



Project acronym: EUROTERMBANK

Project title: Collection of Pan-European Terminology Resources through Cooperation of Terminology Institutions

Contract number: EDC - 22 267 - EUROTERMBANK

Starting date: 01/01/2005 **Ending date:** 31/12/2006

D4.2 PREPARED CONTENT FOR DATABASE POPULATION

Version: *V 1.0*
Date: *31/08/2006*

Dissemination: *Public*

Authors and Contributors

Albina Auksoriūtė, Institute of Lithuanian Language
Audronė Bielevičienė, Institute of Lithuanian Language
Ainars Blaudums, Tilde
Tomasz Borkowski, Translegis Publishing
Eduards Cauna, Latvian Academy of Sciences
Anita Draveniece, Latvian Academy of Sciences
Asta Mitkeviciene, Institute of Lithuanian Language
Danuta Kierzkowska, Information Processing Centre
Balázs Kis, MorphoLogic
Uldis Priede, Tilde
Gábor Prószéky, MorphoLogic
Ivars Puksts, Latvian Academy of Sciences
Krystyna Siwek, Information Processing Centre
Robertas Stunzinas, Institute of Lithuanian Language
György Tardy, MorphoLogic
Andrejs Vasiljevs, Tilde

The document was compiled by Latvian Academy of Sciences and edited by Tilde

Contents

| | |
|---|---|
| Authors and Contributors | 2 |
| Contents | 3 |
| 1. Described and delivered resources..... | 4 |
| 1.1. General information | 4 |
| 1.2. Paper and electronic resources..... | 4 |
| 1.3. External databases..... | 5 |
| 1.4. Copyright status | 5 |
| 1.5. Distribution of terms by domain..... | 5 |
| 1.6. Further plans of description and collection of resources | 7 |
| 2. ETB database..... | 7 |

1. Described and delivered resources

1.1. General information

Until August 31, 2006 there are 101 delivered resources (excluding external databases) comprising 671,640 new terms (this number is changing as new collections are added to the database).

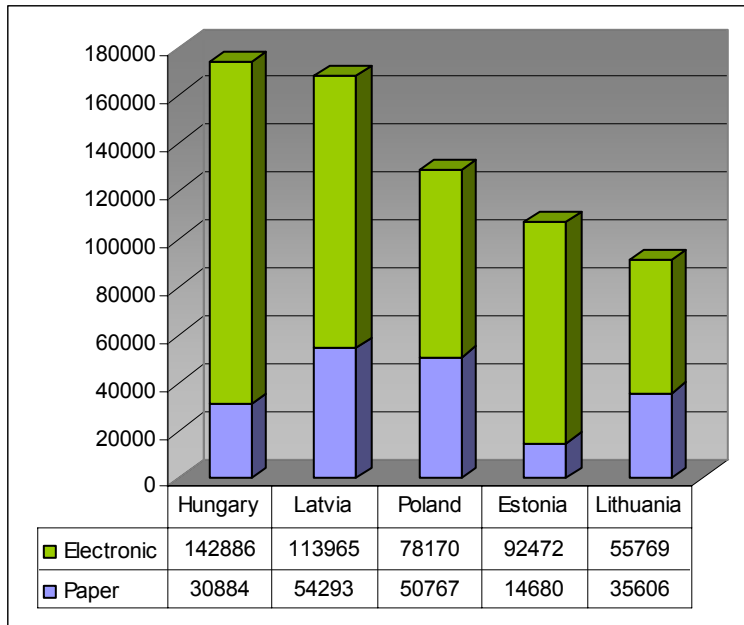


Figure 1. Number of resources submitted by country

Terms are counted as number of terms delivered in each respective partner language.

1.2. Paper and electronic resources

There are 481,410 terms from electronic resources (text files, XML, simple databases etc.) and 190,230 digitized terms obtained scanning paper media (mostly books).

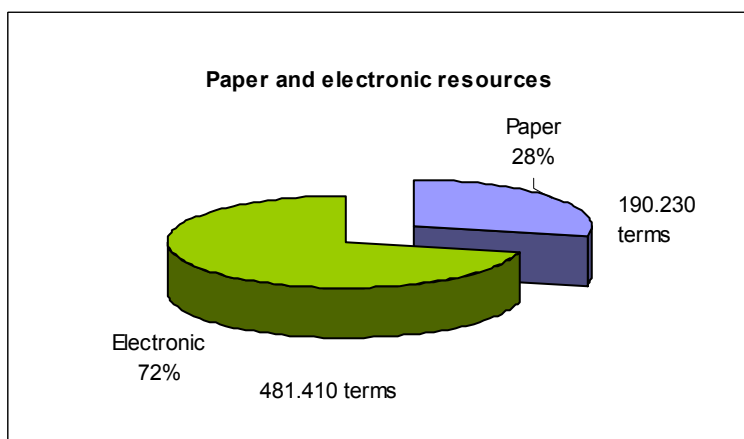


Figure 2. Number of electronic and paper resources

1.3. External databases

ETB partners have obtained consent to use two external databases from their holders:

- LAS terminology database (LV-87) — approx. 145,000 terms
- OSTEN – Open Dictionary of Scientific Terminology (PL-2) — approx. 2000 terms
- GNU Dictionary (PL-57) — approx. 27,000 terms.

