

# Background in Web Development

Michał Nowotka  
job applicant



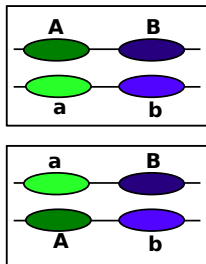
EMBL-EBI  
ChEMBL group

June 8, 2012

- 1 Experience in research
  - Bachelor thesis
  - Master thesis
- 2 Experience at CERN
- 3 Recent experience and current work

# 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.

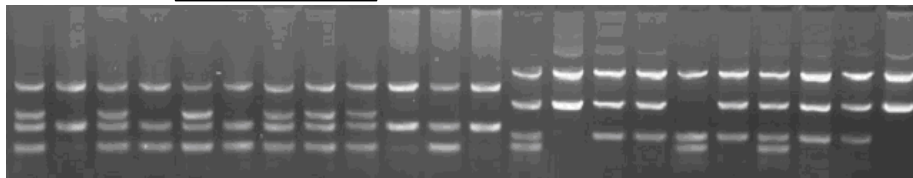


Figure:  
Determining genotype experiment results

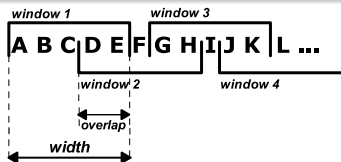
# Idea of short overlapping window

## 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.

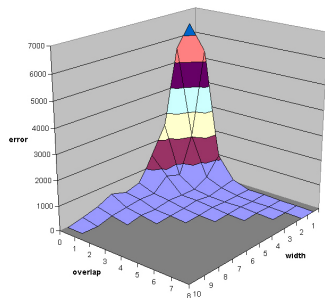
## Solution – Genotypes can be divided into shorter pieces that overlap.

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

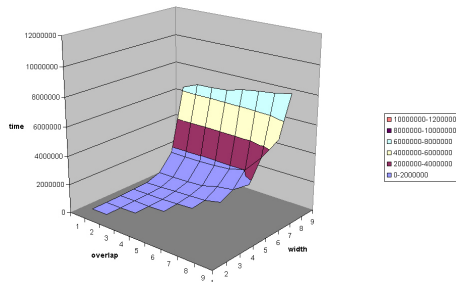


What are the error and execution time as a function of **width** i **overlap** parameters?

# Results

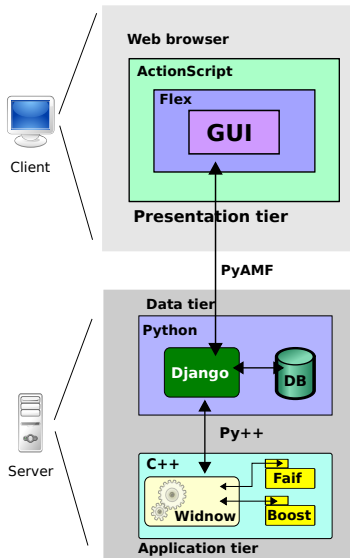


Error as a function of width and overlap parameters



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 availability of 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.
- Standardized 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

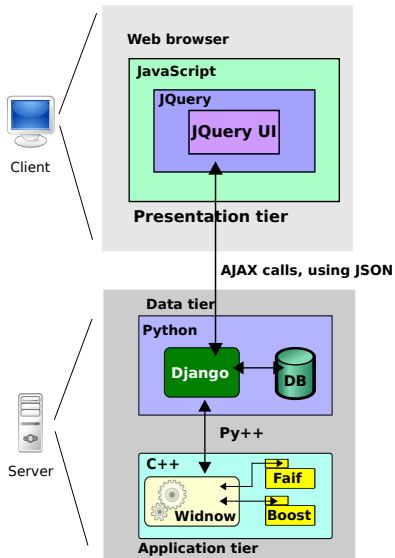


# Inferring functional annotation

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 behavior 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




# Site Status for the CMS sites

Read only: **PRP**

Index	Expanded Table	Gridmap	Alternative views
<p>Click on box on the left to select the CMS site</p> <p>Click on the second column to download the CMS</p> <p>Click on any of the selected to download the CMS</p>			
PRP		PRP	
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
<p>TO_CMS_BOARD</p>			
PRP		PRP	
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>
 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>	 <a href="#">TO_CMS_BOARD</a>

[Help](#)
[Login](#)

Site Status for the CMS Sites v0.13.0 rc1



[Index](#)
[Expanded Table](#)

View: [default](#)


T0 +T1

Status	Site Name
	T0_CH_CERN
	T1_CH_CERN
	T1_DE_KIT
	T1_ES_PIC
	T1_FR_CG2P3
	T1_IT_CNAF
	T1_TW_ASIC
	T1_UK_RAL
	T1_UK_RAL_Disk
	T1_US_FNAL

