

**Kaaskiri riikliku programmi „Eesti keele keeletehnoloogiline tugi (2006–2010)“
sihtvalveerimise evalveerimisraportile**

Riikliku programmi „Eesti keele keeletehnoloogiline tugi (2006–2010)“ sihtvalveerimise evalveerimisraport valmis vastavalt Haridus- ja Teadusministeeriumi ning SA Archimedese vahelisele kokkuleppele 30.06.2011. Hiljem on kohendatud aruande sõnastust.

Pärast evalveerimisraporti valmimist ja edastamist Haridus- ja Teadusministeeriumile on valminud või saanud kättesaadavaks kaks olulist ülevaadet, milles on antud hinnangud keeletehnoloogia hetkeseisule ja arengule Eestis.

Nendeks ülevaadeteks on:

- Eesti keele arendamise strateegia 2004–2010 elluviimine.
- META-NET White Paper Series. Languages in the European Information Society: Estonian.

Neist „Eesti keele arendamise strateegia 2004–2010 elluviimine“ annab põhjaliku kvantitatiivse ülevaate programmiperioodil tehtust alavaldkondade ja -eesmärkide kaupa, META-NET'i ülevaate teemakäsitlus on oluliselt laiem ja hõlmab lisaks keeletehnoloogia vahendite ja ressursside kvantiteedile ka nende kvaliteeti, kättesaadavust, ühilduvust, jätkusuutlikkust jms (vt lühikokkuvõtted allpool).

Neid kahte ülevaadet ja neis sisalduvaid materjale ei ole käsitletud ülalmainitud evalveerimisraportis, kuid neis ülevaadetes toodud seisukohad ja antud hinnangud annavad olulist lisa- ja taustainfot ning täiendavad ja/või toetavad mitmeid raportis väljatoodud seisukohti. Seepärast pidasime vajalikuks viidata käesolevas kaaskirjas ka nendele materjalidele ja olulisema ka välja tuua.

23.09.2011

Viktor Muuli
Madis Saluveer

Eesti keele arendamise strateegia 2004–2010 elluviimine

Tegu on Eesti keelenõukogu poolt koostatud seireraportiga Eesti keele arendamise strateegia (2004-2010) elluviimise kohta, milles on antud ka hinnang eesti keele keeletehnoloogilise toe arendamisele, mille üheks osaks oli ka riiklik programm „Eesti keele keeletehnoloogiline tugi (2006–2010).

9. peatükk. Eesti keele keeletehnoloogiline tugi

Üldhinnang EKASi täitmisele:

Kõik seireperioodiks ettenähtud tööd on käivitatud ja on selles seisus, et on võimalik jätkata tempos ja ulatuses, mis vastab üldeesmärgile: tagada eesti keele edukas funktsioneerimine infoühiskonnas tänapäeval nõutaval tasemel.

Ettepanekud:

- järgneval perioodil jätkata olemasolevate prototüüpide täiustamist ning uute loomist, samuti jätkata andmebaaside ning tarkvara täiendamist ja arendamist;
- üheks olulisemaks ülesandeks on Eesti Keeleressursside Keskuse töö käivitamine ettenähtud mahus. Loodav keeleressursside keskus on infrastruktuur, mis teeb eestikeelsed keeleressursid ja keeletehnoloogilise tarkvara kõigile huvilistele kättesaadavaks. Lisaks käivitab see süsteemi olemasolevate keeleressursside tutvustamiseks, neist teavitamiseks ja potentsiaalsete kasutajate koolitamiseks.

Vt lisa:

<http://ekn.hm.ee/system/files/Eesti+keeles+arendamise+strateegia+elluviimine+2004-2010.pdf>

META-NET White Paper Series

Languages in the European Information Society: Estonian

This white paper is part of a series that promotes knowledge about language technology and its potential. It addresses educators, journalists, politicians, language communities and others.

META-NET, a European Commission Network of Excellence, has conducted an analysis of current language resources and technologies. This analysis focused on the 23 official European languages as well as other important national and regional languages in Europe. The results of this analysis suggest that there are many significant research gaps for each language. A more detailed, expert analysis and assessment of the current situation will help maximise the impact of additional research and minimize any risks.

Executive Summary

This whitepaper for the Estonian language demonstrates that a lively language technology industry and research environment exists in Estonia. Although a number of technologies and resources for Standard Estonian exist, there are fewer technologies and resources for the Estonian language than for the English language. The technologies and resources also have a poorer quality.

According to the assessment detailed in this report, immediate action must occur before any breakthroughs for the Estonian language can be achieved.

Conclusions

The situation of Estonian concerning language technology support gives rise to cautious optimism. Supported by larger research programs in the past, there exists a language technology research scene in Estonia. Unfortunately, the industry consists only of few SMEs. For Estonian, a number of technologies and resources exist, but far less than for English. Language technology support today is by far not in a state that is needed for offering the support a true multilingual knowledge society needs.

To sum up, the results indicate that Estonian stands reasonably well with respect to the most basic language technology tools and resources, such as tokenizers, PoS taggers, morphological analyzers, shallow parsers, reference corpora, treebanks and tools for speech technology. Furthermore, there exist some tools for summarization, machine translation, as well as resources like parallel corpora, and specialized corpora. However, these tools and resources are rather simple and have a limited functionality for some of the areas. For instance, parallel corpora only exist for very few language pairs and for limited text genres.

When it comes to more advanced fields like text semantics, language generation, and annotated multimodal data, Estonian clearly lacks basic tools and resources even if some of these are currently under development.

The research regarding the most advanced tools and resources like discourse processing, dialogue management, semantics and discourse corpora has yielded first results but the resources need still to complement in amount and quality and tools have a quite limited scope.

Most of these tools (except morphological analyzer) have been developed by research institutions. However, most of these tools can be regarded as prototypes, not as mature products. The development was supported by different national programmes of language technology what guarantees the availability of these products.

Vt lisa:

<http://www.meta-net.eu/whitepapers/download/meta-net-languagewhitepaper-estonian.pdf>