Background in Web Development

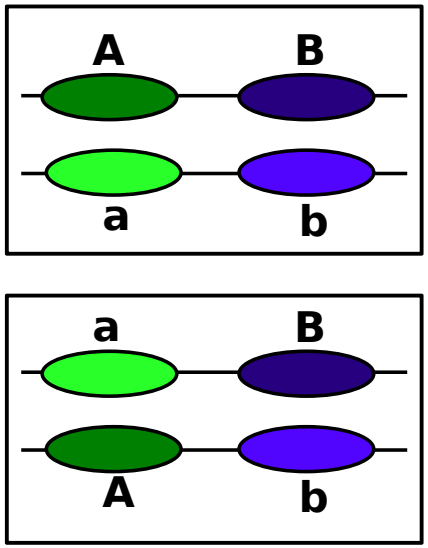
Michał Nowotka
job applicant

EMBL-EBI
ChEMBL group

June 12, 2012

The problem of haplotype frequency estimation – Bachelor thesis

AaBb ?



- Determining haplotypes with laboratory methods is expensive and time-consuming.
- In contrast, there are many cost-effective techniques for determining genotypes.
- In general, it could be impossible to infer haplotypes from genotype data.

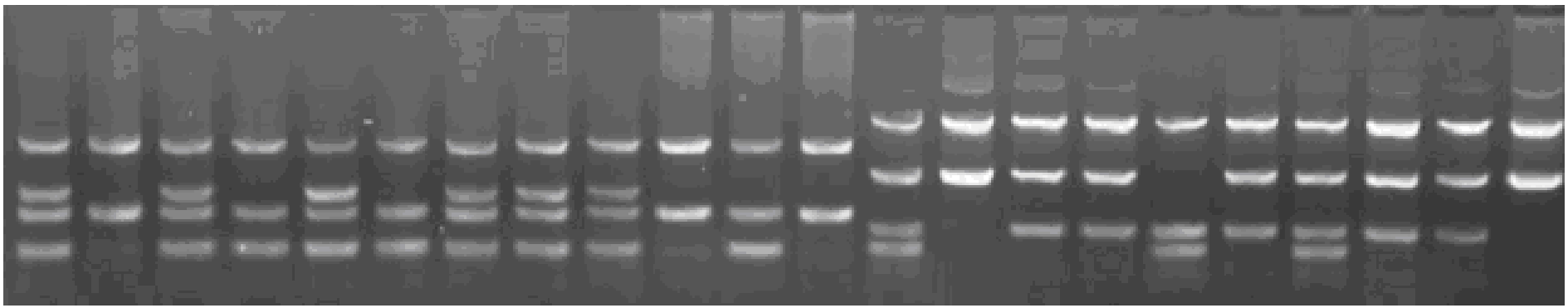


Figure:
Determining genotype experiment results

Idea of short overlapping window

Problem

Every algorithm employing full space search would operate with $O(c^n)$ complexity. This is why it cannot be directly applied to phasing long genotypes.

Solution – Genotypes can be divided into shorter pieces that overlap.

- Piece length is fixed, so is computation time.
- Phasing n pieces has now $O(n)$ complexity.
- Multiple pieces can be phased in parrallel.
- If phasing algorithm is convergent total error should not be large.

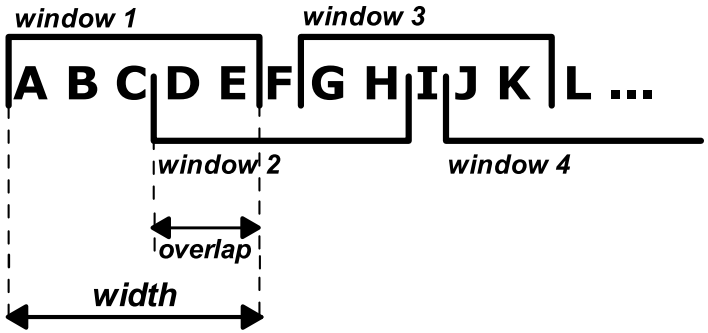
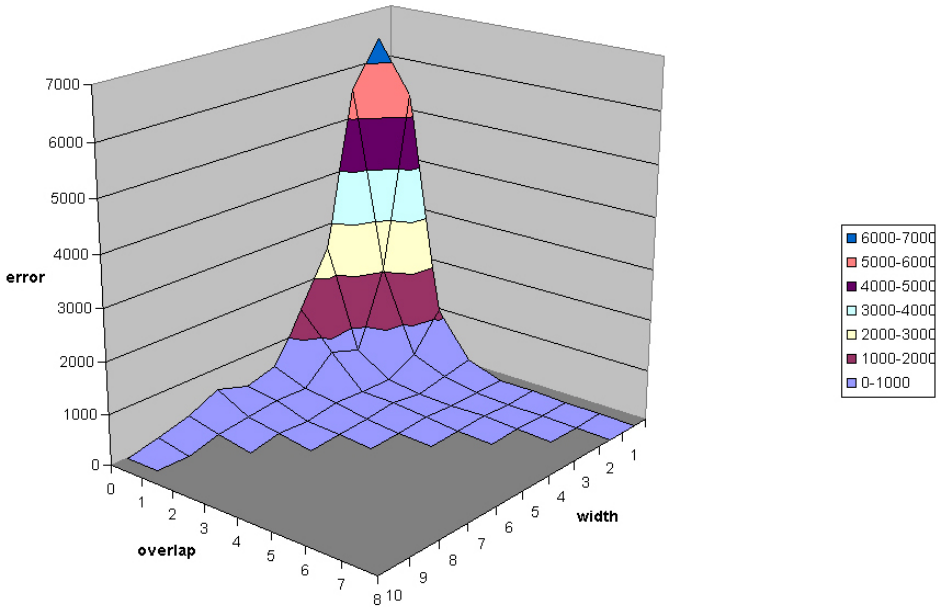
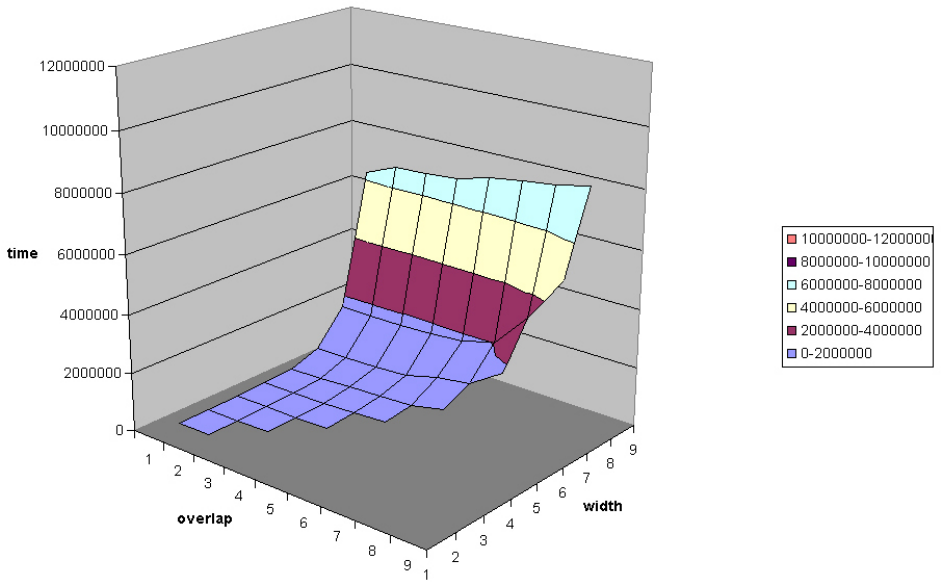