T2

Status	Site Name
	T2_AT_Vienna
	T2_BE_IJHE
	T2_BE_UCL
	T2_BR_SFRAGE
	T2_BR_UERJ
	T2_CH_CSCS
	T2_CN_Beijing
	T2_DE_DESY
	T2_DE_RWTH
	T2_EE_Estonia
	T2_ES_CIEMAT
	T2_FR_IPHC
	T2_HU_Budapest
	T2_IN_TIFR
	T2_IT_Bari
	T2_IT_Legnaro
	T2_IT_Pisa
	T2_IT_Rome
	T2_KR_KNU
	T2_PL_Warsaw
	T2_PT_UIP_Lisbon
	T2_PT_INOQ_Lisbon
	T2_RU_RRC_KI
	T2_RU_SINP
	T2_TR_METU
	T2_TW_Taiwan
	T2_UA_KIPT
	T2_UK_London_Brunel
	T2_UK_London_IC
	T2_UK_Slrd_RALPP
	T2_US_Catech
	T2_US_Florida
	T2_US_MIT

# Old and new SSB


[oddsboard](#)
 Site Status for the CMS sites
 [Found a bug?](#)
[HELP](#)

Index	Expanded Table	Gridmap	Alternative views
-------	----------------	---------	-------------------

Put the mouse over any column header to get the description of the column.  
Clicking on a column header will display the evolution of that column over the last 24 hours.  
 information is more than 24h old

[illegible]

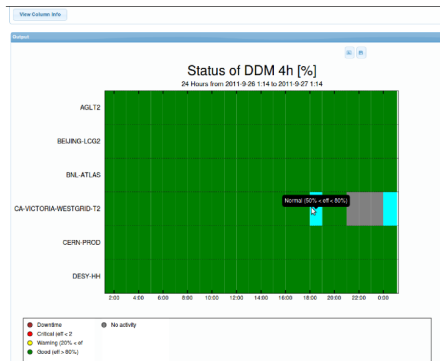
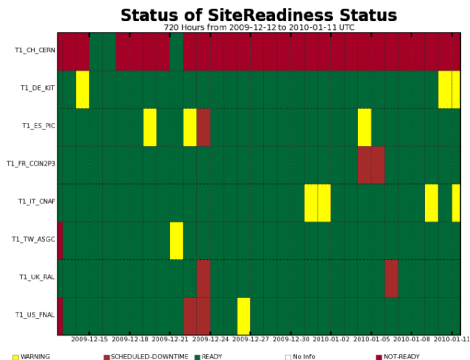
The screenshot shows the CMS dashboard with a table of CMS jobs. The table has columns for Site Name, Visible, JobRobot, SAM TESTS (CE, SRM), Production, Analysis, Site usage (Running, Pending), and Commissioned Links. The table lists various CMS jobs and their status, including 'T0\_CH\_CERN', 'T1\_CH\_CERN', 'T1\_DE\_KEK', 'T1\_ES\_PIC', 'T1\_FR\_COMPS', 'T1\_IT\_CNAF', 'T1\_TW\_ARGC', 'T1\_UK\_RAL', 'T1\_UK\_RAL Disk', and 'T1\_US\_FNAL'. The 'Running' column shows the number of running CMS jobs, and the 'Pending' column shows the number of pending CMS jobs. The 'Commissioned Links' column shows the status of the links, with 'OK' in green and 'S3 combined' in red.

Site Name	Visible	JobRobot	SAM TESTS		Production	Analysis	Site usage		Commissioned Links
			CE	SRM			Running	Pending	(expand this column)
T0_CH_CERN	OK	n/a	OK	OK	100%	n/a	n/a	n/a	current number of running CMS jobs: Rmp: null Desc: null last update: 2011-09-24 22:05:00 UTC last value: 2011-09-26 21:42:55 UTC
T1_CH_CERN	n/a	100%(600)	n/a	n/a	n/a	n/a	n/a	n/a	
T1_DE_KEK	OK	100%(600)	OK	OK	100%(2033)	100%(3)	413	n/a	
T1_ES_PIC	OK	96%(600)	OK	OK	100%(1553)	n/a	12	1	
T1_FR_COMPS	OK	n/a	OK	OK	100%(5820)	n/a	1056	589	
T1_IT_CNAF	OK	100%(600)	OK	OK	100%(7235)	n/a	86	14	S3 combined
T1_TW_ARGC	OK	100%(600)	OK	OK	100%(2995)	100%(79)	8	347	S3 combined
T1_UK_RAL	OK	100%(600)	OK	OK	100%(1850)	n/a	328	2	S3 combined
T1_UK_RAL Disk	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	S3 combined
T1_US_FNAL	OK	100%(1000)	OK	OK	100%(71477)	100%(3403)	n/a	n/a	S3 combined

