

Yandex

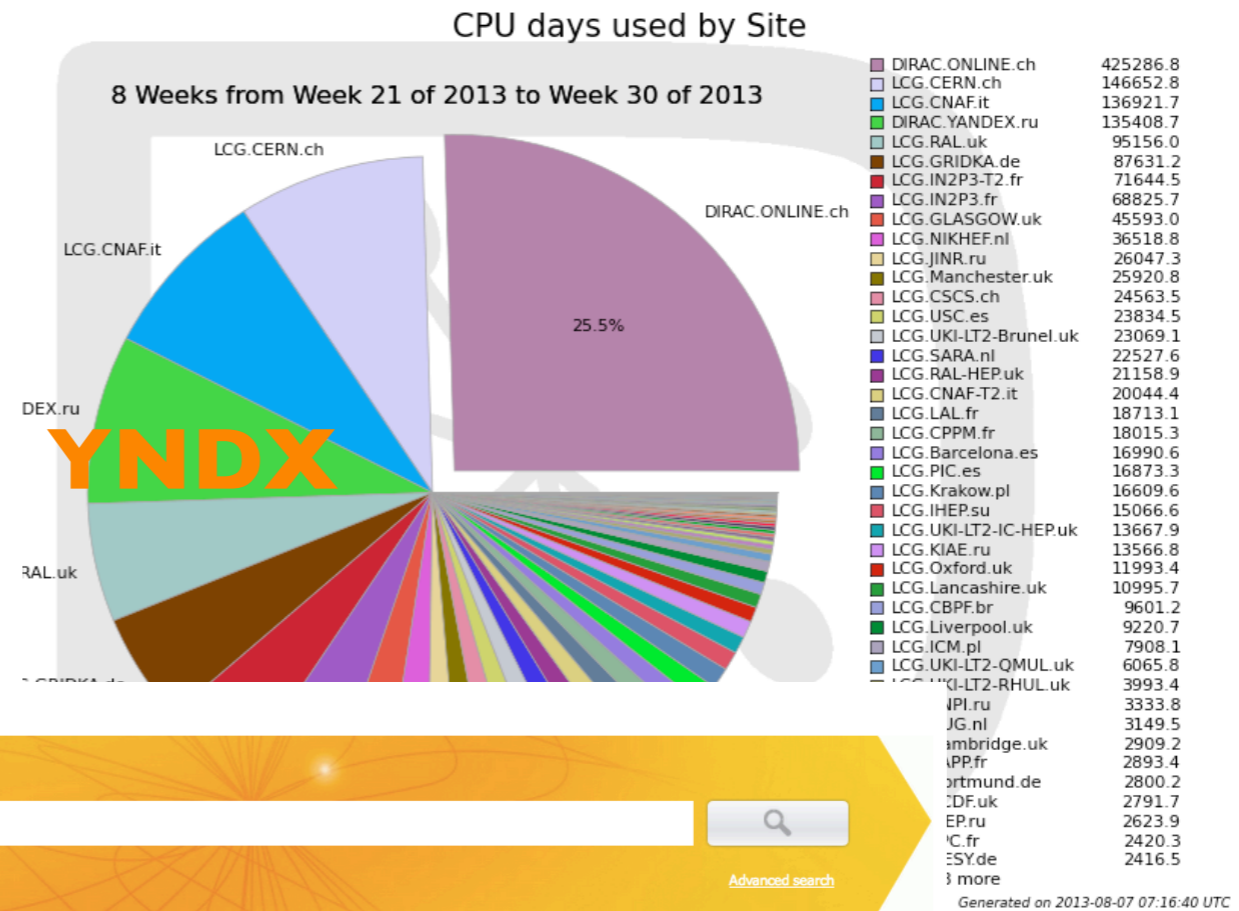


Joint Projects Review

Andrey Ustyuzhanin
Head of CERN-Yandex joint projects

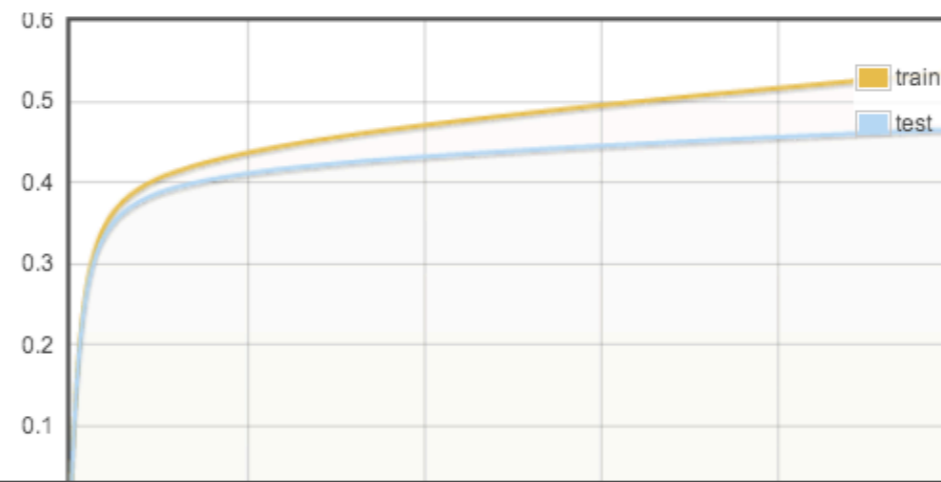
Yandex Services for CERN users

- LHCb Grid Tier 2 Cluster
- Event Index
- Event Filter



Formula «mc_Bu2X-B2psiX_2a»