Figure: What are the error and execution time as a function of **width** and **overlap** parameters?

Results

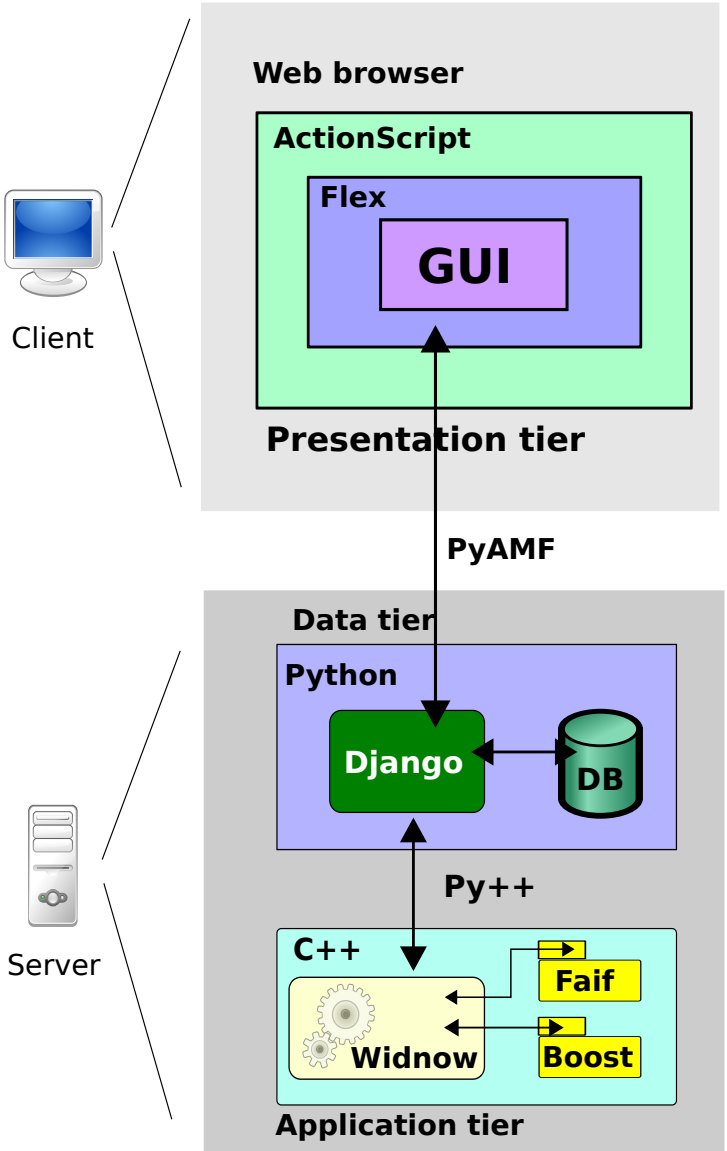


Error as a function of width and overlap parameters



Execution time as a function of width and overlap parameters

Application architecture



Automated functional annotation using classification algorithms and data fusion – master thesis

Functional genomics as a major field in applied bioinformatics

- Functional interpretation is a key step in the analysis of DNA and protein sequences.
- This task cannot be done without extensive functional annotation of the datasets.
- Due to the fast development of high-throughput sequencing technologies, an increasing amount of novel, uncharacterized sequence data have arisen.
- Standarized functional annotation is essential.

The goal

Provide biologists with useful information to take into account when addressing the task of functionally characterizing their sequence data.

Automated functional annotation – the algorithm

Input

Uncharacterized DNA or protein sequence.

- BLAST.
- Gene Ontology lookup.
- Data fusion and inference.

Output

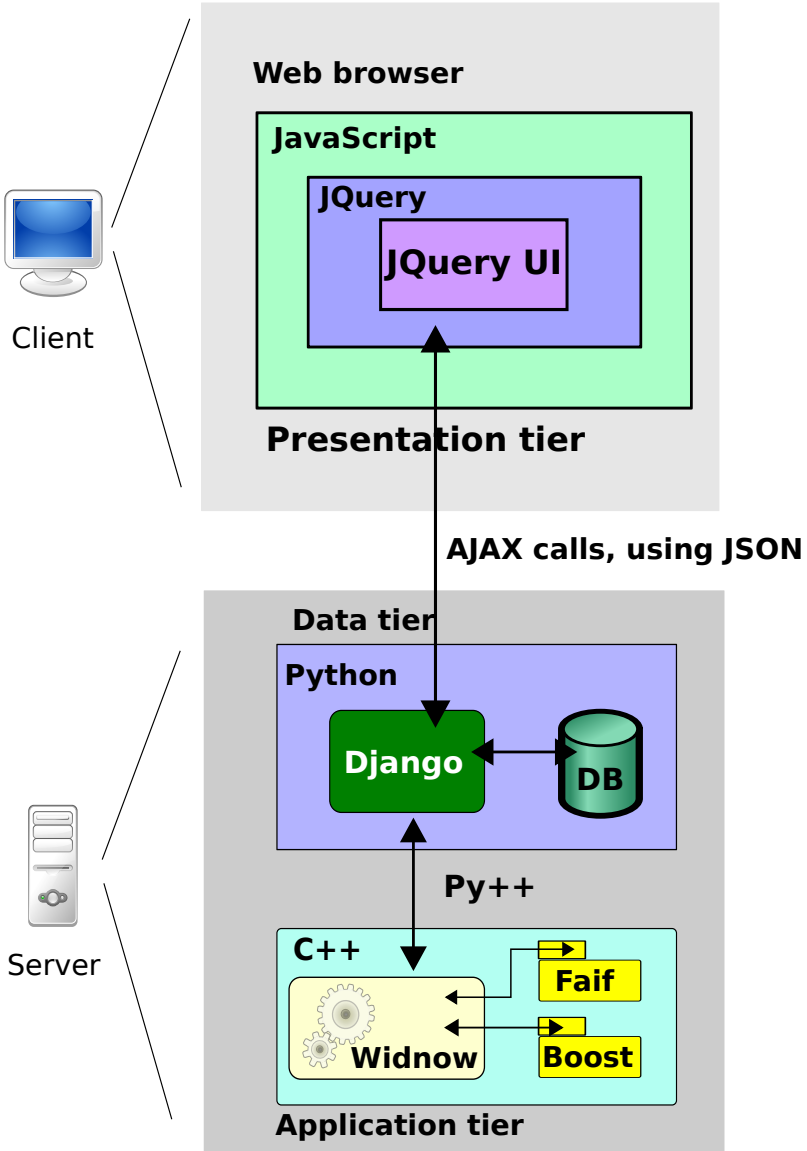
Inferred functional annotation for the input sequence

