

THE MULTITAL PLATFORM

Damien Nouvel, Driss Sadoun, Mathieu Valette

damien.nouvel@inalco.fr, driss.sadoun@postlab.fr, mvalette@inalco.fr

INALCO ERTIM

Languages at INALCO

INALCO (National Institute for Oriental Languages and Civilizations) is a university and research center that teaches around 100 oriental languages from Central Europe, Africa, Asia, America and Oceania.

The **ERTIM** lab works on some of those languages.

Under-resourced languages

In a **worldwide** perspective, currently:

- Around **150 written** languages
 - Approximately **6000 oral** languages
- The vast majority of those are under-resourced
- Who will provide resources? Not the GAFAs...

Goals of the MultiTAL platform <http://multital.inalco.fr>

Issues

- Relevant NLP tools are not so easy to find, in particular for under-resourced languages
- Documentation is not always comprehensive (sometimes native)
- Instructions to install, configure and execute NLP tools are not that simple

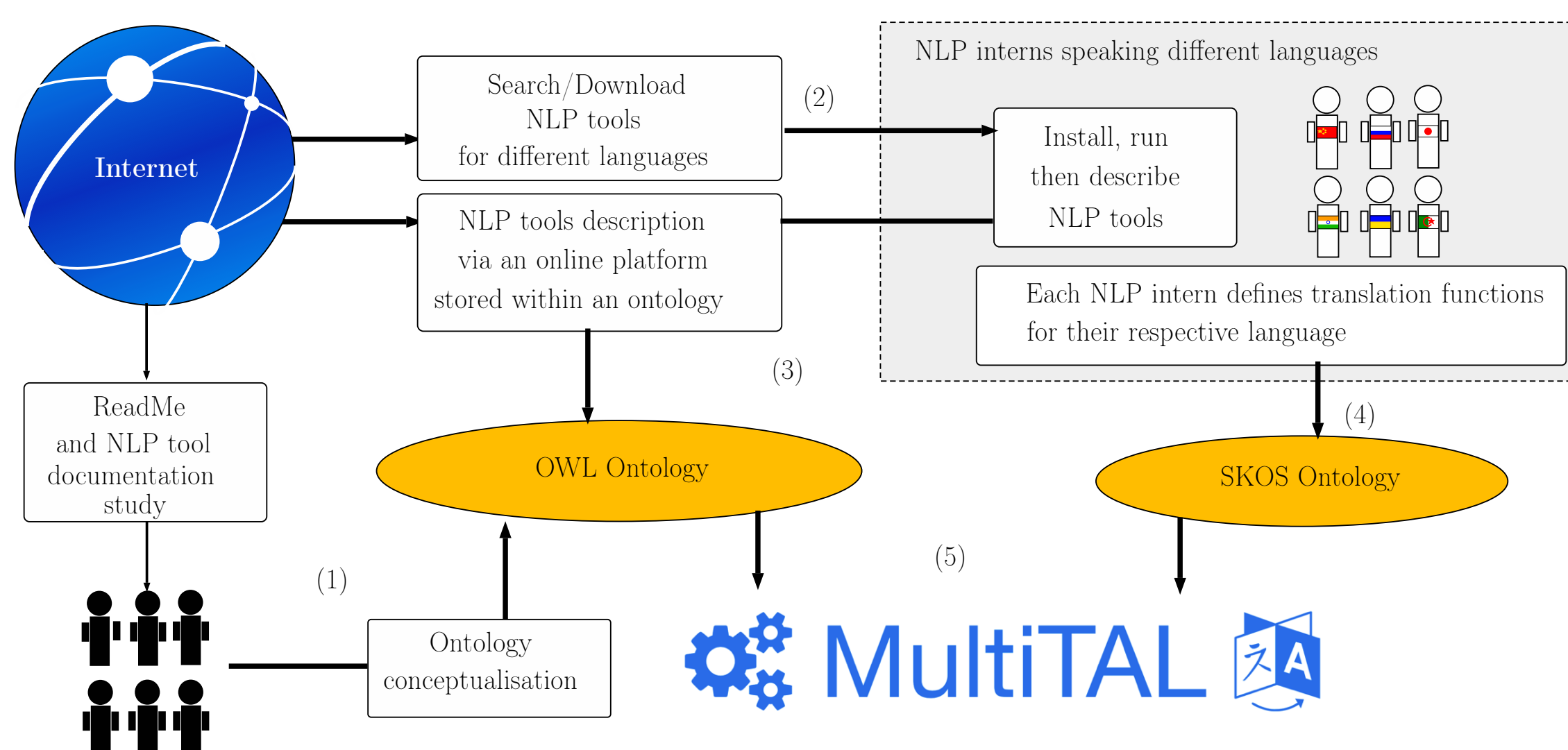
Goal: make technologies more accessible regardless of the expertise or spoken language of end-users

Mean: provide information with a simple, multilingual, structured and standardised tool documentation

Contributions of the MultiTAL platform

- **Multilingual ontology** of NLP tool documentation
- Web platform providing an easy access to **multilingual and structured documentation**
- Starting point to establish **guidelines and best practices** for NLP tool documentation

Platform architecture



Current status

- **45 NLP tasks**
- **47 languages**
- **167 tools** have been tested
 - 91 of them are published
 - 70 of them have failed tests
- **7 languages** for UI (Arabic, Chinese, English, French, Hindi, Japanese, Russian, Tibetan)
- **33 NLP methods**
- **112 tool authors**

Acknowledgements