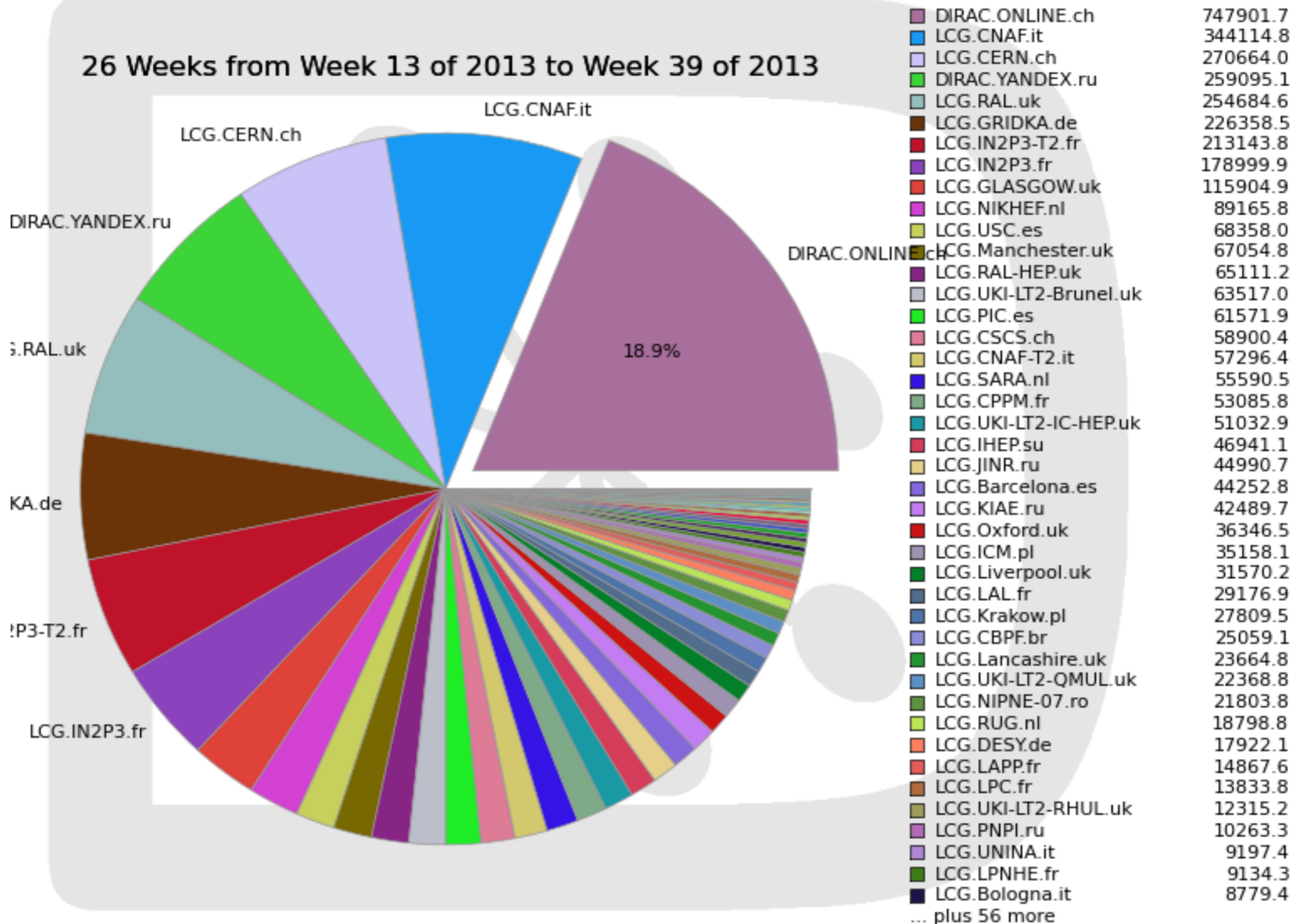
Learning Factors Export



supported by
ndex

CPU days used by Site