Inferring functional annotation

For combining multiple results the Dempster's rule of combination is used.

- Often used as a method of sensor fusion.
- Strongly emphasises the agreement between multiple sources and ignores all the conflicting evidence.
- Better alternative to weighted voting.

Application architecture



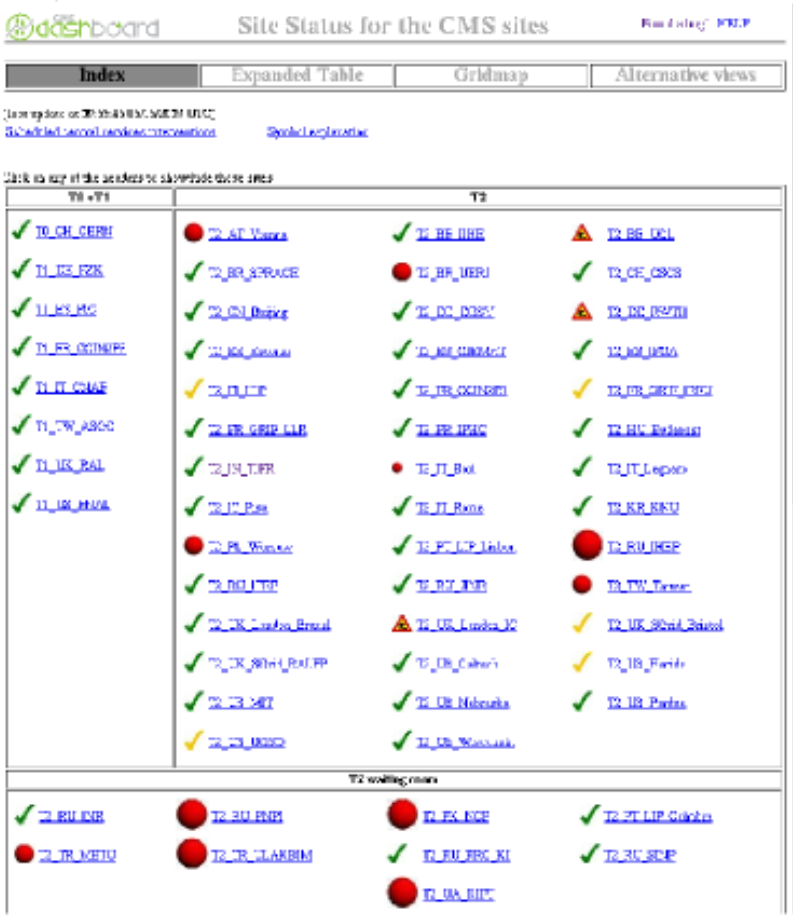
Site Status Board – an application monitoring the behaviour of all the centers of a particular VO

- SSB provides a single entry point that summarizes the status of the sites.
- The main idea is to provide a flexible framework which would allow VOs to define multiple monitoring metrics.
- The metrics can be added, deleted and easily modified.
- The most critical metrics can be combined into a single value for each site corresponding to its status.
- SSB keeps the history of how all the metrics have evolved over time..
- SSB consists of three components: collectors that gather information, a database and a web server.

