

Consolidating European Multilingual Terminology across Languages and Domains

Tatiana Gornostay, Roberts Rozis, Andrejs Vasiļjevs, Inguna Skadiņa

Tilde, Riga, Latvia

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

Abstract. This short paper addresses the problem of consolidating European multilingual terminology resources across languages and domains. In the introduction section we identify the task and the necessity of a consolidated interface for different, usually dispersed, multilingual terminology resources. In the second section we give examples of state-of-the-art approaches to the solution of this task – the Quest tool and the EuroTermBank portal. The third section presents a brief overview and reports on midterm results (achieved during the first year of the project) of an on-going research on consolidating European multilingual terminology resources as part of the ICT PSP EU project META-NORD within the META-NET initiative and the META-SHARE open linguistic infrastructure. Finally, we make conclusions and outline future work.

Keywords: language resource, terminology resource, multilingualism, consolidation, linguistic infrastructure

1 Introduction

Terminology is multidisciplinary and comprises primarily such tasks as the analysis of concepts and conceptual systems; creation of new terms; identification, recognition, extraction of existing terms; compilation of terminology resources, for example, dictionaries, banks, databases, i.e., terminography; application of terminology resources, for example, in translation, including computer-assisted and machine translation; management of terms.

After a long history of terminology work and as the number of terminology resources grow every year, the task of the consolidation of different, usually dispersed, terminology resources becomes more urgent.¹ According to recent research surveys on user practice in terminology work, one of the most required functionality of a terminology resource is a consolidated interface [2, 4].

¹ See, for example, a PhD thesis on consolidation of heterogeneous terminology resources [8].

2 Dispersed Terminology and Its Consolidation

There have been several efforts to provide reasonable solutions in accessing multilingual terminology resources for language workers. For example, the *Quest* tool, a one-stop access to a series of general-interest terminology databases, brings consolidated terminology content to translators in the Directorate-General for Translation of the European Commission [6]. Quest is not a terminology database but a metasearch interface which translators can use to query several databases simultaneously [Ibid.: 9].

One of the major efforts in the consolidation of terminology resources is EuroTermBank² – a centralized online publicly available term bank for the EU languages which provides a federated access to 5 interlinked external term banks [1, 7, 10] (Fig. 1).

Moreover, a free innovative multilingual terminology translation tool *EuroTerm-Bank Terminology Add-in*³ was developed. The tool integrates terminology resources from EuroTermBank into the most widely exploited working environment among language workers – *Microsoft Word 2003/2007* [3].

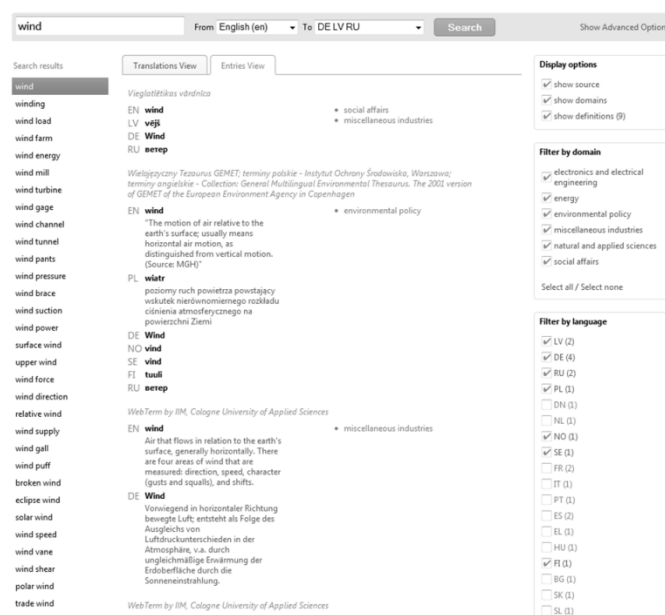


Fig. 1. Consolidated representation of terminology entries in EuroTermBank from different bilingual and multilingual resources

² www.eurotermbank.eu

³ The tool can be downloaded under the following link:
<http://www.eurotermbank.com/downloads.aspx>

3 Consolidation of Terminology in META-NORD

The ICT PSP project META-NORD⁴ contributes to building an open linguistic infrastructure for language resources by identifying and describing language resources in the Baltic and Nordic countries [5, 11] and by populating language resources (after IRP issues are cleared) and their metadata into the open distributed META-SHARE platform.⁵ The first batch of META-NORD language resources was released in November 2011, two more batches are planned in July, 2012 and January, 2012.

META-NORD addresses a growing demand for consolidating dispersed terminology resources. Within the task of consolidating European multilingual terminology across languages and domains, META-NORD aims at:

- extending META-SHARE with monolingual and bilingual multilingual terminology resources across Europe;
- integrating the EuroTermBank 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 coverage of EuroTermBank [9].