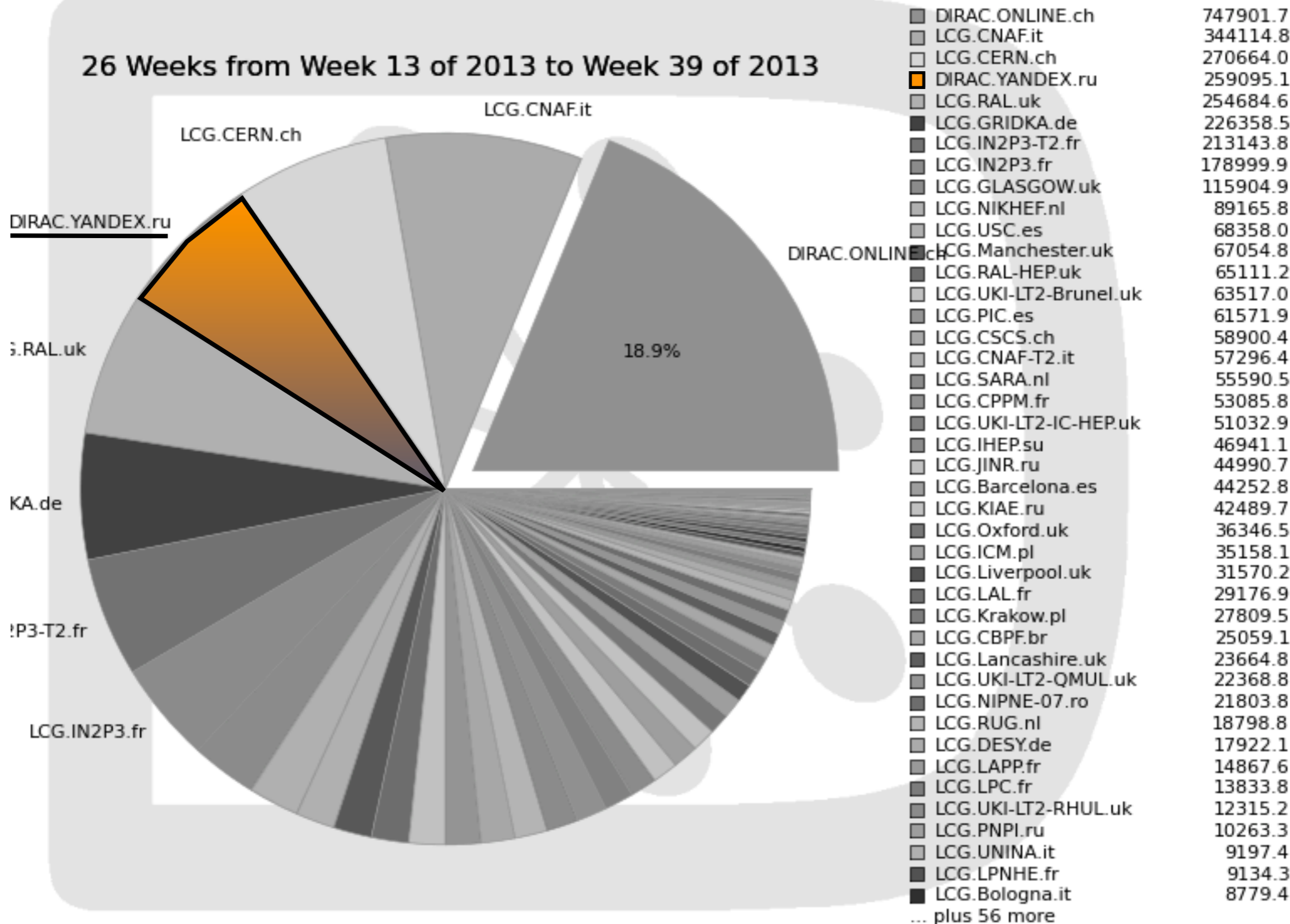
26 Weeks from Week 13 of 2013 to Week 39 of 2013