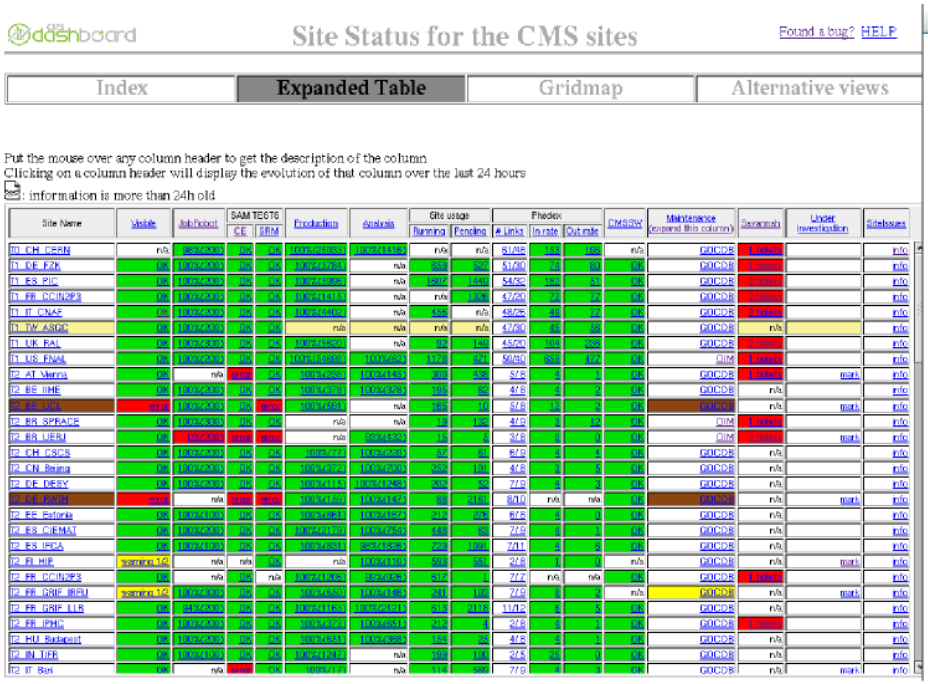
SSB – implemented features

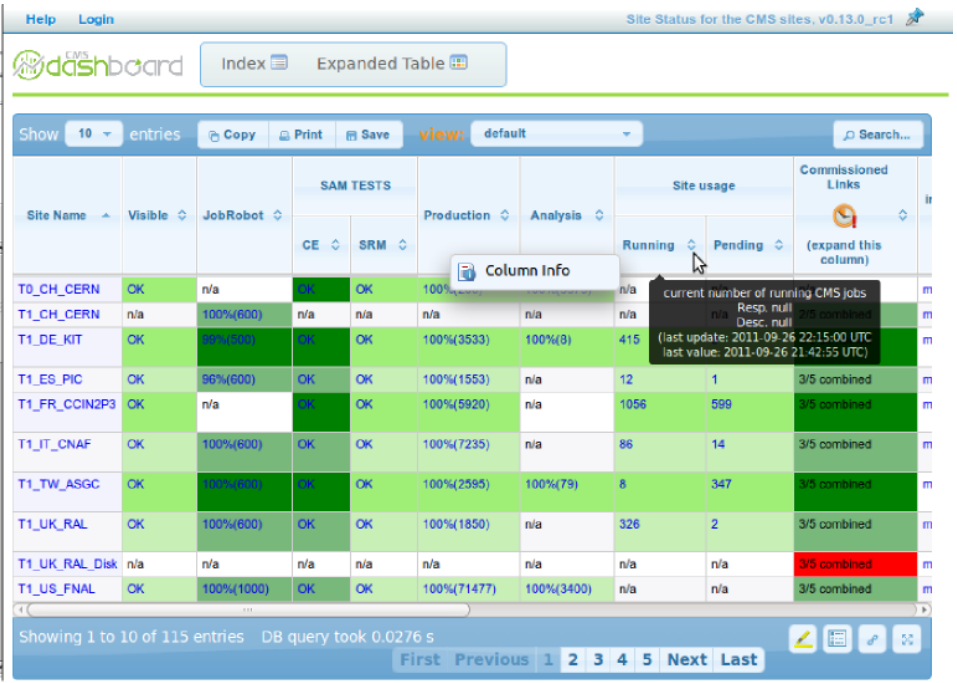
- XSLT replaced by Java Script template system.
- New coherent GUI.
- Filtering, paging, sorting in Expanded Table, computed on server side.
- Expanded Table ready for large amount of data.
- Redesigned backend.
- Client-side plotting.
- Bookmarking, undo/redo.
- Backbone.