1.4. Copyright status

Only resources with resolved copyright status are included in the ETB database. According to ETB guidelines, explicit written consent of copyright holder should be obtained. Copyright papers are submitted for 92 resources and they are expected soon for 4 more resources (translation of agreement or a photocopy of signed consent is expected).

7029 terms are ISO terminology — ETB Consortium has received consent to use these data in e-mail form.

Deliverable D4.1 Contracts with copyright owners provides a thorough overview of the copyright status of all delivered ETB resources.

1.5. Distribution of terms by domain

Most terms cover European Union terminology, computers and IT, technology (incl. building and machinery), economics, law and energy.

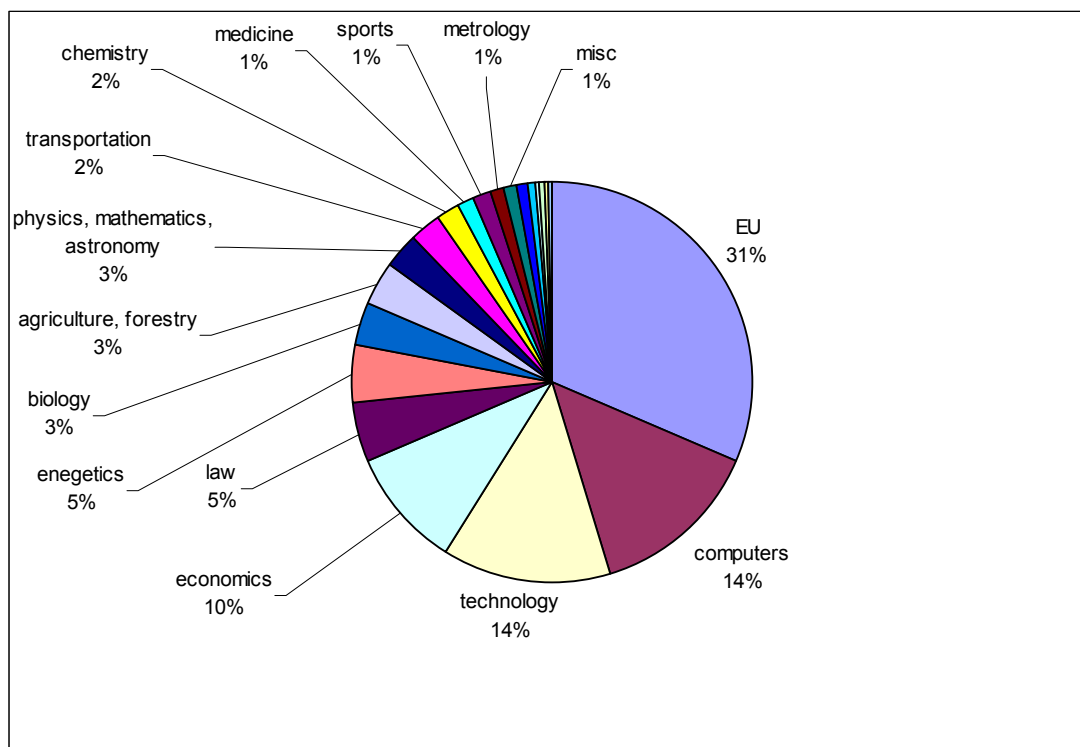


Figure 3. Number of submitted resources by domain

| Domain | Number of terms | % |
|---------------------------------|-----------------|-------|
| EU | 209 931 | 31,3% |
| computers | 94 241 | 14,0% |
| technology | 91 058 | 13,6% |
| economics | 66 105 | 9,8% |
| law | 31 320 | 4,7% |
| energy | 31 100 | 4,6% |
| biology | 23 278 | 3,5% |
| agriculture, forestry | 22 780 | 3,4% |
| physics, mathematics, astronomy | 20 353 | 3,0% |
| transportation | 15 682 | 2,3% |
| chemistry | 13 347 | 2,0% |
| medicine | 9 067 | 1,3% |
| sports | 8 770 | 1,3% |
| metrology | 7 790 | 1,2% |
| misc | 7 350 | 1,1% |
| ecology | 6 772 | 1,0% |
| art | 3 000 | 0,4% |
| linguistics | 2 684 | 0,4% |
| information processing | 2 593 | 0,4% |
| education | 2 000 | 0,3% |
| military | 1 558 | 0,2% |
| fire safety | 531 | 0,1% |
| quality | 330 | 0,0% |

Table 1. Number of submitted resources by domain

1.6. Further plans of description and collection of resources

During the project, the partners have collected information about 385 terminology resources (dictionaries, termlists, glossaries, thesauruses etc.). They were described according to Tedif standard and included in a common ETB resources information database. This work is still ongoing, and number of described resources increases every week. The resource descriptions are a valuable source of reference for the users of ETB portal.

Even though the goals of the EuroTermBank project in terms of collected resource volumes have been reached (e.g. digitization of 100,000 terms), the partners intend to continue working on resource collection and processing to increase the number of resources included in the ETB database thus making it more attractive for the potential users.

2. ETB database

The term collections currently in ETB database are accessible here:

URL: <http://db.eurotermbank.com>

User name: evaluator

Password: 112358