Generated on 2013-09-25 18:19:10 UTC

CPU days used by Site

26 Weeks from Week 13 of 2013 to Week 39 of 2013



Generated on 2013-09-25 18:19:10 UTC

Event Index



Current version runs at <http://volhcb16.cern.ch:9152/>
New version will be available at <https://volhcb02.cern.ch/>



Search is supported by
Yandex

Event Filter (<https://volhcb16.cern.ch/>)

Home	Datasets	Pools	Formulas	Predictions	Tasks	About				
BDT9_Flavio1_i10000_w001_x64	Flavio1	B_s0_TAU_ps, mu_MINIPS, ...	336903	1	2013-04-19 10:12:49	-0.0543496	↓ Expo	× Remo		
D0mumu-test							45:23	1.46709	↓ Expo	× Remo
default							34:52	0.243071	↓ Expo	× Remo
form_1							00:32	-0.000751594	↓ Expo	× Remo
marco_new_test							18:07	0.3142	↓ Expo	× Remo
mc_Bu2X-B2psiX_2							43:10	0.607015	↓ Expo	× Remo
mc_Bu2X-B2psiX_2a							53:19	0.46621	↓ Expo	× Remo
merger_3							37:05	0.0453774	↓ Expo	× Remo
MN_13_Flavio1_i10000_w001_)							25:23	-0.382667	↓ Expo	× Remo
MN_13_Flavio1_i10000_w001_)							12:24	0	↓ Expo	× Remo

Formula «mc_Bu2X-B2psiX_2a»

Learning Factors Export

Time (s)	Train Performance	Test Performance
0	0.0	0.0
1000	0.45	0.40
2000	0.48	0.43
3000	0.50	0.45
4000	0.52	0.46

Event Filter HTTP API

/api/upload_dataset

/api/train_formula

/ajax/tasks_status

/api/get_trained_formula_binary

/api/remove_dataset

...