Old and new SSB

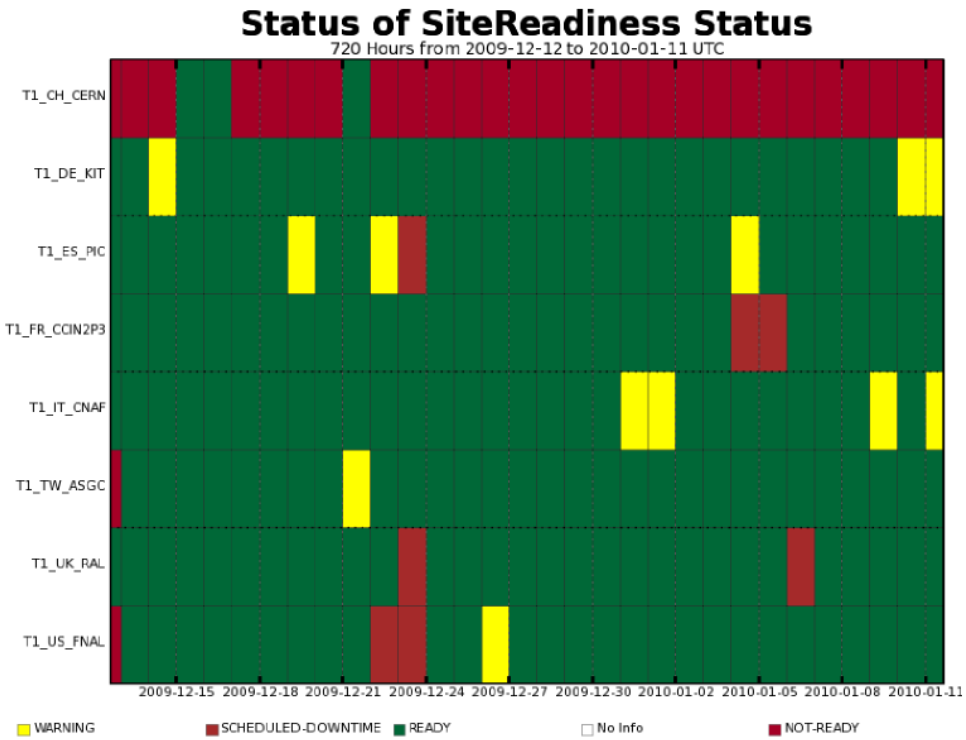


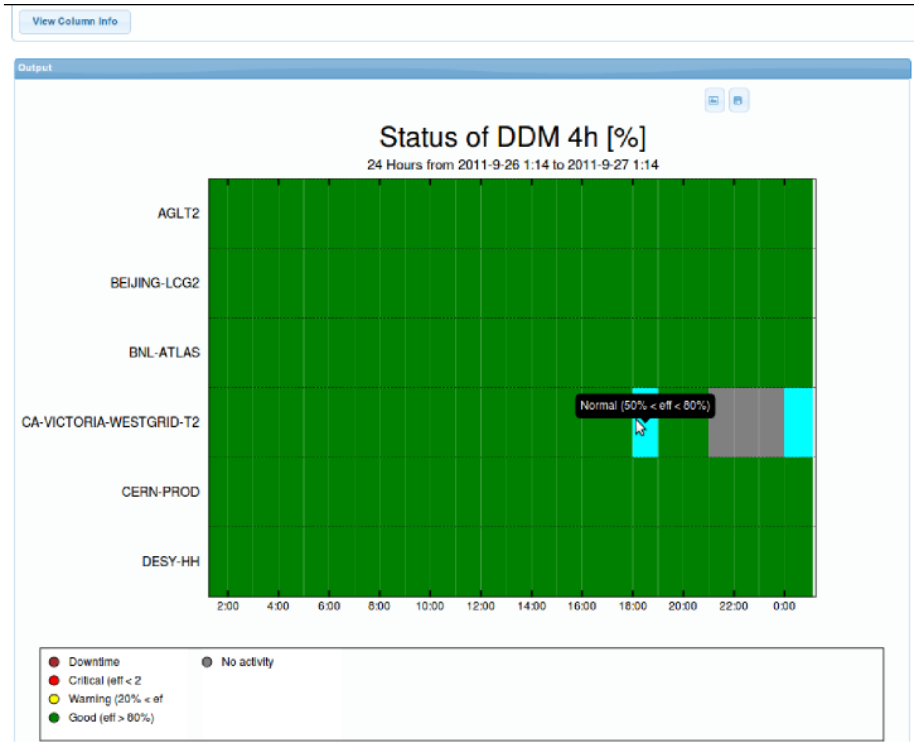
Old and new SSB





Old and new SSB



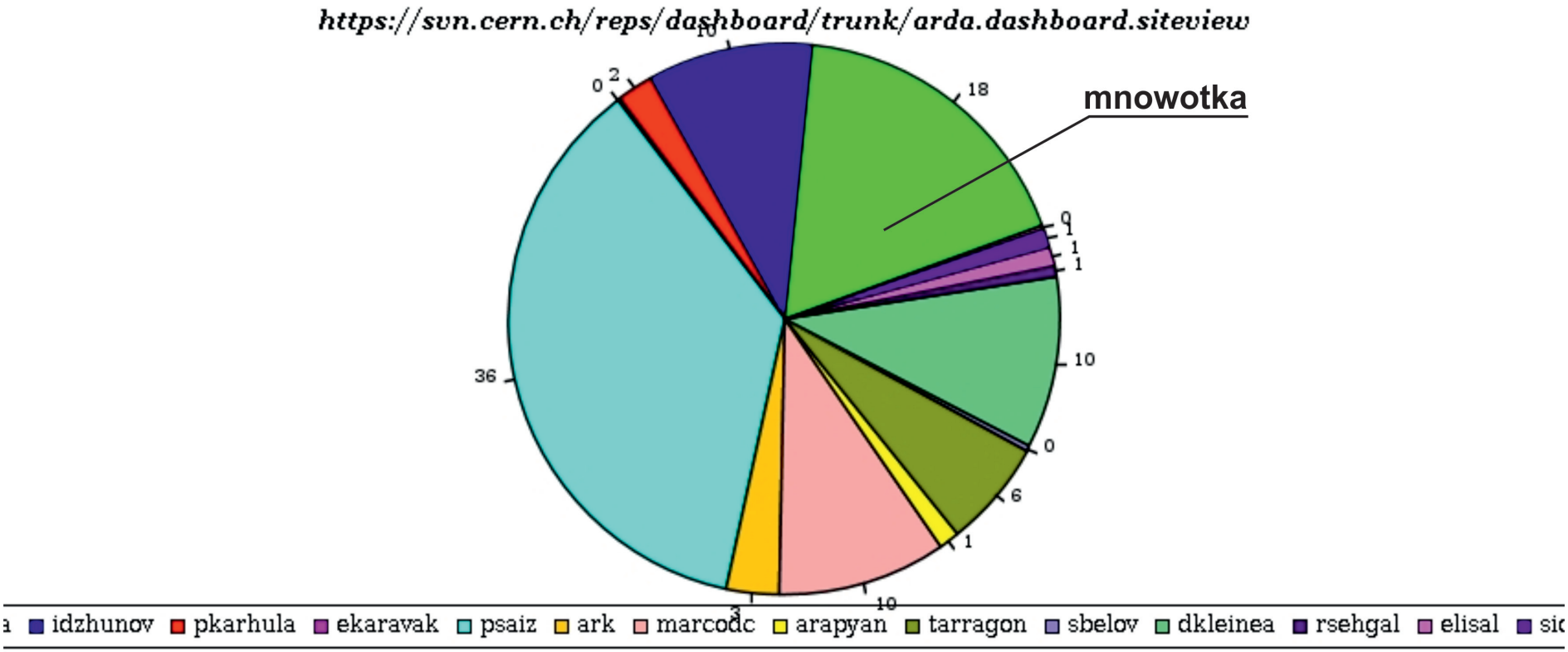


SSB – collector metainformation

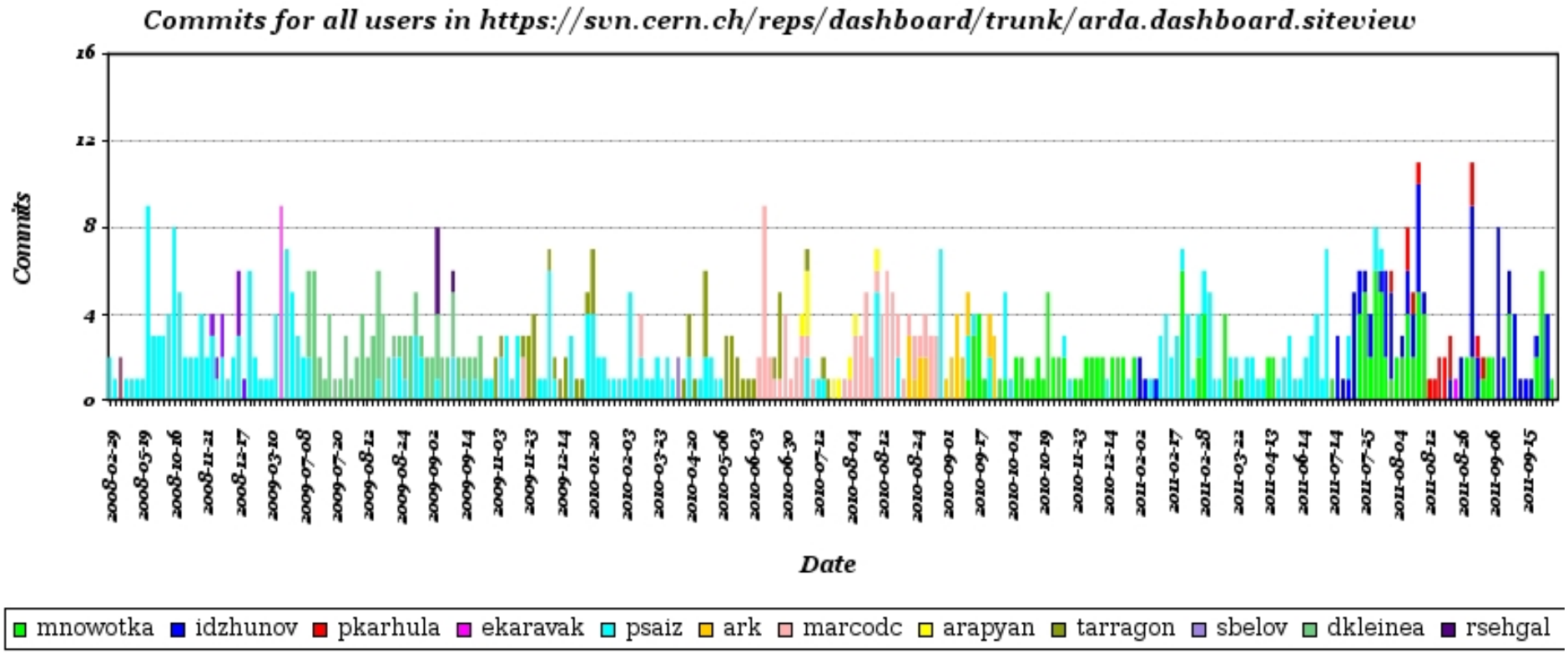
SSB – TODO

- Tests (jQunit, Selenium).
- Database synchronization.
- Web based installation wizard.
- Getting rid of FOUCs.
- Refactoring of DAO.
- Expanded Table should refresh periodically and highlight recent changes.
- NoSQL for Sieview Data.

