knowARC –
applications for industry, medical organizations, and bioinformatics research

EGEE conference
Geneva, September 26, 2006

Henning Müller

Overview

- My background
- Problems with grids for a non expert user
- Needs for users in various domains
  - Industry
  - Bioinformatics research
  - Medical/large administrations
- KnowARC goals
- Conclusions

My background

- Work at the interface of hospitals and University
  - Research in image processing/image analysis/image information retrieval
  - No dedicated research computing infrastructure (only desktops)
- BUT: Large infrastructure is available
  - ~6000 computers in the hospitals
  - How to get access to this infrastructure

Our problems with grid use

- For a non expert user!
- Maintaining an infrastructure is expensive
  - Computers and software maintenance
- Adapting/griddifying software is expensive
  - Particularly when pre-defined applications are used (Insight toolkit, Matlab, …)
- Often grid middleware is invasive
  - Impossible to reuse an existing infrastructure
- System administrators are not always aware of possibilities and problems

Application selection for KnowARC

- **Application areas with varying needs**
  - Less-traditional application areas
- **Applications in industry**
  - Science and computing
- **Application in basic research**
  - Bioinformatics University of Lübeck, Germany
- **Application in large organizations**
  - University and hospitals of Geneva

Application industry

- Science + computing
- Service provider for industry
  - Automotive, …
- Production-quality middleware
  - Stable, easy to use, graphical interfaces, maintainable without experts

Application basic research

- Bioinformatics University of Lübeck
- Often, a well defined workflow exist
  - Often not in typical batch processing
- Repetitive tasks can be improved strongly with computing infrastructure
  - Often some infrastructures are available but not shared
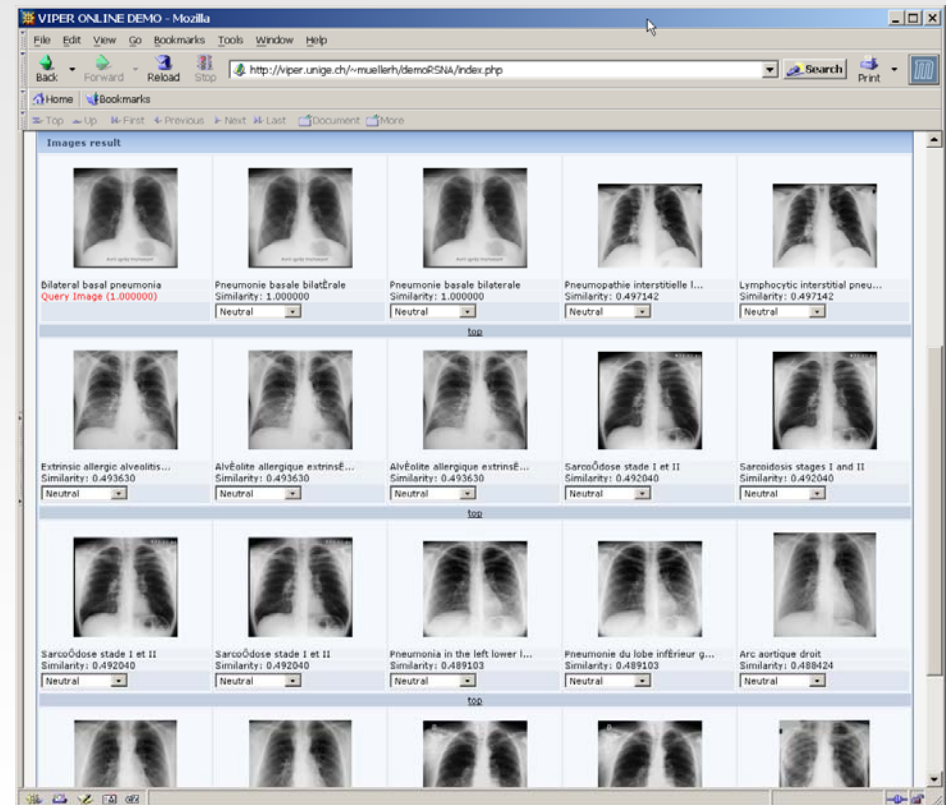- Treatment of larger datasets improves knowledge discovery

Application in large organisations

- Many administrations have similar characteristics
  - Very heterogeneous set of computers is available (6000 in Geneva, 1-5 years old)
  - Independent entities
  - Research is not the priority
    - But could profit from resources
  - Security is main motivation
  - Politics more important than technical issues
  - Windows infrastructure, although research is Windows/Linux/MacOS

Application Image Retrieval

- Fairly easy to griddify
- Allows new approaches
  - Local interest points, …
- Limited computing limits techniques
- 40'000 images produced per day

Goals of KnowARC

- Make grid middleware easier accessible to non-experts
  - Little maintenance
  - Non-invasive
  - Graphics user interfaces
  - Virtualizations, accessible under Windows
    - XEN, Damn Small Linux, …
  - …

- Identify problems of domain experts with the middleware and include these into the design
  - Surveys, …

Conclusions

- **Goal of KnowARC is to make grid middleware easily usable and maintainable**
  - Interfaces, installation, adaptation
  - Include various domains into the middleware design (industry, research, administrations)
- **Our goal is to better use our current infrastructure**
  - No investments
  - Acquire knowledge on griddification
  - Develop new research approaches