



Hands-on: Terminology in the Semantic Web

Patricia Martín-Chozas and **Elena Montiel-Ponsoda**
Ontology Engineering Group
Universidad Politécnica de Madrid, Spain

SD-LLOD-22

✉ pmchozas@fi.upm.es

🐦 @pmchozas

📅 1-6-22

📍 Cercedilla

Outline

1. Terminology tools
2. Standardisation approaches
 - a. Semantic Web standards
3. Main terminological resources in the SW
4. TermitUp testing
5. Modelling with VocBench
6. Querying EuroVoc
7. Additional info

Tools

1. **Term Extraction:** [TermoStat Web](#),
[TermSuite](#), [TBXTools](#), [Text2TCS](#)
2. **Keyword Extraction:** [RAKE](#)
3. **Extraction+Corpus:** [SketchEngine](#)
4. **Publishing:** [Terme à LLOD](#)
5. **Generation+Publishing:** [TermitUp](#)

Widely used standardisation models

1. [TBX](#), [TBX2RDF](#)
2. [ISO](#)
3. [SKOS](#)
4. [Ontolex](#) and it's [extensions](#) such as:
 - a. [Vartrans](#)
 - b. [Termlex](#) proposal

Main terminological* resources in the SW

1. [EuroVoc](#)
2. [UNESCO Thesaurus](#)
3. [IATE](#)**
4. -nets: [ConceptNet](#), [BabelNet](#), [WordNet](#)***
5. [Wikidata](#)***

* We treat thesauri as terminological resources, since they contain terminological info

** Not complete, but a dump

*** Not terminological resources but useful

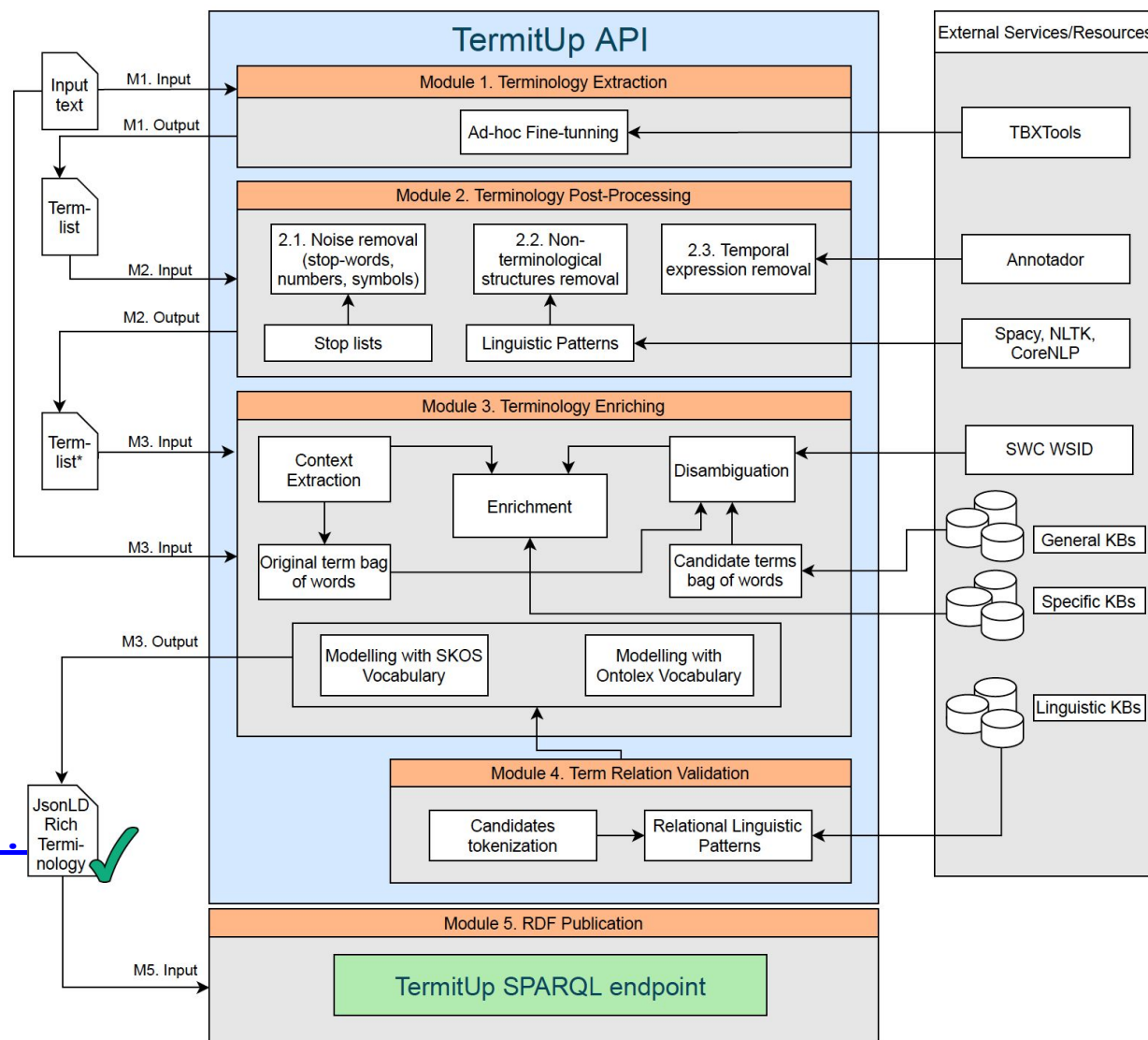
See [LLOD cloud](#) for more resources.