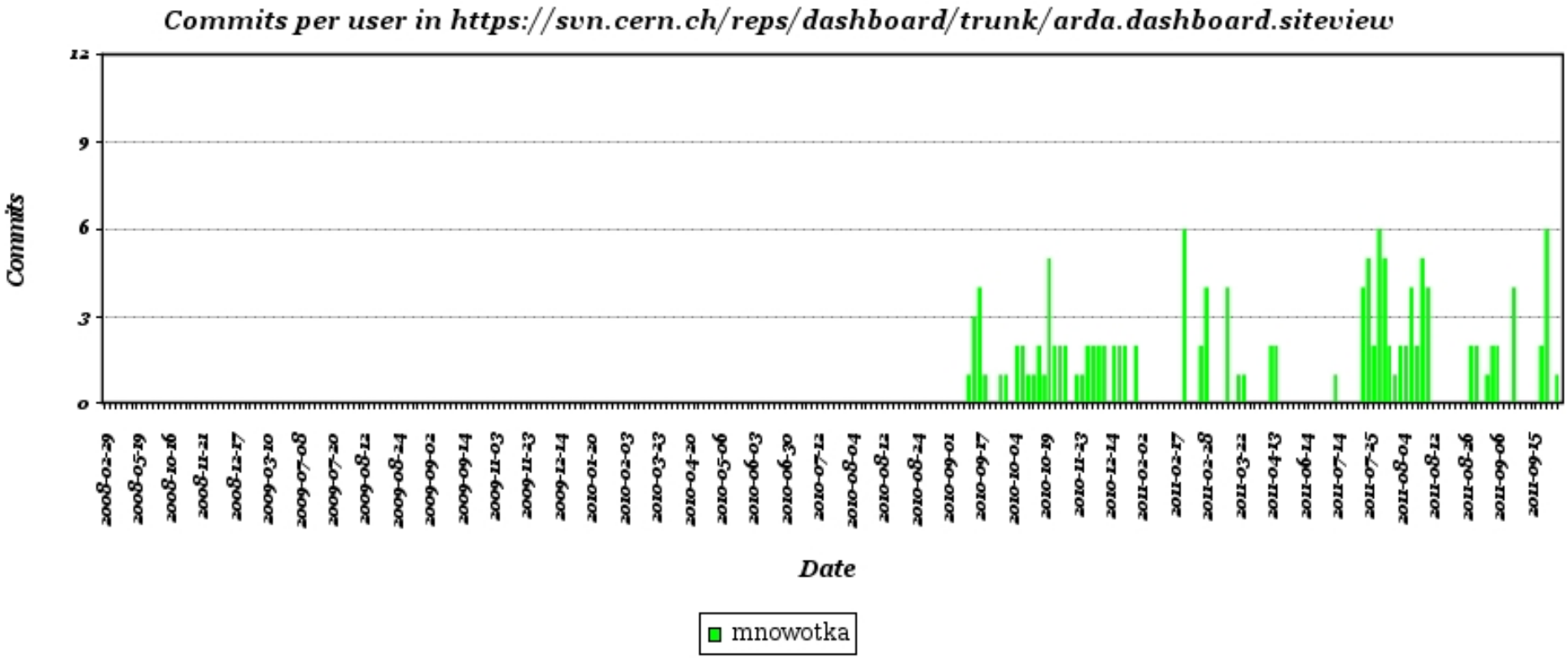
SSB – Impact chart



SSB – Commits


Commits for all users in https://svn.cern.ch/reps/dashboard/trunk/arda.dashboard.siteview
Commits
16
12
8
4
0
2008-02-29
2008-05-19
2008-10-16
2008-11-21
2008-12-17
2009-03-10
2009-07-08
2009-07-20
2009-08-12
2009-08-24
2009-09-02
2009-09-14
2009-11-03
2009-11-23
2009-12-14
2010-01-20
2010-02-03
2010-03-23
2010-04-20
2010-05-06
2010-06-03
2010-06-30
2010-07-12
2010-08-04
2010-08-12
2010-08-24
2010-09-01
2010-09-17
2010-10-04
2010-10-19
2010-11-23
2010-12-14
2011-02-02
2011-02-17
2011-02-28
2011-03-22
2011-04-13
2011-06-14
2011-07-14
2011-07-25
2011-08-04
2011-08-12
2011-08-26
2011-09-06
2011-09-15
Date
mnowotka
idzhunov
pkarhula
ekaravak
psaiz
ark
marcodc
arapyan
tarragon
sbelov
dkleinea
rsehgal

SSB – My commits



Framework

Benefits for the dashboard framework:

- Coherent set of tools and libraries.
- Proofs of concepts.
- Authentication mechanisms implemented in framework.
- Better documentation.

MonAlisa

- Installation on every node.
- Instalation and tuning of ML Repository.
- Alarms.
- New Metrics.

MonAlisa

dashboard

MonALISA Repository

MonALI
MONitoring Agents us
Integrated Services Arch

Repository Home | Administration Section | Events XML Feed | MonaLisa GUI

Repository

- MonALISA Repository
- Global Views
- Statistics
- Services
- Machines
- Collectors
- Repository info
- Installed packages

