Tarkvara TAK

Jaak Vilo

STACC Software Technology and Applications Competence Center

26. november 2009

STACC

Software Technologies and Applications

Competence Center

Jaak Vilo

STACC Software Technology and Applications Competence Center

November 26, 2009

Estonia needs R&D in ICT for sustainable software industry

/computer science/

STACC Software Technology and Applications Competence Center

From University to Industry

- Research relevant to companies
- Topics interesting for students
- Results both for academia **and** for industry
- **Funding** from industry, matching from EAS

STACC Software Technology and Applications Competence Center

Timeline

- 2008 August – started planning
- 2008 September – LOI (partners!)
- 2008 Nov - 2009 Jan – planning, negotiating, writing proposal
- 2009 Feb – Hearings; May – decisions
- July – STACC
- August – Consortium Agreement
- Sep-Dec Detailed plans, project launches

STACC Software Technology and Applications Competence Center

Partners

webmedia
INSPIRING PERFORMANCE

REGIO
www.regio.ee

logica

CYBERNETICA

QureTEC

skype

DELFI

IDA-TALLINNA KESKHAIGLA

TARTU ÜLIKOOL
1632
UNIVERSITAS TARTUENSIS

Swedbank

know it

1918
TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY

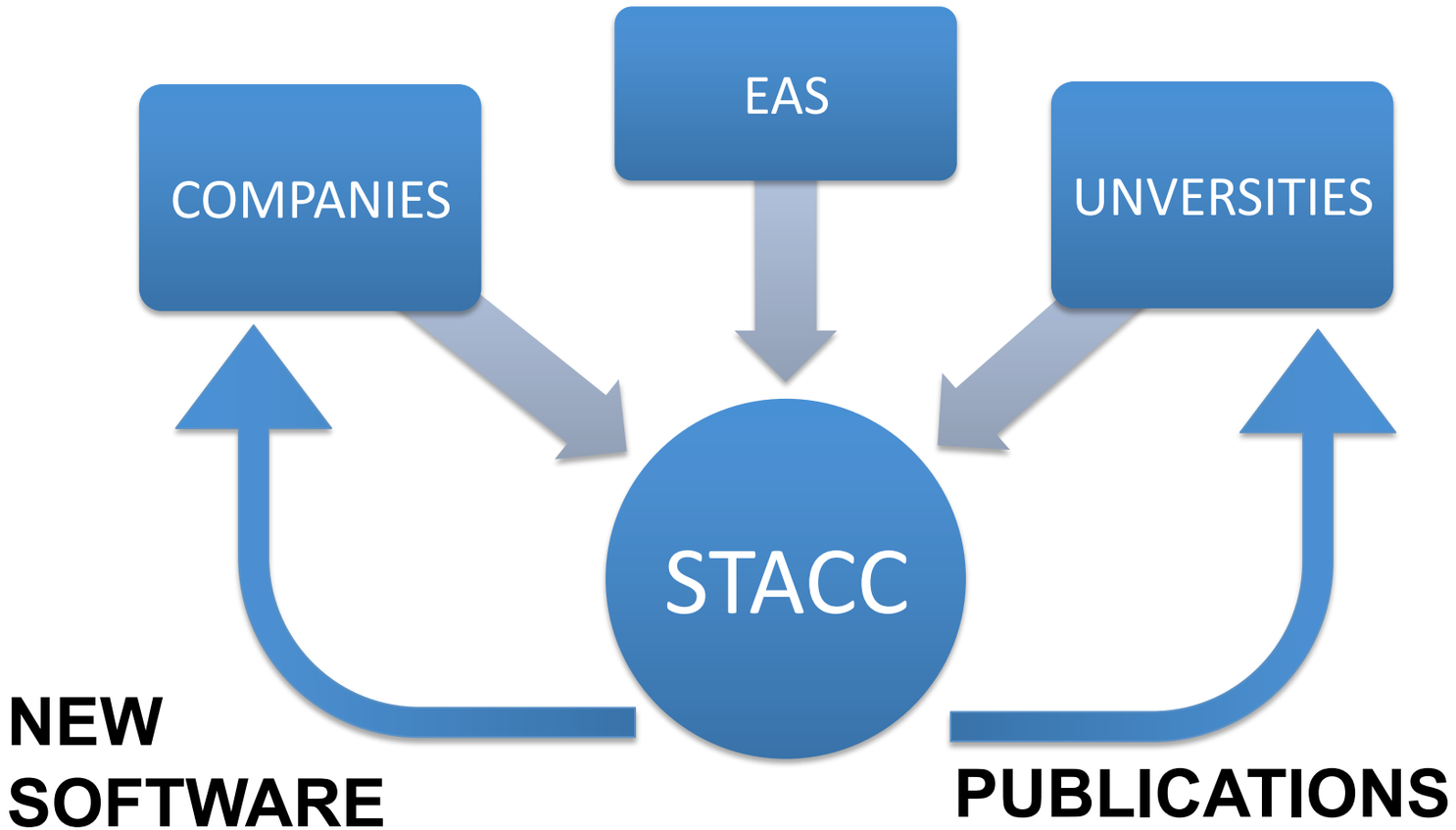
STACC
Software Technology and
Applications Competence Center

Started from a business problem

- Need for better software
- Co-operation between academia and companies
- Enterprise Estonia Competence Center program

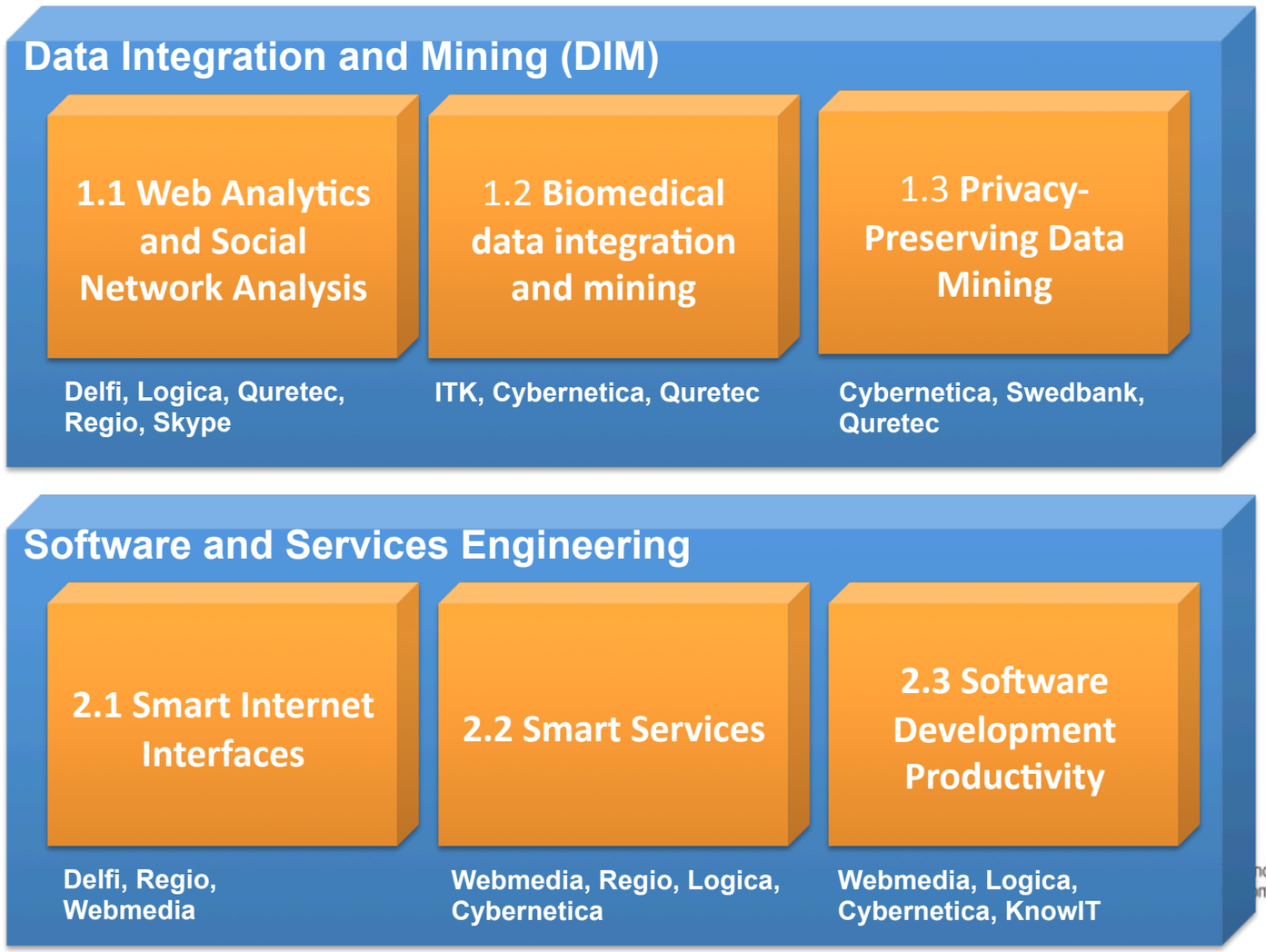
STACC Software Technology and Applications Competence Center