TermitUp

<https://termitup.oeg.fi.upm.es/>

Prêt-à-LLOD

<https://pret-a-llod.github.io/>



Martín-Chozas, P., Vázquez-Flores, K., Calleja, P., Montiel-Ponsoda, E., & Rodríguez-Doncel, V. *TermitUp: C and Enrichment of Linked Terminologies*. 2021. Special Issue on Latest Advancements in Linguistic Linked Semantic Web Journal (Forthcoming) <http://www.semantic-web-journal.net/system/files/swj2885.pdf>



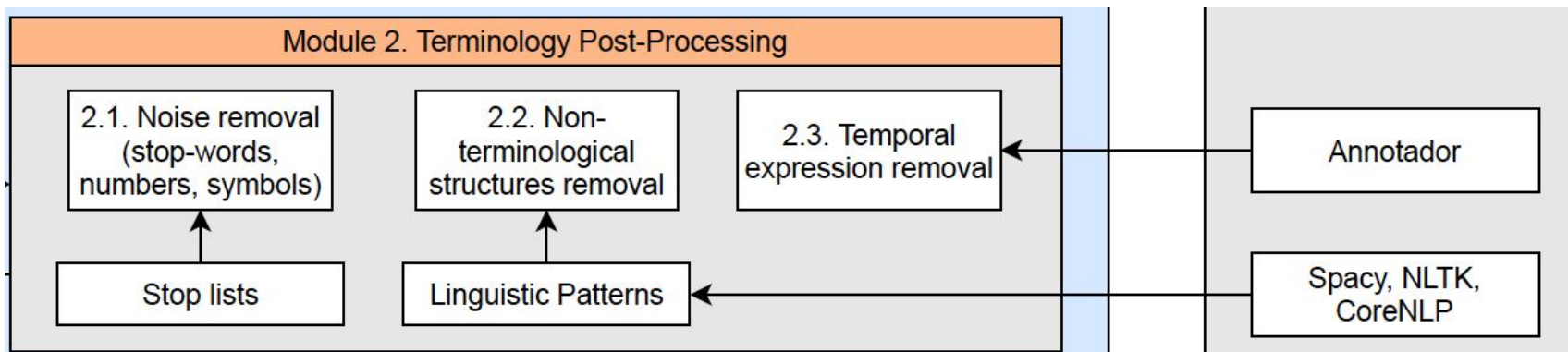
Module 1

- statistical terminology extraction
- English and Spanish
- TBXTools: open source python library
- **input:** corpus (plain text)
- **output:** terms (plain text)



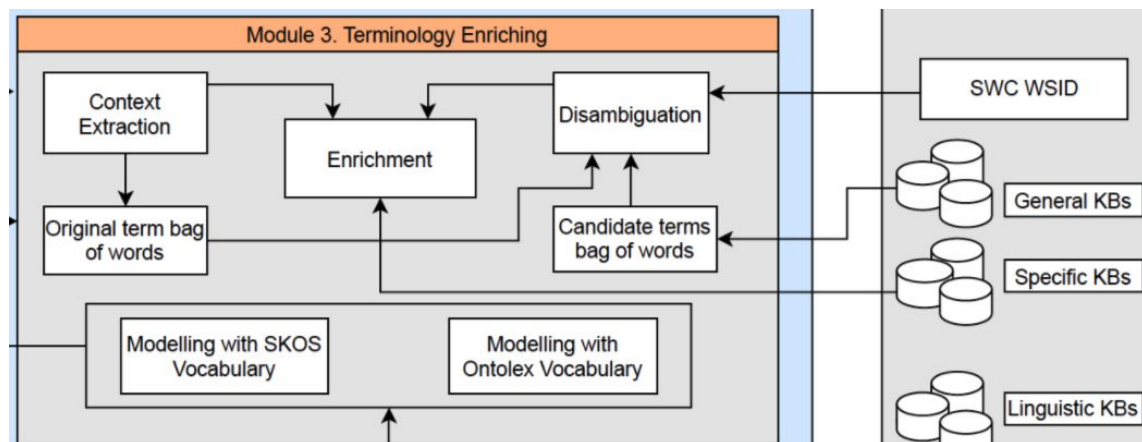
Module 2

- term post processing (cleaning)
- linguistic patterns
- English and Spanish
- external APIs ([Annotador](#)) and other Python libraries for NLP
- **input:** raw terms (plain text)
- **output:** clean terms (plain text)