more info at: <https://volhcb16.cern.ch/static/API.pdf>

Example:

```
anaderi@lxplus:~$ curl --digest http://volhcb16.cern.ch/api/find_pool?id=22
{
  "data": {
    "fields":
    "B_DIRA_OWNPV,B_ENDVERTEX_CHI2,B_FDCHI2_OWNPV,B_IPCHI2_OWNPV,B_PT,D_ENDVERTEX_
    CHI2,is_signal,Jpsi_ENDVERTEX_CHI2,Jpsi_PT,K_MINIPCHI2,K_PT,muminus_MINIPCHI2,
    muplus_MINIPCHI2,pi_MINIPCHI2,pi_PT",
    "fields_count": "15",
    "id": "22",
    "name": "D0mumu_a",
    "rows_count": "53239",
    "size_on_disk": "10596801",
    "type": "CSV",
```


Event Filter API Python wrapper

Trivial Example of API usage

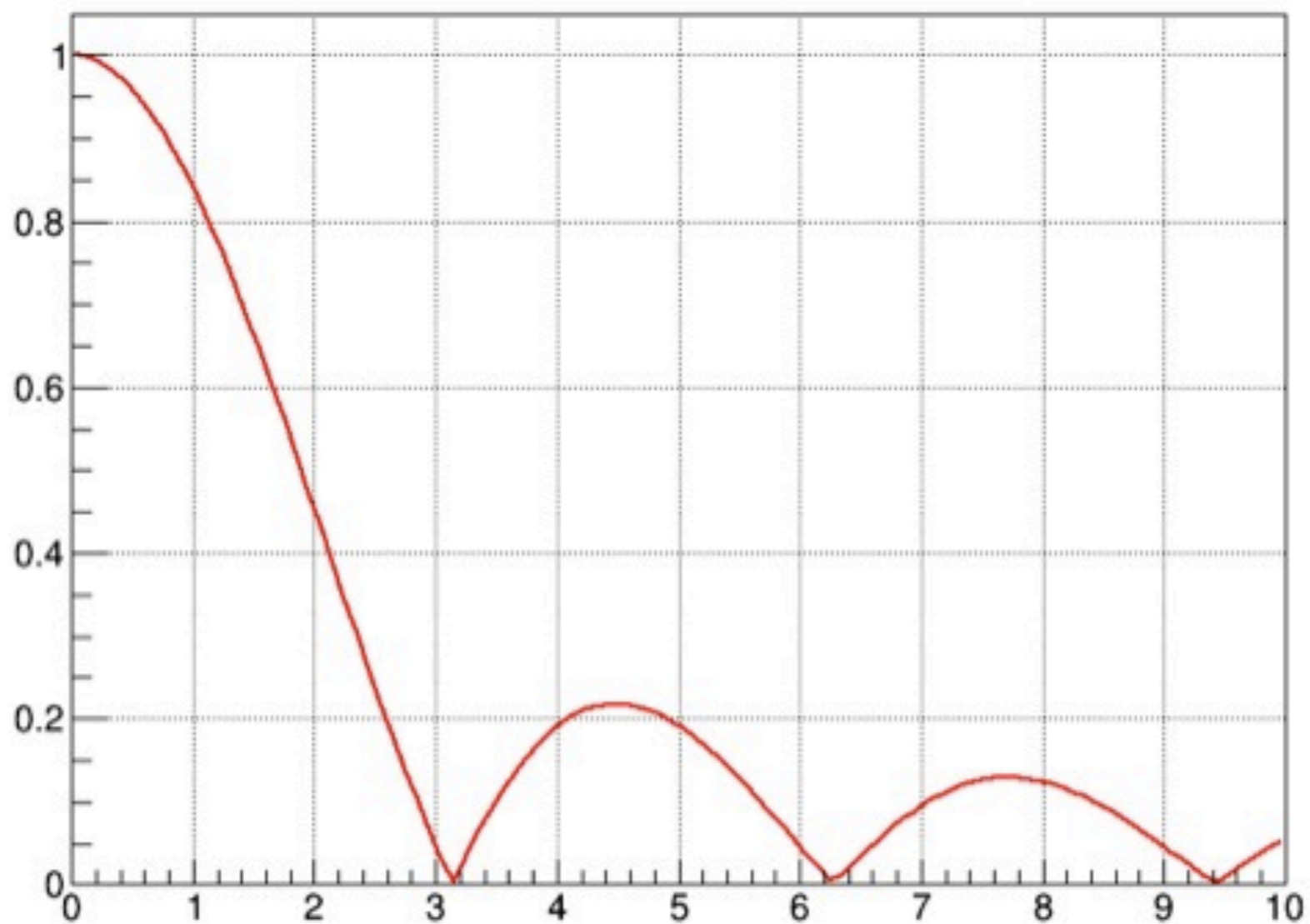
```
import Yandex_EF_API
try:
    api = Yandex_EF_API(url_base=URL, user='', password='')
    result = api.upload_dataset(train_filename, TEST_NAME)
    sleep(60) # ... or check it is ready
    result = api.train_formula(pool_id, TEST_NAME,
                               fields=[
                                   'B_DIRA_OWNPV',
                                   'B_ENDVERTEX_CHI2',
                                   'B_FDCHI2_OWNPV' ],
                               train_rows=1000)
    result = api.download_formula(
        formula_file_csv,
        formula_name=TEST_NAME,
        file_type='csv')
except Exception, e:
    print "Oops!", e
```