Co-operative R&D center



STACC Software Technology and Applications Competence Center

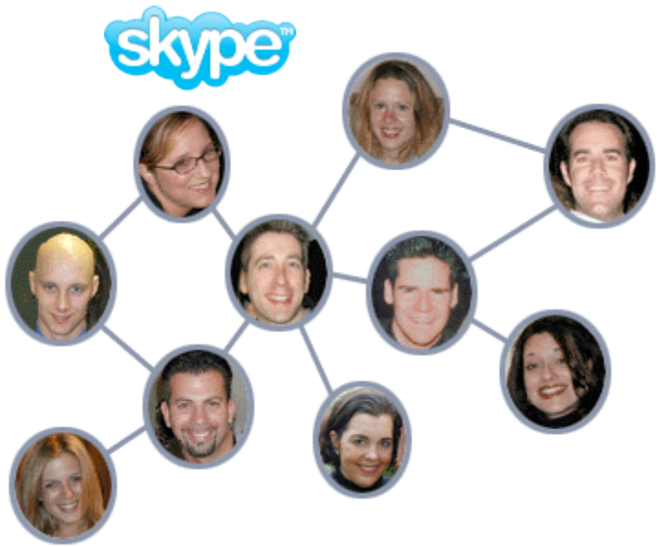
Research Tracks and Programs

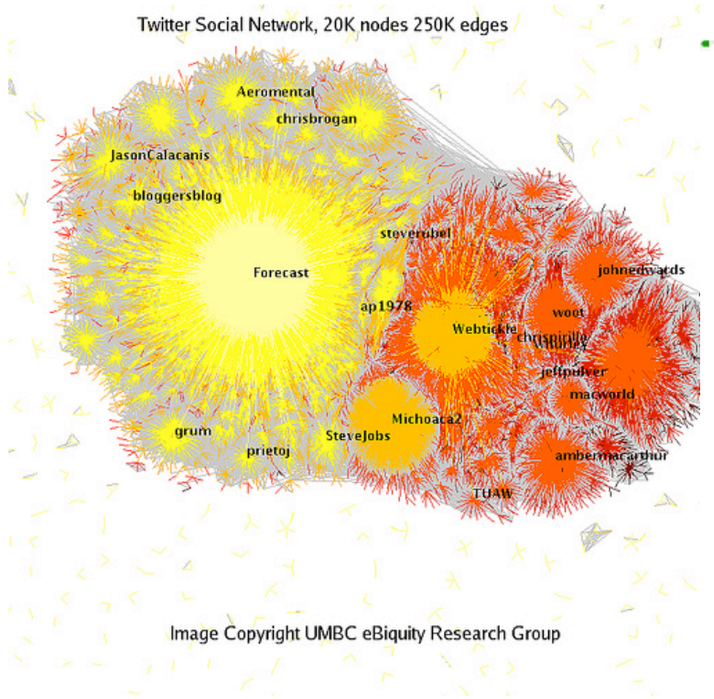


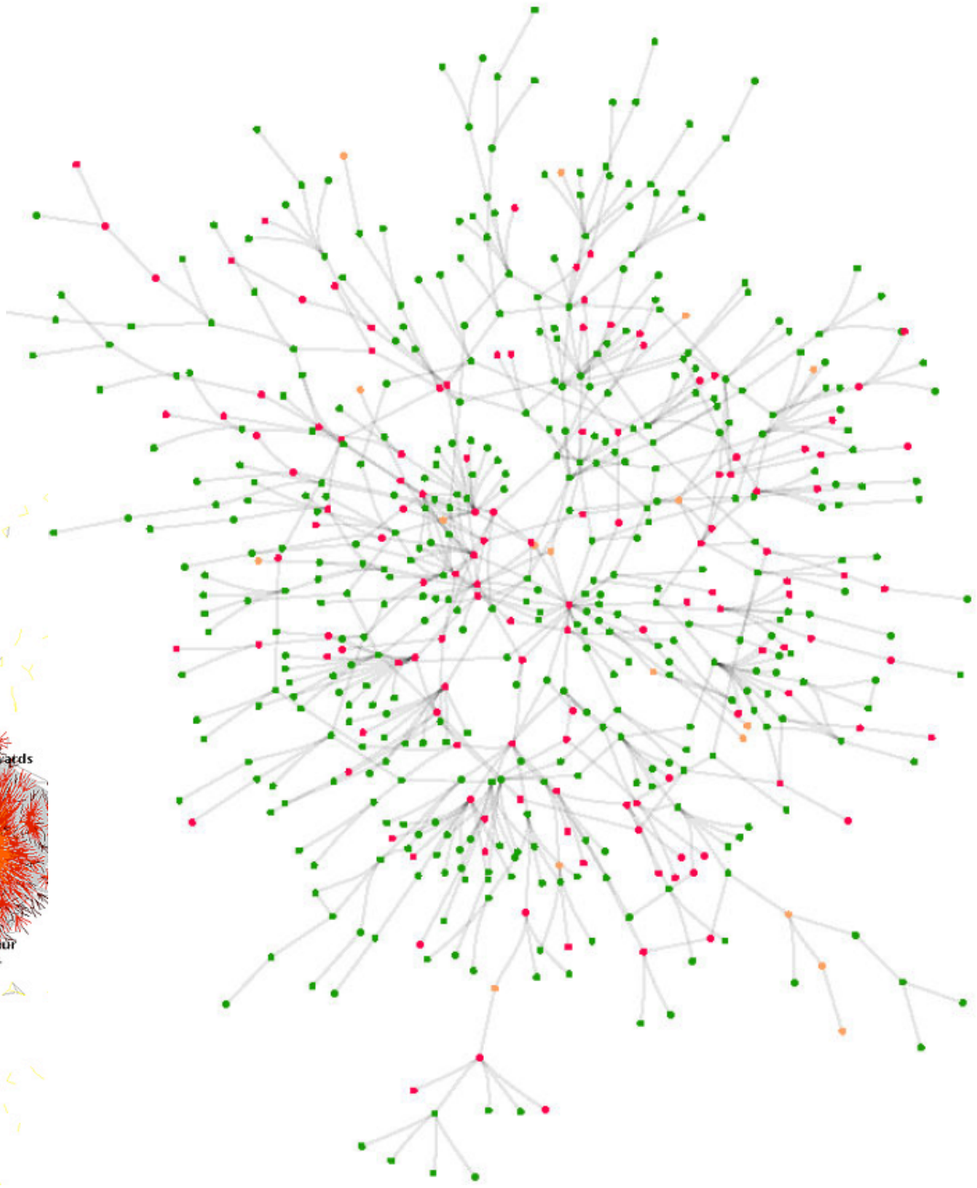
nology and
mpetence Center

1.1 Web Analytics and Social Network Analysis

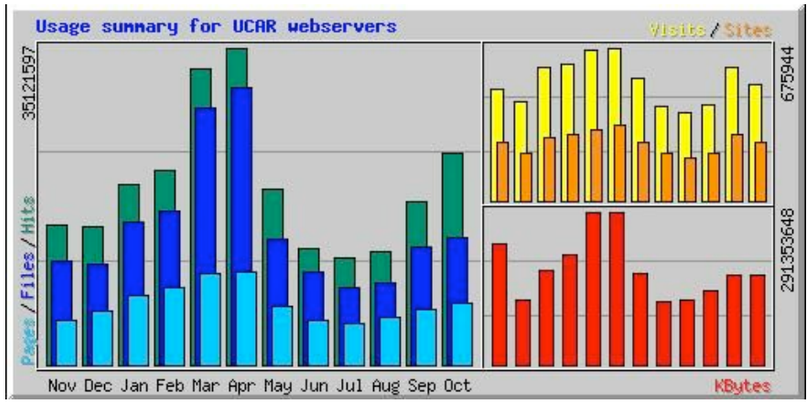
- Methods for **analysing** the **structure and dynamics** of **very large social networks and raw web usage data** in order to discover user clusters, user goals, service or product consumption patterns, customer churning patterns, spam and fraud patterns and other patterns of individual or collective user behaviour.
