Cloud-Based Infrastructure
of Terminology Services and Resources

Tatiana Gornostay, Andrejs Vasiļjevs, Roberts Rozis

Tilde, Riga, Latvia

{tatiana.gornostay, andrejs, roberts.rozis}@tilde.lv

Abstract. This short paper presents the concept of the cloud-based infrastructure of terminology services and resources. The first section addresses the needs of an instant access to the most up-to-date terms, user participation in the acquisition, processing, and sharing of multilingual terminological data, and efficient solutions for the reusing of terminology resources based on the cloud-based terminology services for human and machines as users. The second section focuses on the task of the consolidation and harmonisation of different, usually dispersed, terminology resources within the open infrastructure of language resources. This paper also includes the poster to be presented at the EAFT Terminology Summit 2013.[1]

Keywords: terminology, service, resource, cloud-based, infrastructure, acquisition, processing, sharing, reusing, consolidation, harmonisation

1   Introduction

Terminology is everywhere: we come across with terms every day, for example, when we are visiting a doctor, building a house, buying a car, etc. Terminology is a spine of the professional communication and of a document within its life cycle, including the creation, publication, translation, localisation, and other document management processes. Furthermore, terminology is of vital importance for brand consistency and customer satisfaction within businesses.

Terminology is developing rapidly and every day the volume of terminology grows along with the explosion of information available on the web. However, current static models for the acquisition, processing, and sharing of terminological data cannot keep up with this increasing demand. Moreover, in the context of the multilingual Europe, the role of terminology is even more important than ever to insure that people communicate efficiently and precisely.

---

[1] See Figure 1 for the poster to be presented at the EAFT Terminology Summit 2013.

Thus, there are evident needs for an instant access to the most up-to-date terms, collaborative models for the acquisition, processing and sharing of multilingual terminology, and facilities for the reusing of terminology resources in various applications within different usage scenarios. However, the current static models for these core terminology tasks cannot keep up with the growing demand due to the information explosion, terminology volume growth, etc.

2 Infrastructure of Terminology Services

An ongoing TaaS[2] project has been initiated this year to address the abovementioned needs and establish a sustainable cloud-based platform that provides the following online core terminology services for key terminology tasks:

 ➢ Automatic extraction of monolingual term candidates using the state-of-the-art terminology extraction techniques from the documents uploaded by users;

 ➢ Automatic recognition of translation equivalents for the extracted terms in user-defined target language(s) from different public and industry terminology resources (for example, TAUS, IATE, EuroTermBank, and others);

 ➢ Automatic acquisition of translation equivalents for terms not found in existing terminology resources from parallel and/or comparable web data using the state-of-the-art terminology extraction and bilingual terminology alignment methods;

 ➢ Facilities for the platform users for cleaning up, or revising (i.e., editing, deleting), of automatically acquired terminology;

 ➢ Facilities for terminology sharing and reusing: APIs and export tools for sharing the resulting terminological data with major term banks and reuse in various applications within different usage scenarios.

The TaaS platform will demonstrate the efficacy of reusing acquired and user-cleaned terminology resources within the following usage scenarios:

 ➢ Simplify the process for language workers to prepare, store and share of task-specific multilingual term glossaries

 ➢ Provide instant access to term translation equivalents and translation candidates for professional translators through computer-assisted translation (CAT) tools

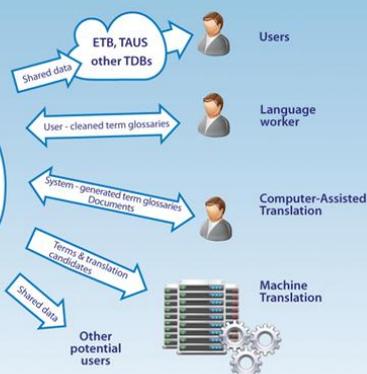 ➢ Domain adaptation of statistical machine translation (SMT) systems by dynamic integration with TaaS provided terminological data

3 Infrastructure of Terminology Resources

After a long history of terminology work that resulted in a large number of terminology resources, the task of the consolidation and harmonisation of different, usually dispersed, terminology resources becomes more urgent. An ongoing META-NORD[3] project addresses this demand. Within the project task of consolidating European multilingual terminology across languages and domains, the project aims at:

> ➢ Extending the META-SHARE Open Linguistic Infrastructure with monolingual, bilingual, and multilingual terminology resources across Europe;

> ➢ Integrating the EuroTermBank[4] platform into META-SHARE by adapting EuroTermBank to relevant data access and sharing mechanisms;

> ➢ Populating EuroTermBank with additional terminology resources and thus broadening the language domain coverage of EuroTermBank.

META-SHARE will contain among others:

> ➢ EuroTermBank as a local terminology repository consisting of its own terminology resources and metadata that will follow the META-SHARE schema;

> ➢ A local inventory consisting of metadata for the terminology resources stored at the local repository;

> ➢ A META-SHARE inventory consisting of the metadata for the terminology resources stored in the repository.[5]

4 Conclusions

The ongoing work lays the ground for the fruitful cooperation in the acquisition, processing, sharing, and consolidating of terminology resources. During last year more than 10 terminology resources were identified to be interlinked with the META-SHARE Open Linguistic Infrastructure via the EuroTermBank portal. The work on the integration of EuroTermBank into META-SHARE has started in the beginning of this year and the number of interlinked terminology resources is increasing.

---

[3] META-NORD: Baltic and Nordic Parts of the European Open Linguistic Infrastructure (www.meta-nord.eu), the work within the project leading to these results has received funding from the European Union under grant agreement n° 270899.

[4] EuroTermBank (www.eurotermbank.com) is one of the major efforts in the consolidation of terminology resources. It is a centralised online publicly available term bank for the languages of the European Union. EuroTermBank provides a federated access to other interlinked external term banks.

[5] See more about the overall architecture of META-SHARE at: www.meta-net.eu/meta-share/architecture.

Figure 1. Cloud-Based Infrastructure of Terminology Services and Resources