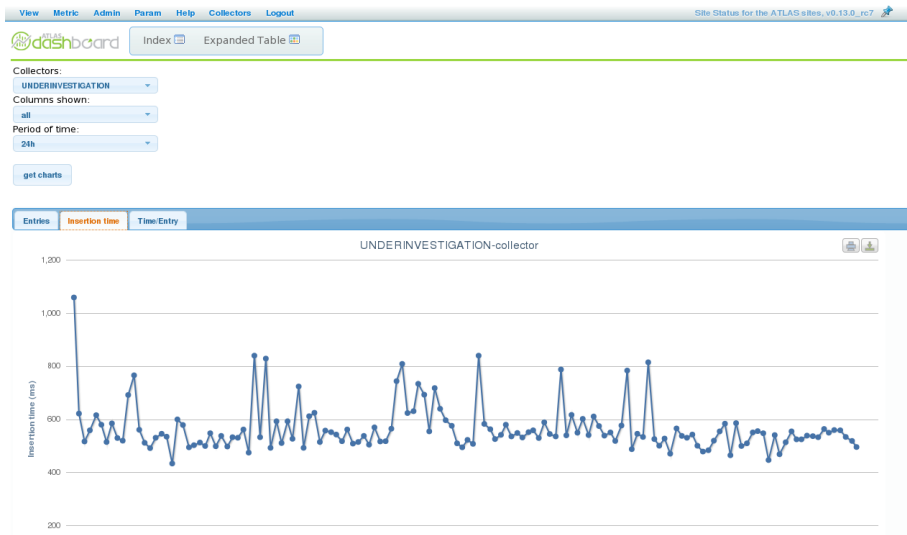
Showing 1 to 10 of 115 entries. DB query took 0.0276 s

First Previous 1 2 3 4 5 Next Last

# Old and new SSB



# SSB – collector metainformation



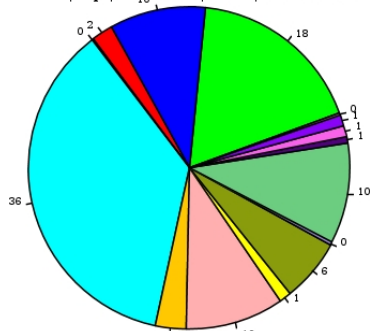


# SSB – TODO

- 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

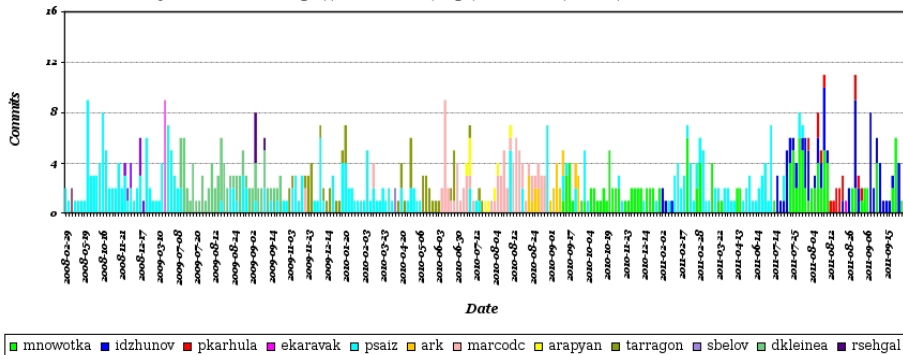
<https://svn.cern.ch/repos/dashboard/trunk/arda.dashboard.siteview>



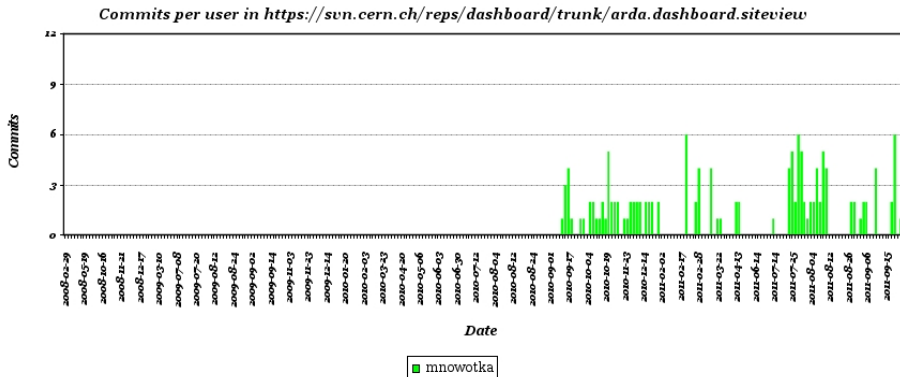
a ■ idzhunov ■ pkarhula ■ ekaravak ■ psaiz ■ ark ■ marcodc ■ arapyan ■ tarragon ■ sbelow ■ dkleinea ■ rsehgal ■ elisal ■ sic