- Application to **user interface personalization and re-organization**, personalized search, targeted advertising, peer-to-peer network monitoring, and derivation of e-Business metrics associated to advertising, customer acquisition and retention/churning, and business intelligence.

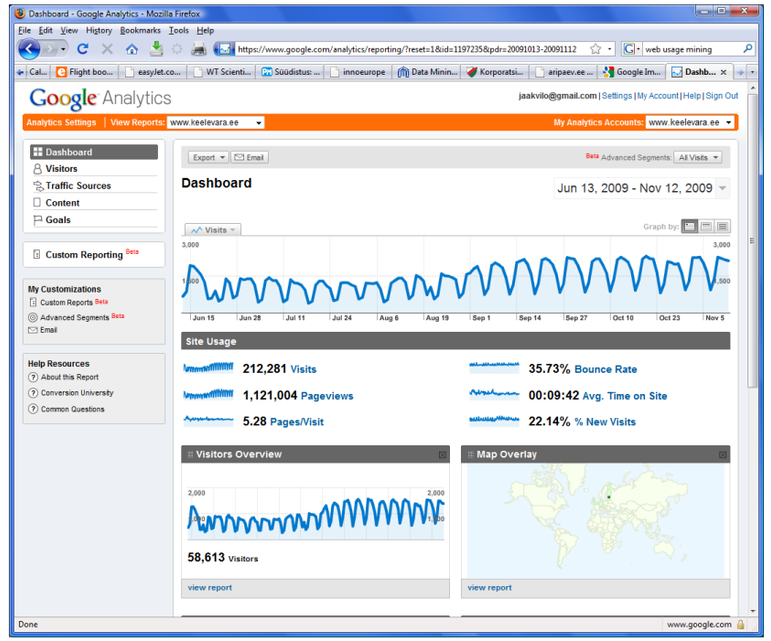

skype


Twitter Social Network, 20K nodes 250K edges
Aeromental
chrisbrogan
JasonCalacanis
bloggersblog
steverubel
Forecast
johnedwards
ap1978
woot
Webtickle
chrisprille
jeffpulver
macworld
Michoaca2
grum
SteveJobs
prietoj
ambermacarthur
TUAW
Image Copyright UMBC eBiquity Research Group

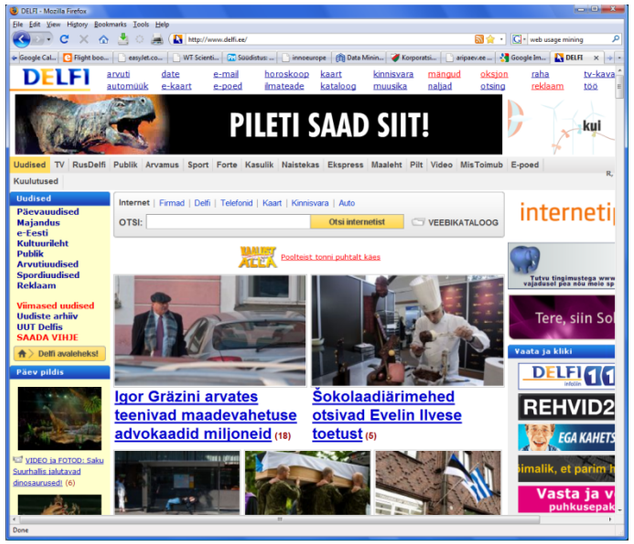


Web usage mining



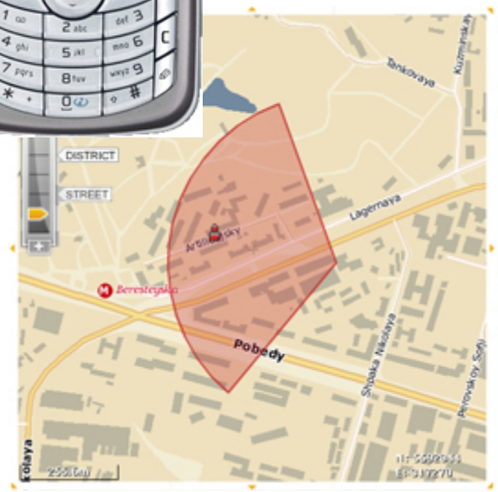


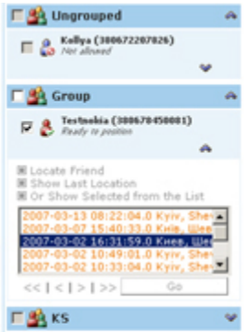
STACC Software Technology and Applications Competence Center