EuroTermBank will be integrated into the META-SHARE platform as a distributed terminology repository (Fig. 2).

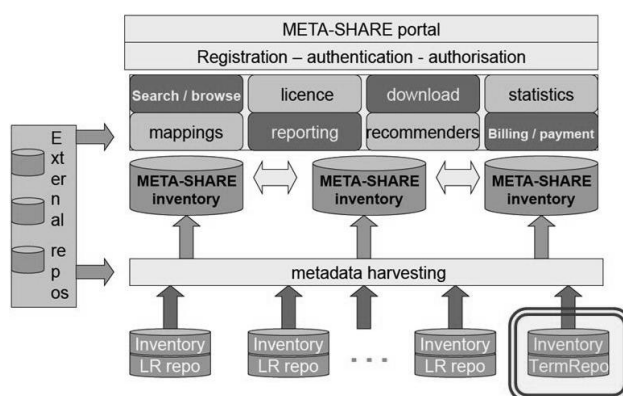


Fig. 2. Terminology repository within the META-SHARE network

META-SHARE will contain among others: (1) EuroTermBank as a local terminology repository consisting of its own terminology resources and metadata that will follow the META-SHARE schema; (2) a local inventory consisting of metadata for the terminology resources stored at the local repository; (2) a META-SHARE inventory consisting of the metadata for the terminology resources stored in the repository.⁶

⁴ www.meta-nord.eu

⁵ www.meta-share.eu

⁶ See more about the overall architecture of META-SHARE at: www.meta-net.eu/meta-share/architecture.

4 Conclusions and Future Work

META-NORD lays the ground for fruitful cooperation in identifying, consolidating, and sharing terminology resources across Europe. During the first year of the project 10 terminology resources have been identified to be interlinked with the META-SHARE portal via EuroTermBank. The work on integrating EuroTermBank into META-SHARE has started recently and the number of interlinked terminology resources can be increased during the second year of the project. It is also anticipated that the META-NORD initiative can be further extended to other European countries by other projects within the META-NET network – CESAR, METANET4U, and T4ME.

References

1. Auksoiriūtė A., Belogrīvs I., Bielevičienė A. [et al.]. Towards Consolidation of European Terminology Resources: Experience and Recommendations from the EuroTermBank Project. Signe Rirdance and Andrejs Vasiljevs (eds.). Tilde, Riga, Latvia, 2006.
2. Blancafort, H. and Gornostay, T. Calling Professionals: Help Us to Understand Your Needs! A questionnaire-based online survey about terminology and corpora practices. Electronic resource: http://www.ttc-project.eu/images/stories/TTC_Survey_2010.pdf.
3. Gornostay T., Vasiljevs A., Rirdance S., and Rozis R. Bridging the Gap – EuroTermBank Terminology Delivered to Users' Environment. In Proceedings of the 14th Annual Conference of the European Association for Machine Translation, Saint-Raphaël, France, 2010.
4. Gornostay, T. Terminology management in real use. In: Proceedings of the 5th International Conference “Applied Linguistics in Science and Education”, Saint-Petersburg, Russia, 2010.
5. Skadiņa, I., Vasiljevs, A., Borin, L., Smedt, De K., Lindén, K., Rögnavaldsson, E. META-NORD: Towards Sharing of Language Resources in Nordic and Baltic Countries. In Proceedings of the workshop IJCNLP 2011 on Language Resources, Technology and Services in the Sharing Paradigm, Chiang Mai, Thailand, 2011.
6. Translation tools and workflow: European Commission Directorate-General for Translation. Luxembourg, Office for Official Publications of the European Communities, 2007.
7. Vasiljevs A. and Rirdance S. Consolidation and unification of dispersed multilingual terminology data. In Proceedings of the International Conference “Recent Advances in Natural Language Processing”, Borovets, Bulgaria, 2007.
8. Vasiljevs, A. Consolidation of Heterogeneous Terminology Resources. PhD thesis, University of Latvia, Riga, Latvia, 2011.
9. Vasiljevs, A., Forsberg, M., Gornostay, T. [et al.]. Creation of an Open Shared Language Resources Repository in the Baltic and Nordic Countries. In Proceedings of the 8th International Conference on Language Resources and Evaluation, Istanbul, Turkey, 2012.
10. Vasiljevs, A., Rirdance, S. and Liedskalnins, A. EuroTermBank: Towards Greater Interoperability of Dispersed Multilingual Terminology Data. In Proceedings of the 1st International Conference on Global Interoperability for Language Resources, Hong Kong, 2008.
11. Vasiljevs, A., Skadiņa, I., Sandford, B.P., Smedt, De K., Borin, L. META-NORD: Baltic and Nordic Branch of the European Open Linguistic Infrastructure. In Proceedings of the 18th Nordic Conference on Computational Linguistics, Riga, Latvia, 2011.