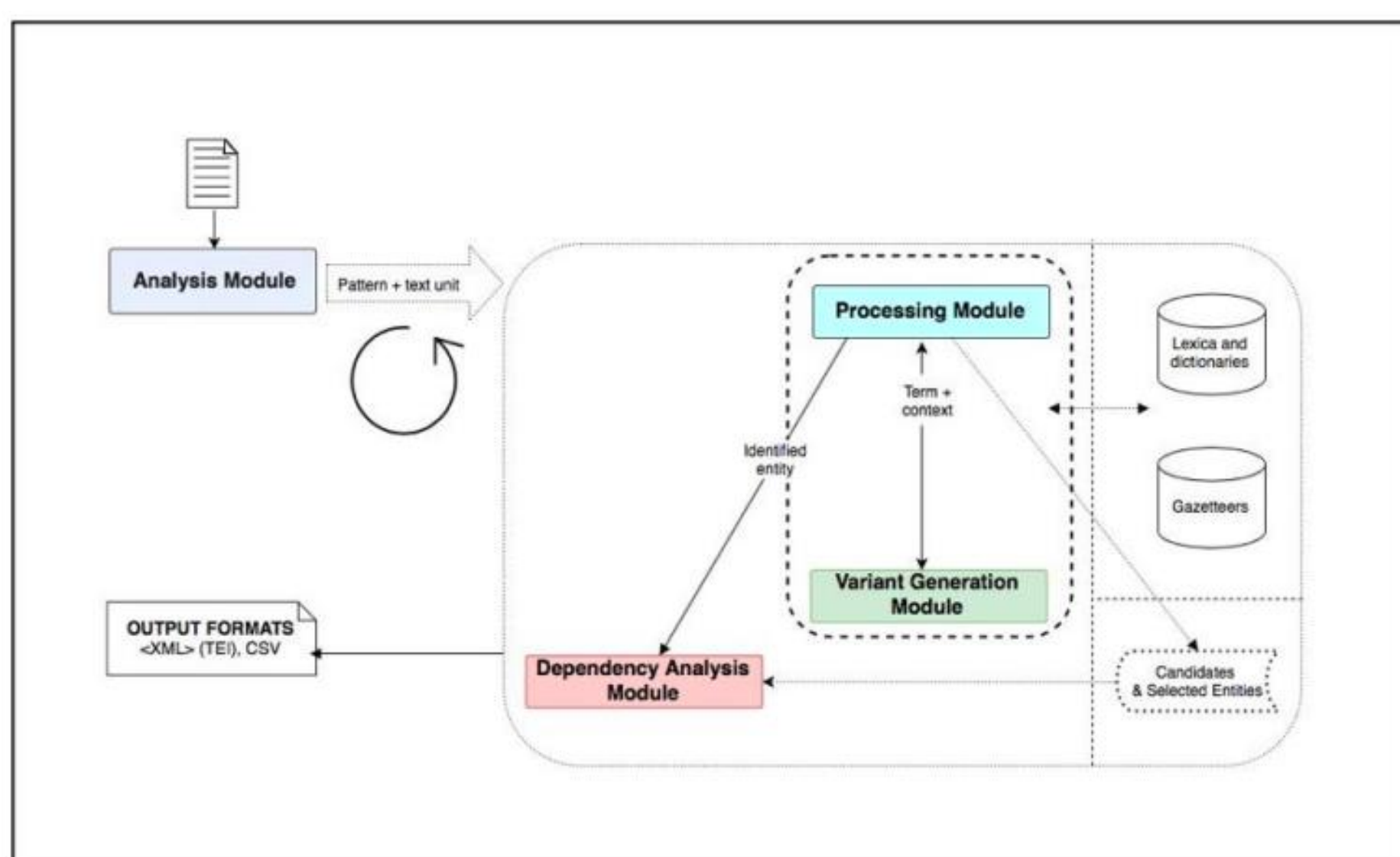


Benchmarking

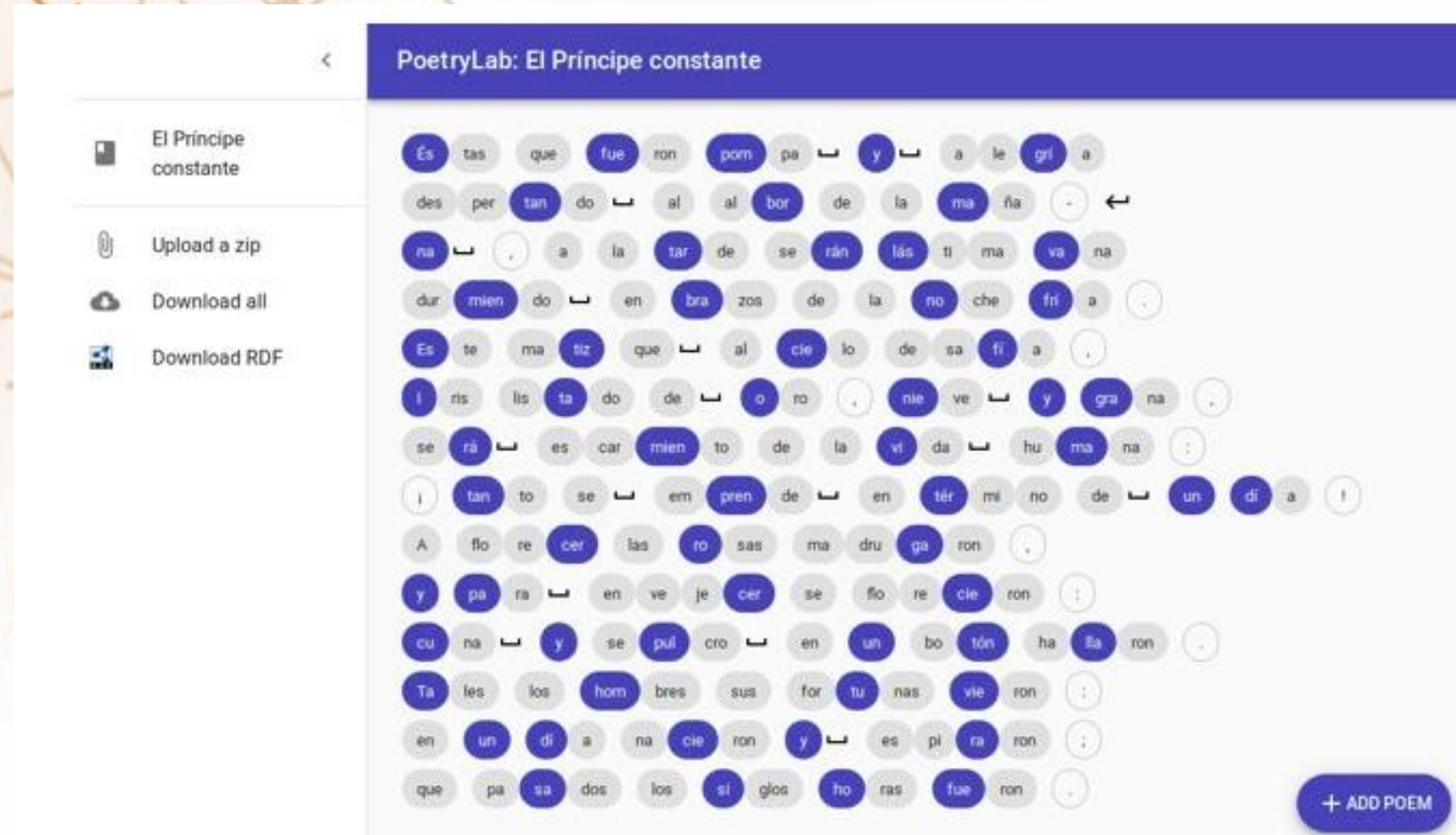
Task	Tool	Success (%)
Syllabification (EDFU)	PoetryLab(rantanplan)	99.99
	Navarro-Colorado	98.74
Metrical patterns (Navarro-Colorado)	PoetryLab(rantanplan)	96.22
	Navarro-Colorado	94.87
	Gervás	88.73
	Agirrezabal	90.84

HISMETAG, Hispanic Medieval Tagging

- Recognition of named entities in texts written in medieval Spanish
- PLN techniques for linguistic processing
- Normalization of historical spelling and variants generation
- Tagging of names automatically detected and proposed names



Usage Example with the UI



<http://prototipo-postdata.linhd.es/poetry-lab.php>