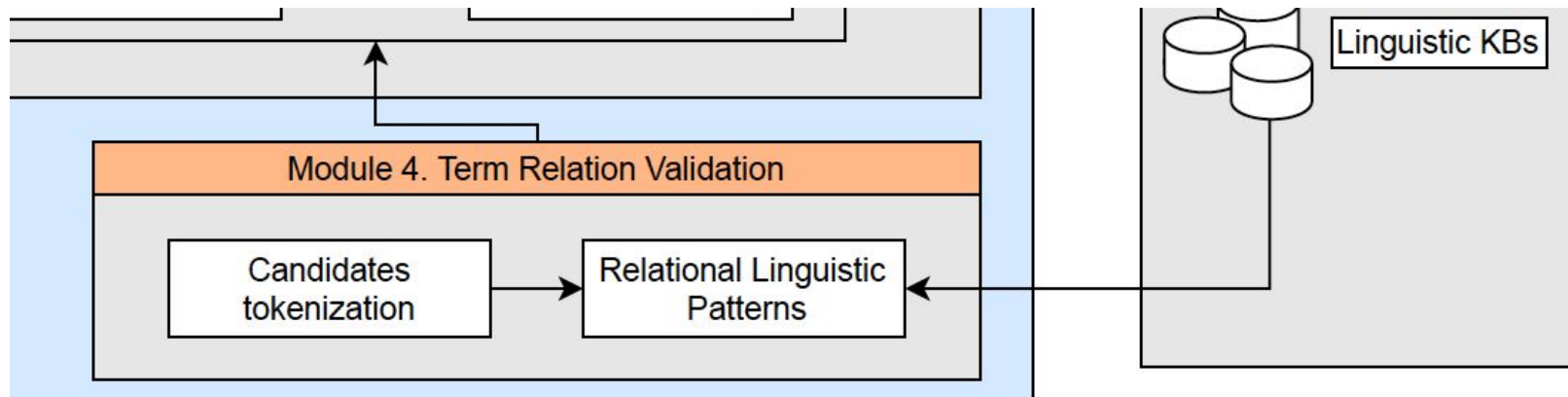
Module 3

- terminology enrichment
- disambiguated multilingual data from existing resources, some published in the LLOD ([WSD from SWC](#))
- translations, synonyms, usage examples, definitions, relations...
- **input:** clean terms (plain text)
- **output:** enriched terms (RDF: SKOS/Ontolex)



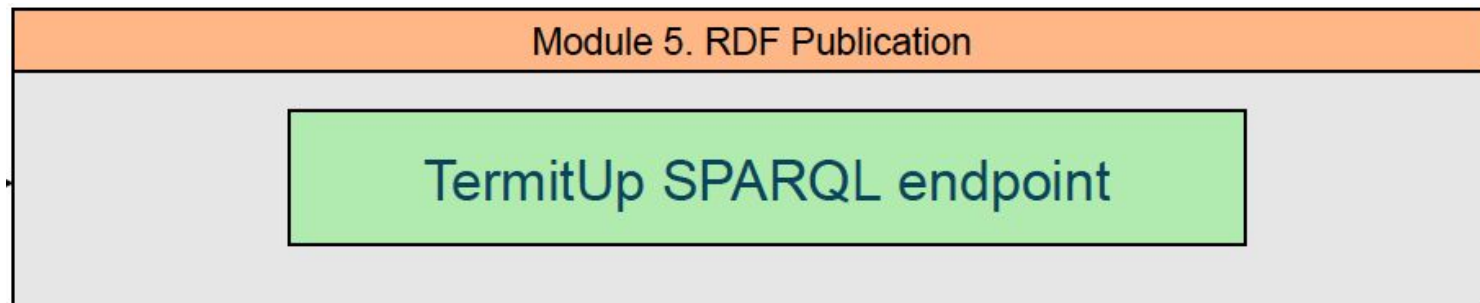
Module 4

- relation validation
- validates retrieved relations (hierarchical)
- using linguistic knowledge base + linguistic patterns
- **input:** terms (plain text)
- **output:** terms + relations (plain text)



Module 5

- publication (SKOS/Ontolex)
- SPARQL endpoint <https://termitup.oeg.fi.upm.es/sparql>
- freely queryable



<http://tiny.cc/965suz>

Terminology Modelling Tips

- 1. Study your data:**
 - a. type of data (content)**
 - b. domain**
 - c. purpose**
- 2. Study your model: specification**
- 3. Choose your model according to the type, domain and purpose.**

Useful tools for RDF

1. [JSON playground](#)
2. [RDF validator](#)
3. [LOV](#)
4. [prefix.cc](#)
5. [VocBench](#)
6. [Protégé](#)



Hands-on: Terminology in the Semantic Web

Patricia Martín-Chozas and **Elena Montiel-Ponsoda**
Ontology Engineering Group
Universidad Politécnica de Madrid, Spain

SD-LLOD-22

✉ pmchozas@fi.upm.es

🐦 @pmchozas

📅 1-6-22

📍 Cercedilla