



Contribution ID: 29

Type: Demo

Using the BalticGrid-II infrastructure: SemtiKamols - a linguistic analyser of Latvian language

Tuesday, 23 September 2008 16:04 (0 minutes)

Describe the activity, tool or service using or enhancing the EGEE infrastructure or results. A high-level description is needed here (Neither a detailed specialist report nor a list of references is required).

We would like to present SemtiKamols - a linguistic analyzer used for syntactic and morphologic analysis for Latvian language. SemtiKamols is an one of applications, developed and used by Baltic States research communities that benefit from using the BalticGrid Second Phase EU Project (BalticGrid-II) infrastructure, fully interoperable with the pan-European e-Infrastructures established by EGEE. Application is run within Migrating Desktop -intuitive uniform interface to Grid resources.

Report on the impact of the activity, tool or service. This should include a description of how grid technology enabled or enhanced the result, or how you have enabled or enhanced the infrastructure for other users.

SemtiKamols is a linguistic analyser of Latvian written texts, being the part of University of Latvia project that aims to develop semantic resources and methodologies for automatic meaning extraction from Latvian texts. This ultimate goal requires the lower levels of the language analysis, namely, morphology, syntax and lexical semantics, to be properly implemented. Besides basic grammar parsing for Latvian, the project utilizes techniques from Semantic Web and related formal logic-based approaches to enable formal reasoning about the facts mentioned in the text.

The algorithm is of a high-complexity and requires exponential CPU time. For example, analysys of all digitally available Latvian language texts (about 4GB of text) requires thousands of CPU hours (about 2-3 CPU years). Using Grid infrastructure makes possible analyzing of large amounts of data - each sentence is analysed independently, so 'gridification' of this problem is quite easy and gives obvious benefits.

Abstracts for online demonstrations must provide a summary of the demo content. Places for demos are limited and this summary will be used as part of the selection procedure. Please include the visual impact of the demo and highlight any specific requirements (e.g. network connection). In general, a successful demo is expected to have some supporting material (poster) and be capable of running on a single screen or projector.

We would like to show a demo of the SemtiKamols application, starting from defining of input files and computing parameters, submission of the job to the BalticGrid-II infrastructure, monitoring the job state, and ending with visualization of results. We hope to attract the audience's attention by the interesting scientific content of our demo combined with intuitive interface. The application will be run within the Migrating Desktop Platform -intuitive graphical user interface that hides the complexity of the grid middleware and

makes access to the BalticGrid resources easy and transparent. Some supporting materials: as posters, as well as several types of brochures will be provided. A single screen and a network connection is needed to present the demo on-line.

Describe the added value of the grid for your activity, or the value your tool or service adds for other grid users. This should include the scale of the activity and of the potential user community, and the relevance for other scientific or business applications.

The Baltic Grid Second Phase project is a continuation of the BalticGrid, designed to increase the impact and to further extension of pan-European e Infrastructure. This will be achieved by an extension of the BalticGrid infrastructure (fully interoperable with EGEE resources and complementary with the EGEE infrastructure) to Belarus; identifying and addressing the specific needs of new scientific communities such as nano-science and engineering sciences; and by establishing new Grid services.

The usage of grid infrastructure of the project has a significant impact on scientific communities, validated through everyday work of the scientists studying the Baltic Sea environment, bioinformatics, high energy physic, astrophysics and other sciences.

The example of application that benefit from Grid infrastructure is SentiKamols, which thanks to usage of project resources is able e.g. to analyse all digitally available Latvian language texts.

Primary authors: Mr PALAK, Bartosz (Poznan Supercomputing and Networking Center); Mr PODINS, Karlis (University of Latvia)

Presenters: Mr PALAK, Bartosz (Poznan Supercomputing and Networking Center); Mr PODINS, Karlis (University of Latvia)

Session Classification: Demos and Posters

Track Classification: Demo