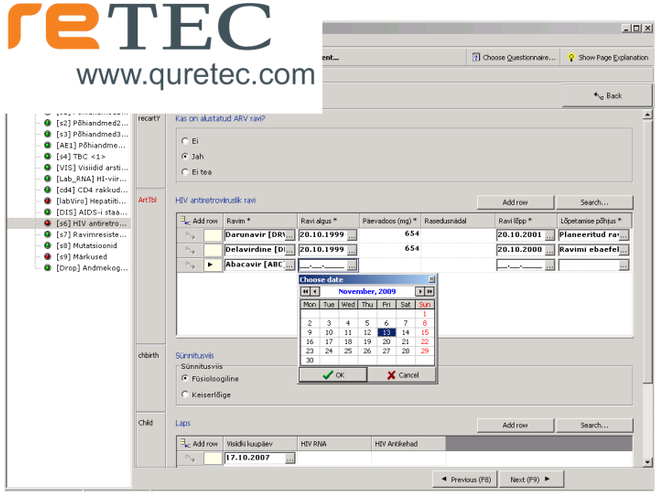
REGIO

Reach-u

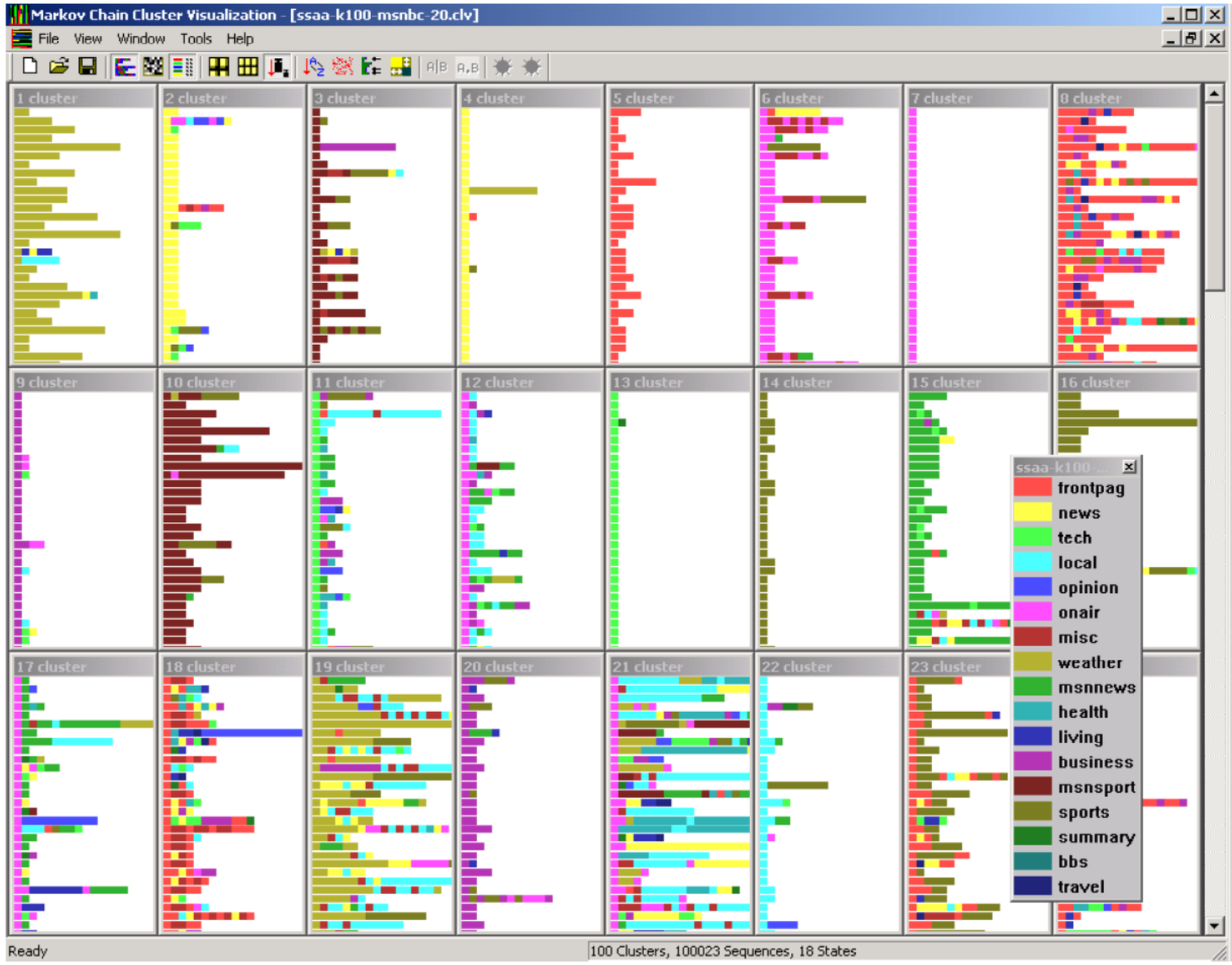




QureTEC
www.quretec.com



STACC Software Technology and Applications Competence Center


Markov Chain Cluster Visualization - [ssaa-k100-msnbc-20.clv]
File View Window Tools Help
1 cluster
2 cluster
3 cluster
4 cluster
5 cluster
6 cluster
7 cluster
8 cluster
9 cluster
10 cluster
11 cluster
12 cluster
13 cluster
14 cluster
15 cluster
16 cluster
17 cluster
18 cluster
19 cluster
20 cluster
21 cluster
22 cluster
23 cluster
ssaa-k100...
frontpag
news
tech
local
opinion
onair
misc
weather
msnnews
health
living
business
msnsport
sports
summary
bbs
travel
Ready
100 Clusters, 100023 Sequences, 18 States

Activities

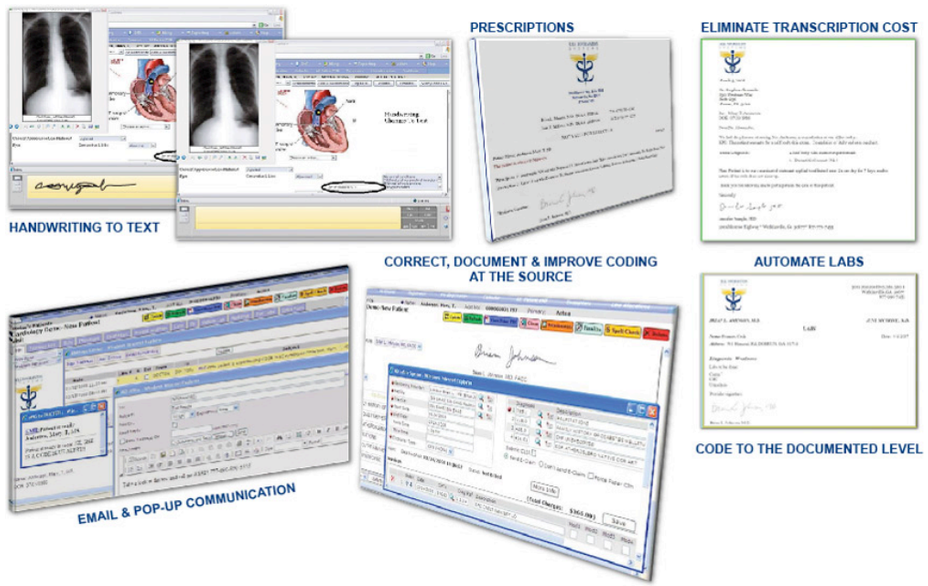
1. Social Network Analysis
   - mining of very large (400M nodes, x 10 edges) graphs for elementary properties
2. Web (and software) log mining
   - warehousing
   - clustering, viusalisation, decision support …
3. User Intent Prediction
   - real time learning of user intent and goals
   - collaborative filtering