close all

This page: bookmark, URL

Dashboard machines' status

What is this a

Machines status

			CPU				Networking (eth0)		Busiest disk			httpd workers		httpd now		httpd instance avg			httpd instance total		
Host	sms state	Online	Load	usr	sys	iowait	idle	IN	OUT	Util (%)	Device	IOPS	Running	Idle	Req/s	Traffic	Req/s	Traffic	Bytes/req	Requests	Traf
dashb-virtual06			-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	
dashboard03	production		0.04	0.667	0.103	0.131	99.08	0.176 Mbps	0.711 Mbps	1.601	sda	5.38	2	98	7.491	0	3.8	46.1 KB/s	12390	107736	1.
dashboard17	production		0.03	1.662	0.75	0.355	97.07	87.4 Kbps	30.91 Kbps	0.997	hda	5.516	1	7	0.017	85.31 B/s	0.04	173 B/s	4342	170	72
dashb-virtual09			-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	
dashb-virtual11			0.16	1.217	2.1	0	96.45	0.824 Kbps	3.868 Kbps	0.042	sda	0.817	-	-	-	-	-	-	-	-	
dashb-virtual04			0.42	13.73	11.01	0	74.33	1.718 Kbps	3.393 Kbps	0.083	sda	3.083	-	-	-	-	-	-	-	-	
dashboard11	maintenance		0.08	0.36	0.2	0.015	99.41	12.53 Kbps	22.2 Kbps	0.12	sda	4.083	-	-	-	-	-	-	-	-	
dashboard21	production		0.4	1.437	0.674	0.345	97.02	0.101 Mbps	0.529 Mbps	0.992	hda	5.549	-	-	-	-	-	-	-	-	
dashboard09	production		0.01	0.965	0.125	0.021	98.86	80.69 Kbps	0.52 Mbps	0.213	sda	3.7	3	72	6.85	0	3.59	57.5 KB/s	16384	102092	1.
dashboard27	maintenance		0.01	0.282	0.218	0.161	99.26	2.349 Kbps	4.166 Kbps	0.43	hda	3.017	-	-	-	-	-	-	-	-	
dashboard31	production		0.32	1.925	0.536	17.41	79.02	95.24 Kbps	91.88 Kbps	19.78	hda	5.699	74	54	-	-	-	-	-	-	
dashboard22	maintenance		0.05	0.33	0.362	0.201	99.07	2.458 Kbps	3.152 Kbps	0.572	hda	2.666	-	-	-	-	-	-	-	-	
dashb-virtual07			-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	
dashboard25	production		0.17	3.052	0.904	0.396	95.17	83.56 Kbps	51.35 Kbps	0.965	hda	8.632	-	-	-	-	-	-	-	-	
dashb-virtual02			-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	
dashboard23	production		0.19	0.128	0.287	9.902	89.64	2.456 Kbps	3.215 Kbps	16.48	hda	2.066	-	-	-	-	-	-	-	-	
dashb-virtual05			-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	
dashboard02	production		0.01	0.069	0.184	0.127	99.57	3.987 Kbps	4.478 Kbps	1.427	sda	6.3	-	-	-	-	-	-	-	-	

Other applications

- Dashboard for Google Earth.
- SiteView.

Presentations

Group meetings presentation:

- jQuery.
- Charting.
- Deployment and load balancing.
- noSQL.
- Architecture of JS applications.

Twiki

Twiki articles:

- JS tools and libraries (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/Libs).
- MVC architecture (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/ModelViewController).
- Dashboard services documentation (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/Services).
- MonAlisa installation procedure (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/MonAlisa).
- Authentication mechanism in dashboard framework (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/Auth).
- Form handling (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/FormHandling).
- Google Earth emergency (https://twiki.cern.ch/twiki/bin/view/ArdaGrid/DashbEarth).

Other

- Contributing to CHEP papers.
- Attending to Daily Ops.
- Attending to CMS Ops.
- Summer Student.

What I learned

- Java Script technologies.
- Dashboard Framework.
- CERN School of Computing.
- Sys Admin stuff.
- Many interesting lectures (including those by Richard Stallman and James Watson).
- French language course.
- Working in multinational environment.
- Working in large organisation.
- Living abroad.
- Faster than light neutrino.

Horus.pl

Development of business applications intended for use by corporate clients:

Orange

T-mobile

Play

Netia

Horus Workflow

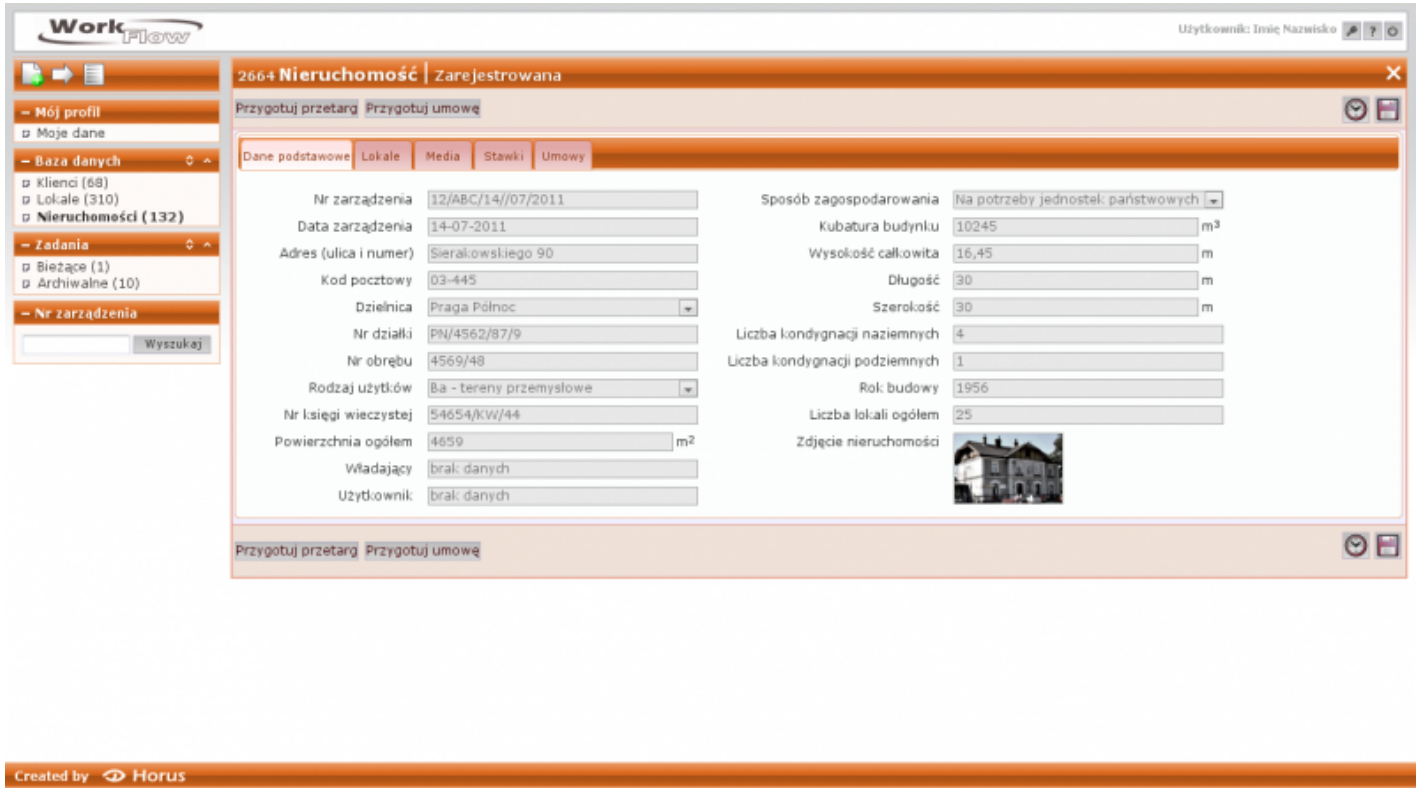
Horus Workflow

Horus Workflow is used to define and monitor workflow in business processes. It supports the implementation of any number of administrative processes, personnel, management or sales.