Example

In [2]:

```
from ROOT import TF1
import rootnotes
c1 = rootnotes.default_canvas()
fun1 = TF1( 'fun1', 'abs(sin(x)/x)', 0, 10)
c1.SetGridx()
c1.SetGridy()
fun1.Draw()
c1
```

abs(sin(x)/x)



Out[2]:

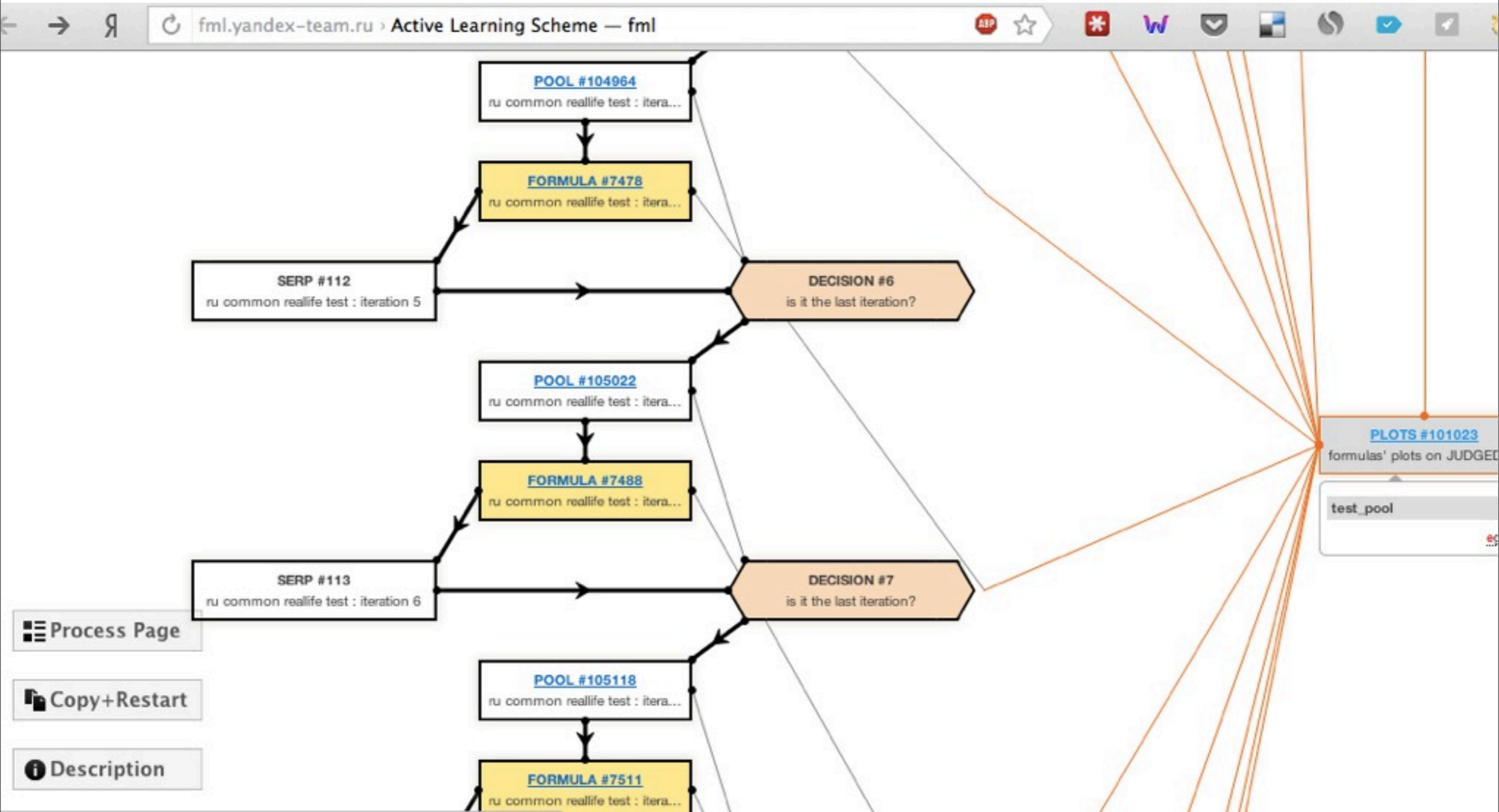
<http://bit.ly/I8svtRP>

Services Roadmap

Version	Date	Event Index Features
0.5	September 2013	Index whole Collision12
1.0	November 2013	Quick indexed events analysis
2.0	2014	Index data 2010-2012

Version	Date	Event Filter Features
0.5	September 2013	API + Python wrappers
1.0	November 2013	Different cost functions selection Baseline training & Cross validation
2.0	2014	Graphical research pipeline interface

Yandex Data Factory



Яндекс

www.yandex.ru

fmi filter by factor name or index

today 06:20 **US formula reproduction** by egoist
 Waiting for approve -- Experiment
 started EvalTask #14030 «FlashRule, isGameClassify, isGameRegex, isGameUrlRegex»
 started Pool #81053

yesterday 13:23 **created Православный статик ранк для Турции** factors TG_STATIC | TG_DOC by egoist
 failed Formula #666 «50K 7b 7g VarAmp7.w012 /10000»