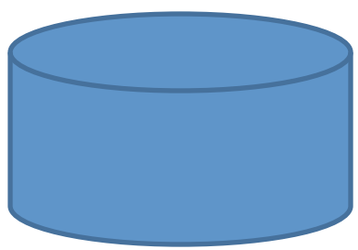
STACC Software Technology and Applications Competence Center

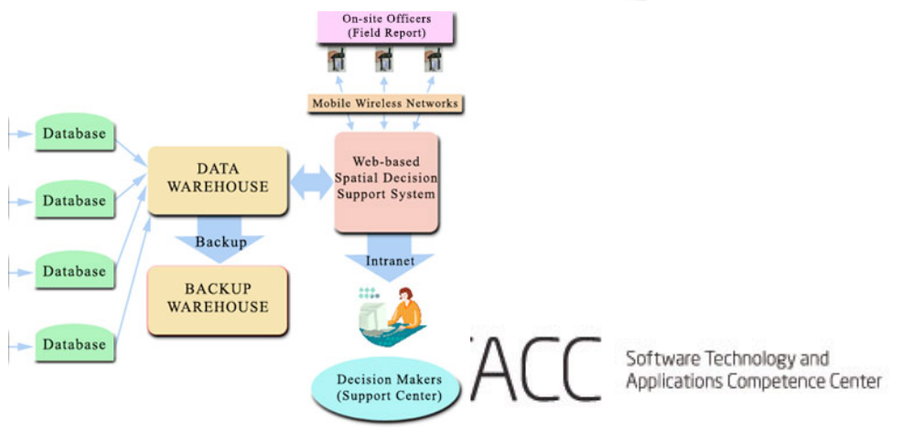
1.2 Biomedical data integration and mining

- To develop **data integration and analysis methods** for **electronic patient records** and **biomarker data** with the goal of improving the (early) diagnosis and medical treatment of the diseases.
  - Complex disease of COPD as the proof of principle
  - Mining Electronic patient records and E-Health data

- UT – data mining; TTU – medical know-how and data
- Hospital and e-health solutions


HANDWRITING TO TEXT
PRESCRIPTIONS
ELIMINATE TRANSCRIPTION COST
CORRECT, DOCUMENT & IMPROVE CODING AT THE SOURCE
AUTOMATE LABS
EMAIL & POP-UP COMMUNICATION
CODE TO THE DOCUMENTED LEVEL




On-site Officers (Field Report)
Mobile Wireless Networks
Database
Database
Database
Database
DATA WAREHOUSE
Backup
BACKUP WAREHOUSE
Web-based Spatial Decision Support System
Intranet
Decision Makers (Support Center)
ACC
Software Technology and Applications Competence Center

Activities

1. COPD patient cohort buildup based on smoking etc characteristics
2. Data warehousing and decision support for hospital e-health data
3. Introducing ontologies into clinic
4. Text mining of medical records

STACC Software Technology and Applications Competence Center

1.3 Privacy-Preserving Data Mining

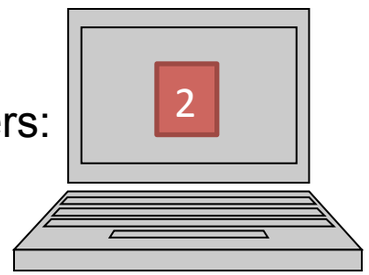
To develop and to evaluate **privacy-enhancing methods** for data storage and processing, along two complementary directions:

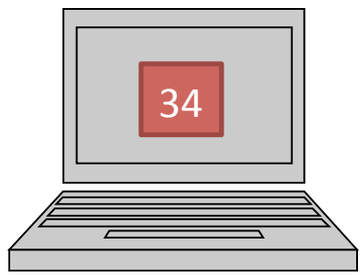
1. **Security of micro-data releases and query auditing**: Detection and elimination of possible privacy breaches in data to be published, and detection of queries to (medical and financial) databases that may breach the privacy of individuals.
2. **Privacy-preserving data aggregation**: Development of secure and practically efficient methods for aggregating data from multiple sources, that leak nothing beyond the end aggregate results.

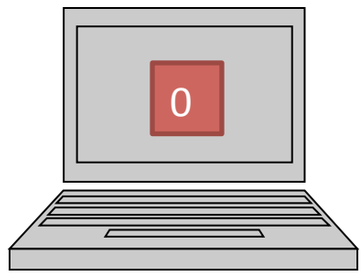
Privacy preserving DM

Q: sexual behavior of HIV patients?

sex partners:



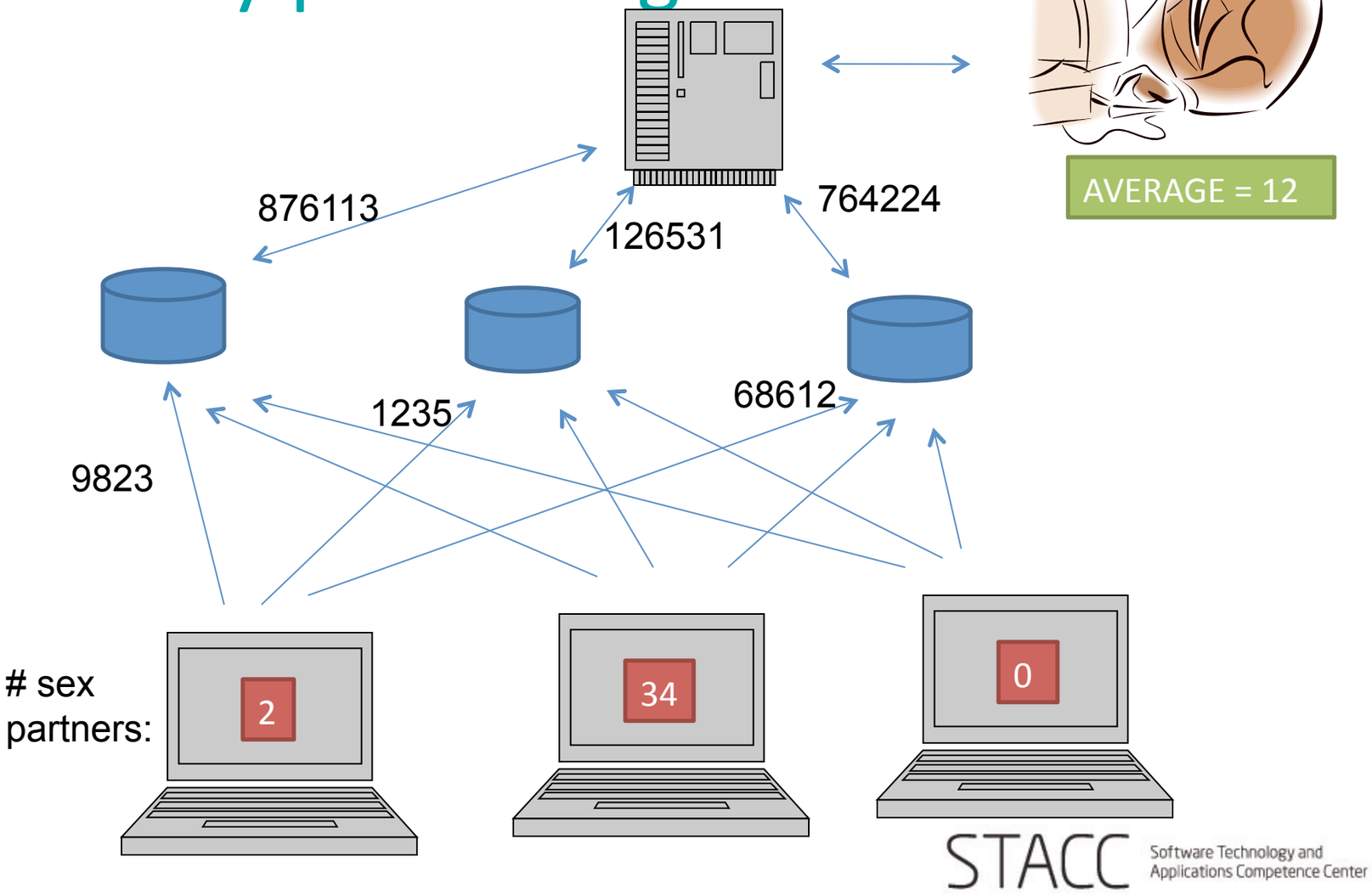




STACC Software Technology and Applications Competence Center

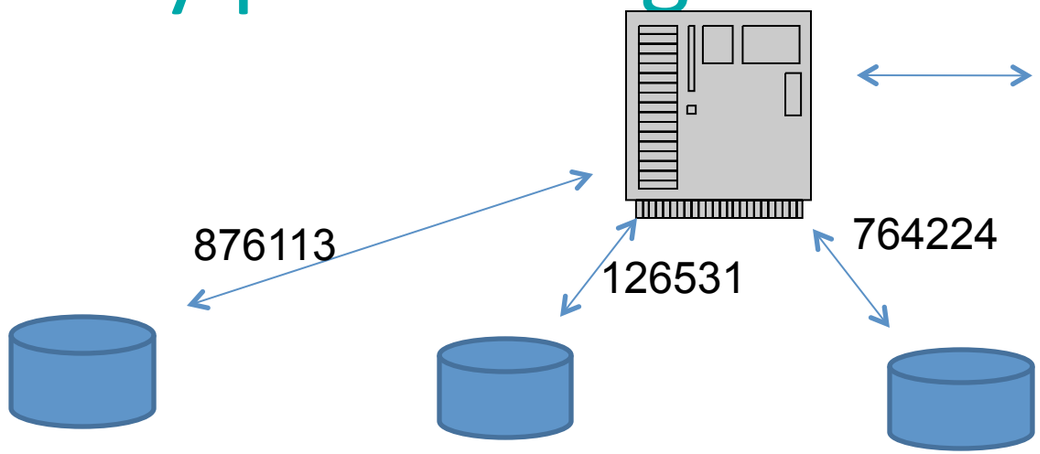
Privacy preserving DM

AVERAGE = 12



STACC Software Technology and Applications Competence Center

Privacy preserving DM



AVERAGE = 12

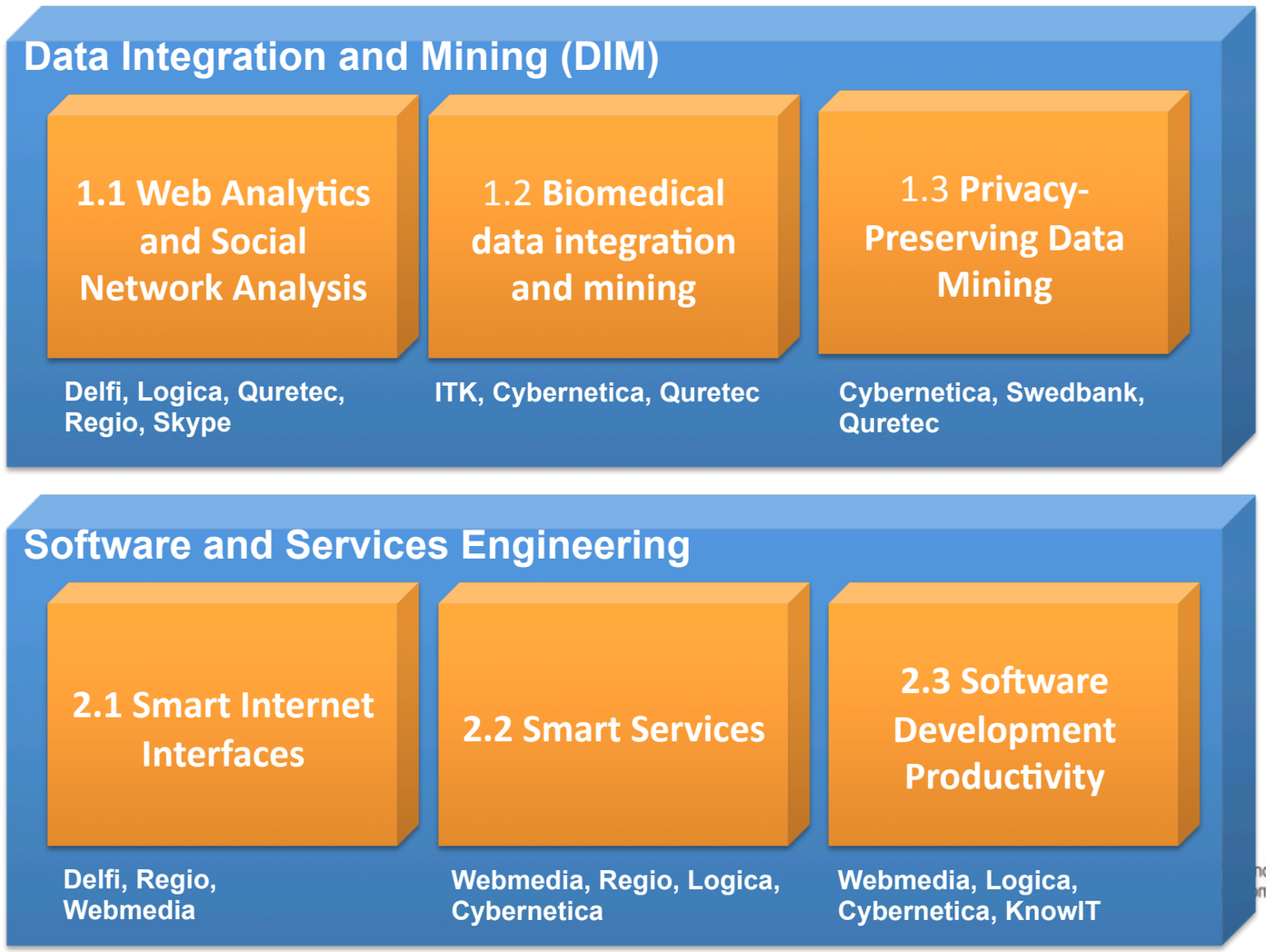
ShareMind

STACC Software Technology and Applications Competence Center

Activities in 1.3

1. Micro-data protection mechanisms
2. An environment for developing privacy-preserving applications
3. Demonstration: questionnaire system
4. Protocol analysis for secret-shared applications