# SSB – Commits

Commits for all users in <https://svn.cern.ch/repos/dashboard/trunk/arda.dashboard.siteview>



# 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



## MonALISA Repository



[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 s

### Machines status

		CPU				Networking (eth0)			Busiest disk			httpd workers		httpd now		httpd instance avg			httpd instance total			
Host	sms state	Online	Load	usr	sys	lowlat	idle	IN	OUT	Util (%)	Device	IOPS	Running	Idle	Req/s	Traffic	Req/s	Traffic	Bytes/req	Requests	Tml	
dashb-virtual06			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
dashb-board03	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.	
dashb-board17	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	-	-	-	-	-	-	-	-	-	-
dashb-board11	maintenance		0.08	0.36	0.2	0.015	99.41	12.53 Kbps	22.2 Kbps	0.12	sda	4.083	-	-	-	-	-	-	-	-	-	-
dashb-board21	production		0.4	1.437	0.674	0.345	97.02	0.101 Mbps	0.529 Mbps	0.992	hda	5.549	-	-	-	-	-	-	-	-	-	-
dashb-board09	production		0.01	0.965	0.125	0.021	98.88	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.	
dashb-board27	maintenance		0.01	0.282	0.218	0.161	99.26	2.349 Kbps	4.166 Kbps	0.43	hda	3.017	-	-	-	-	-	-	-	-	-	-
dashb-board31	production		0.32	1.925	0.536	17.41	79.02	95.24 Kbps	91.88 Kbps	19.78	hda	5.699	74	54	-	-	-	-	-	-	-	-
dashb-board22	maintenance		0.05	0.33	0.362	0.201	99.07	2.458 Kbps	3.152 Kbps	0.572	hda	2.666	-	-	-	-	-	-	-	-	-	-
dashb-virtual07			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
dashb-board25	production		0.17	3.052	0.904	0.396	95.17	83.56 Kbps	51.35 Kbps	0.965	hda	9.632	-	-	-	-	-	-	-	-	-	-
dashb-virtual02			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
dashb-board23	production		0.19	0.128	0.287	9.902	89.64	2.456 Kbps	3.215 Kbps	16.48	hda	2.066	-	-	-	-	-	-	-	-	-	-
dashb-virtual05			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
dashb-board02	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 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 course.
- Working in multinational environment.
- Working in large organisation.
- Living abroad.
- Faster than light neutrino.



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

**WorkFlow** Użytkownik: Janusz Nawrocki

**266+ Nieruchomość | Zarejestrowana**

Przygotuj przetarg Przygotuj umowę

**Dane podstawowe Lokale Media Stawki Umowy**

Nr zarządzenia	12/48C/14/07/2011	Sposób zagospodarowania	Na potrzeby jednostek państwowych
Data zarządzenia	14-07-2011	Kubatura budynku	10245 m <sup>3</sup>
Adres (ulica i numer)	Sierakowskiego 90	Wysokość całkowita	16,45 m
Kod pocztowy	03-445	Długość	30 m
Dzielnica	Praga-Północ	Szerokość	30 m
Nr działki	216/4562/87/9	Liczba kondygnacji naziemnych	4
Nr obrębu	4569/48	Liczba kondygnacji podziemnych	1
Rodzaj użytku	Ba - tereny przemysłowe	Rok budowy	1956
Nr księgi wieczystej	54654/K.VV/44	Liczba lokali ogółem	25
Powierzchnia ogółem	4659 m <sup>2</sup>	Zdjęcie nieruchomości	
Władający	brak danych		
Użytkownik	brak danych		

Przygotuj przetarg Przygotuj umowę

# Horus Workflow – technologies

Used technologies and libraries:

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



# TMS Brokers Brokerage House



## 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++



**GO4Challenge 2012**  
WYŚCIG INWESTORÓW RYNKU FOREX

**O KONKURSY**  
Zasady i terminy

**RANKINGI**  
Sprawdź  
wyniki

**NAGRODY**  
O co walczysz?

**POLEĆ ZNAJOMYM**  
Zdobądź dodatkową nagrodę!

**EDUKACJA**  
Poznaj rynek FOREX!

**WEŹ UDZIAŁ »**

## RANKINGI

RANKING GENERALNY

RANKING TYGODNIOWY

RANKING DZIENNY

Sprawdź swoją pozycję w rankingu:

**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 KONKURSYE?

**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

L<sup>A</sup>T<sub>E</sub>X source of this presentation can be downloaded from github:

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

# Thank you for your attention.