 started Pool #81053

14.01.13 **Request Popularity** by kvn
 23:21 Sucks -- Waiting for approve

14.01.13 **Матчинг запросов с запросами, по которым кликали на данный док...** by egoist
 completed Formula #4559 «simulate b7g7.20130111-16-54.VarAmpImp7»
 completed EvalTask #13736 «HostAvgUrlTime»

14.01.13 **UriBeBack** by egoist
 started Formula #4559

14.01.13 **Search base fuck up**
 12:24 Bad indexrefmap -- Pool Collection affected

14.01.13 **Переходы с wikipedia** by egoist
 canceled EvalTask #113355 «SR»

14.01.13 **HostWordsFrequency** by egoist
 canceled Formula #4559
 completed EvalTask #13736 «HostAvgUrlTime»
 failed Pool #81052 «VarAmpImp7»
 added Pool #81051 «simulate b7g7.20130111-16-54.VarAmpImp7»
 failed Pool #81050 «qwe»

14.01.13 **Links WordHost** by egoist
 23:00 Success -- implementation verification

14.01.13 **Production Formula Pipeline**
 12:24 Formula #221 -- ru central

New 34
 My experiments
 Failed
 New experiment

Yandex

Andrey Ustyuzhanin

Head of CERN-Yandex joint projects

anaderi@yandex-team.ru

Thank you