Research Tracks and Programs

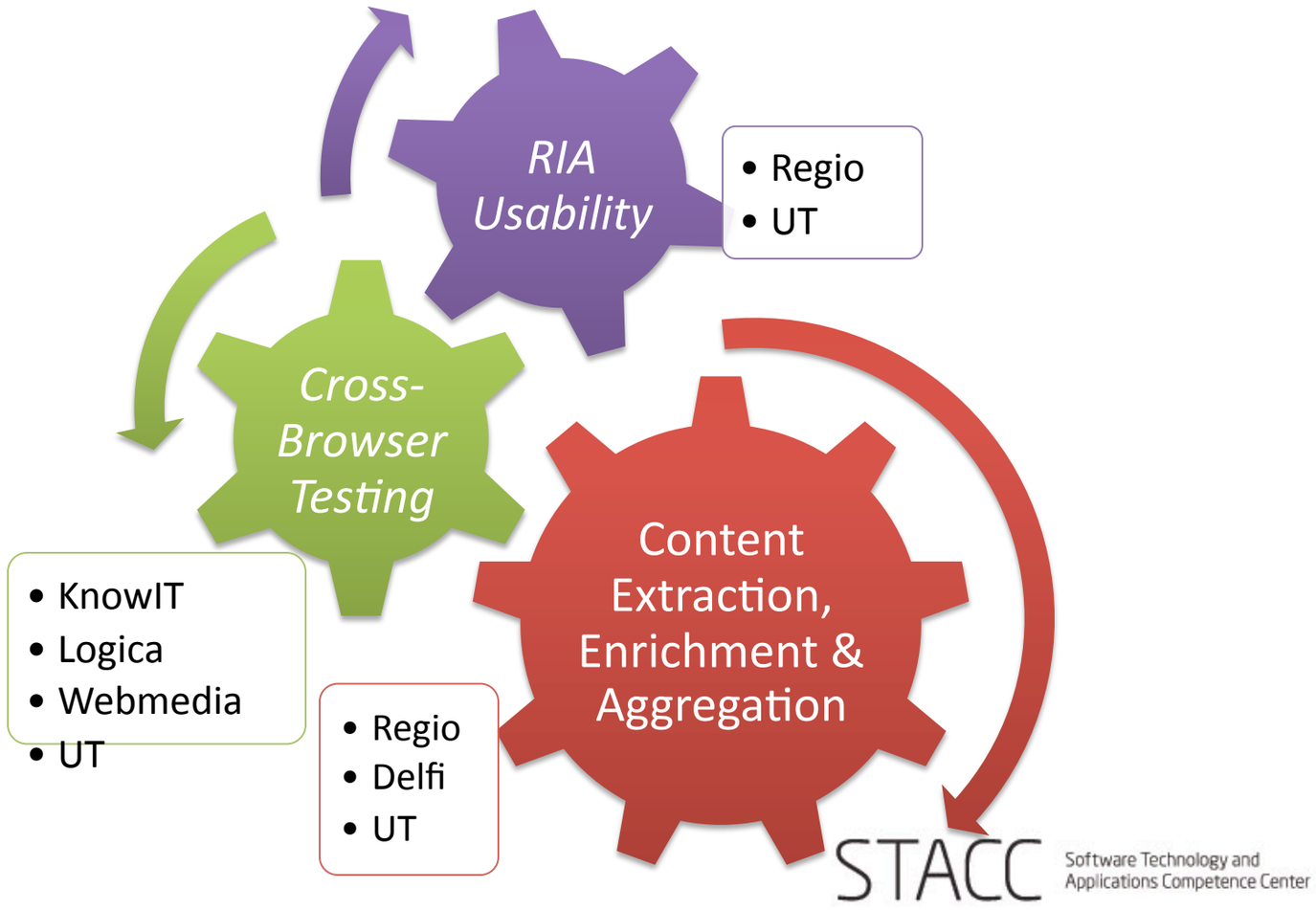


nology and
mpetence Center

2.1 Smart Internet Interfaces

- Vision: To enable the cost-effective development of user interfaces that deliver rich and personalized content over the Internet
- Such “Smart Internet Interfaces” are able to:
  - Enrich content with value-added services
  - Aggregate content from multiple sources
  - Adapt content with respect to user profiles and behaviour
- All this while ensuring:
  - Usability
  - Cross-browser and cross-device compatibility

2.1 Smart Internet Interfaces (cont.)



2.2 Smart Services

- Vision: To enable the cost-effective maintenance and enhancement of service networks
- This is achieved by:
  - Providing tools for rapid reconciliation of heterogeneous interfaces
  - Monitoring services in order to detect and resolve changes in their behaviour
  - Mining networks of services in order for:
    - Detect and analyse Service Level Agreement (SLA) violations
    - Predict SLA violations
    - Identify redundancy and re-design opportunities