Horus Workflow System Features:

- Support for managing tasks
- The ability to define own processes
- Support for document management processes
- Support for a variety of organizational structures
- Monitoring of user activity (change history)
- Management of the company organizational structure

Horus Workflow – application screenshot



Horus Workflow – technologies

Used technologies and libraries:

- Spring
- Maven
- JBoss
- Hudson / Jenkins
- Coffee Script
- JQuery UI

TMS Brokers Brokerage House

TMS Brokers
DOM MAKLERSKI

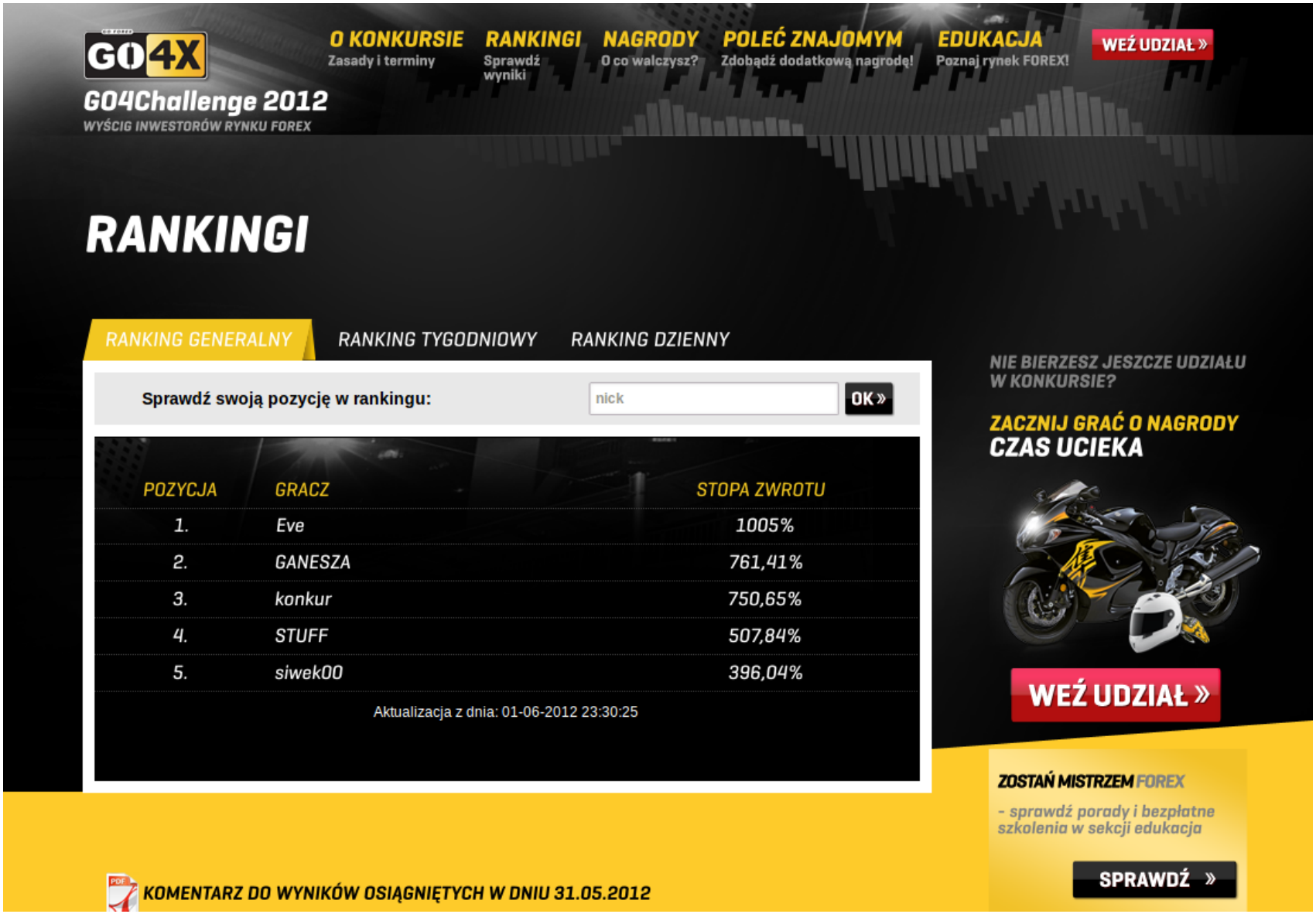
Tasks and responsibilities:

- Development of financial reporting software
- Supporting promotional campaigns
- MetaTrader API programming

TMS Brokers – technologies

Used technologies and libraries:

- JQuery UI
- Highcharts and Highstock
- Python
- Django
- C++


GO4X
O KONKURSIE
Zasady i terminy
RANKINGI
Sprawdź wyniki
NAGRODY
O co walczysz?
POLEĆ ZNAJOMYM
Zdobądź dodatkową nagrodę!
EDUKACJA
Poznaj rynek FOREX!
WEŹ UDZIAŁ »
GO4Challenge 2012
WYŚCIG INWESTORÓW RYNKU FOREX
RANKINGI
RANKING GENERALNY
RANKING TYGODNIOWY
RANKING DZIENNY
Sprawdź swoją pozycję w rankingu:
nick
OK »
POZYCJA
GRACZ
STOPA ZWROTU
1. Eve 1005%
2. GANESZA 761,41%
3. konkur 750,65%
4. STUFF 507,84%
5. siwek00 396,04%
Aktualizacja z dnia: 01-06-2012 23:30:25
NIE BIERZESZ JESZCZE UDZIAŁU W KONKURSIE?
ZACZNIJ GRAĆ O NAGRODY
CZAS UCIEKA
WEŹ UDZIAŁ »
ZOSTAŃ MISTRZEM FOREX
- sprawdź porady i bezpłatne szkolenia w sekcji edukacja
SPRAWDŹ »
KOMENTARZ DO WYNIKÓW OSIĄGNIĘTYCH W DNIU 31.05.2012

Github

Source

LaTeX source of this presentation can be downloaded from github:

```
git://github.com/mnowotka/Chembl-job-web.git
```

Thank you for your attention.