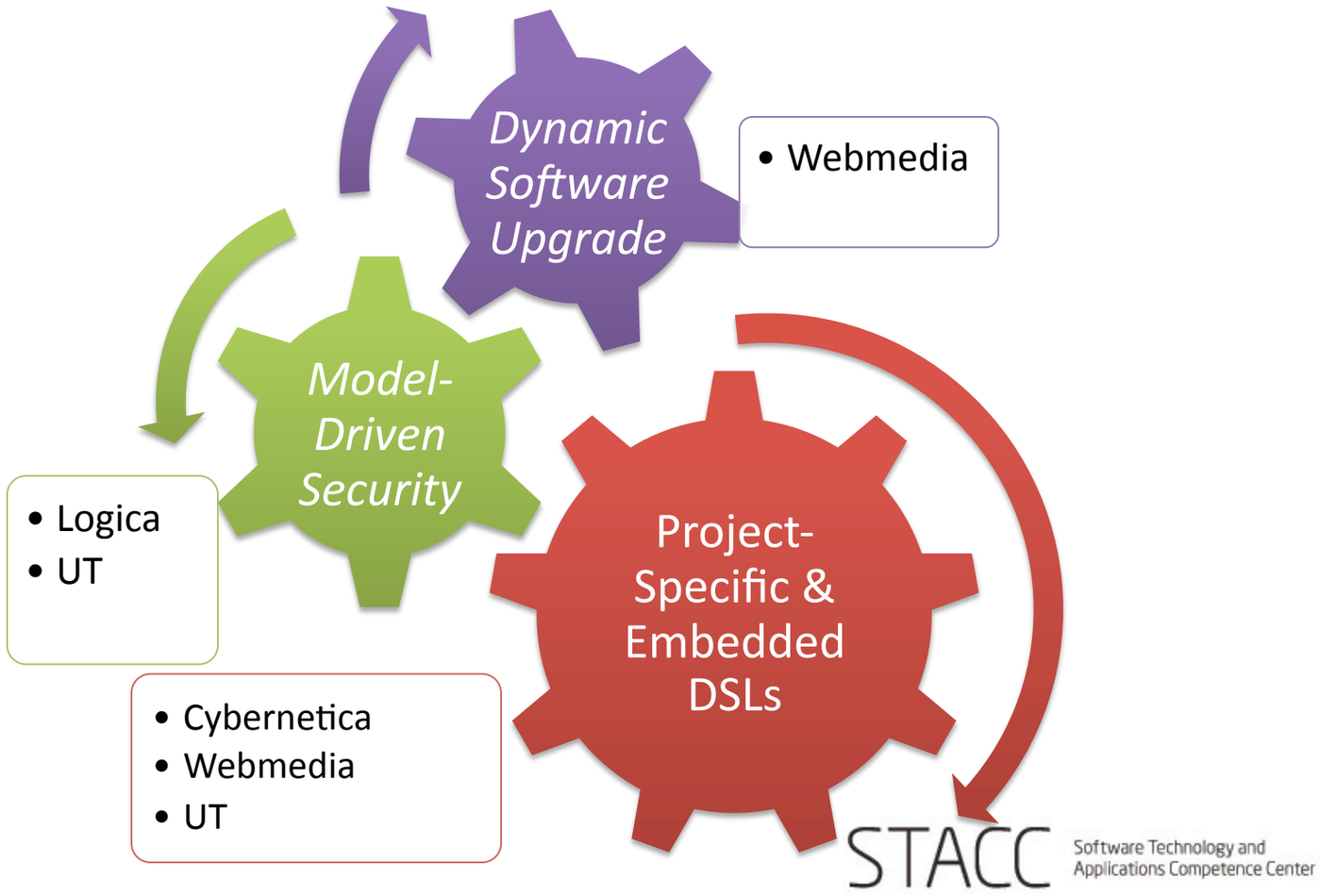
Webmedia, Logica

STACC Software Technology and Applications Competence Center

2.3 Software Development Productivity

- To enhance the productivity of software developers by supporting business-driven and agile development methods
- Specifically, :
  - Dynamic software upgrade and continuous integration
  - Domain-Specific Languages (DSLs)
  - Model-Driven Development (MDD)
  - Real-time software development visibility

2.3 Software Dev. Productivity (cont.)



2.3 Software Dev. Productivity (cont.)

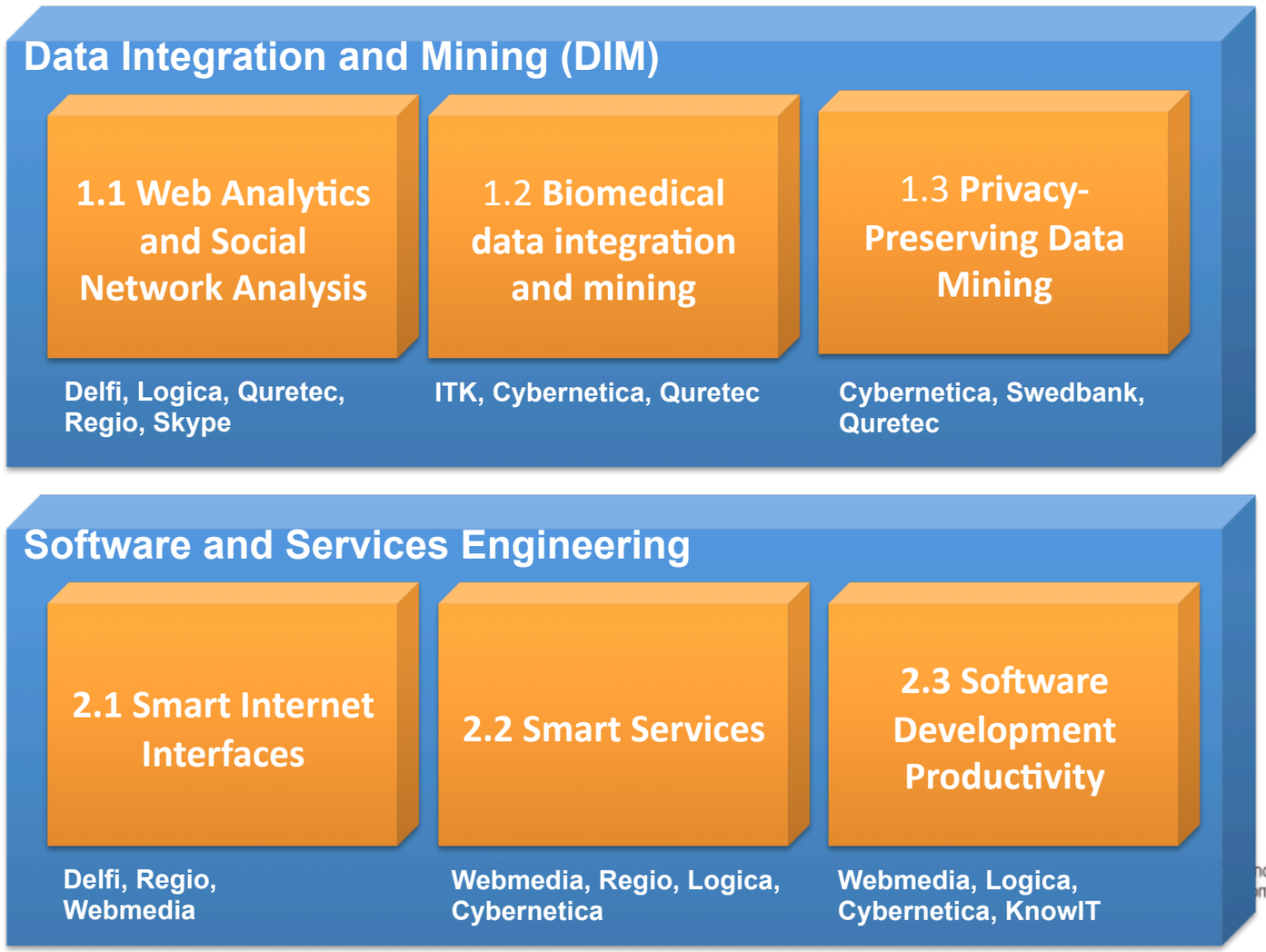
Example of research problem: Detect syntax an typing errors in embedded DSL expressions, e.g. SQL. For example, find errors here:

```
String query = “SELECT FORENAME, SURNAME,   POSITION
FROM EMPLOYEES, JOBS WHERE EMPLOYEE.JOB_ID = JOB.ID”;
if (managementOnly)
   query+=“ AND JOB.TYPE = ”;
query+=“ ORDER BY ID DESC”;
statement.prepareStatement (query);
ResultSet rs = query.execute();
```

2.3 Software Dev. Productivity (cont.)

- Other examples (Model-Driven Security):
  - Generate database access constraints from UML models in order to prevent unauthorised access to a database
  - Automatically hide or inhibit user interface elements that a user does not have the right to see or to modify

Research Tracks and Programs



nology and
mpetence Center

Formula for success

- **Challenges and demands from industry**
- **Researchers and students from university**
- **Funding**
- **Commercial and academic impact**

STACC Software Technology and Applications Competence Center

STACC - people

Council:
Monika Oit, Cybernetica
Teet Jagomägi, Regio
Ivo Mägi, Webmedia
Andreas Sisask, Logica Eesti
Margus Jäger, Quretec
Sten Tamkivi, Skype
Kaspar Loog, Know IT Estonia
Kuldar Kuremaa, ITK
Indrek Jakobson, TTÜ
Indrek Ots, Tartu University
Aho Augasmägi, Swedbank
Priit Piile, Delfi

SAB:
Stefan Tai
Hannu Toivonen
Ülo Parts

Research leaders:
Jaak Vilo
Ruth Sepper
Jan Willemson
Marlon Dumas
Peep Küngas

CEO:
Indrek Vainu

Researchers,
Engineers,
Programmers

Currently ~ 20

Total: ~35-40

STACC Software Technology and Applications Competence Center