

MEX Project

MyExp Ontology

AKSW 2015

What is the context?

**Machine Learning and Statistical
Experiments**

What are the topics?

- (1) Publishing scientific experiment results**
- (2) Understanding of data/results and general architecture**
- (3) Comparing public results**
- (4) Manage experimental data**

What are the gaps into the topics?

Topic	Gaps
Publishing Scientific Experiment Results	There are no guidelines to publish experimental data
Understanding of Data/Results And General Architecture	Hard to understand experimental procedure or find out values for specific ML parameters, for instance.
Comparing Public Results	There is no way to easily compare 2 or more ML/Statistical similar experiments
Manage Experimental Data	Too hard to manage results, even more during training phase. Tendency to have too many data

How to tackle with the gap (Item 1)?

(1) Publishing Scientific Experiment Results

How to tackle with the gap (Item 1)?

- **There is no guideline**

How to tackle with the gap (Item 2)?

**(2) Understanding of Data/Results And
General Architecture**

How to tackle with the gap (Item 2)?

- **really hard to understand people's work**
- **no metadata**
- **no common-format**
- **sometimes just the numbers (alone) do not mean that you've achieved really good results**

How to tackle with the gap (Item 3)?

(3) Comparing Public Results

How to tackle with the gap (Item 3)?

- **most of the time you have to look the paper and try to describe the experiment by yourself**
- **takes awhile to get it**

How to tackle with the gap (Item 4)?

(4) Manage Experimental Data

How to tackle with the gap (Item 4)? meanwhile...

- **Normally a lot of data are generated, derived by the scientific developing process**
- **people tend to create specific formats (txt, csv) without metadata**

How to tackle with the gap (Item 4)?

as result...

- 4190 txt files are created (for instance)
- It is hard to compare possible group of experiments
- Of course it can be solved by implementing some database and just look up for (query). However, still have the gap to share....always need more and more hands on work

How to tackle with the gap (Item 4)? as result...

NORMALLY JUST YOU

(and maybe your friend...)

UNDERSTAND THE EXPERIMENT RESULTS

(...for few days!)



And We Have a Plus!

Many people just write papers trying to publish experimental data having basic mistakes!

We need a automatized way to check it beforehand and generate alerts...

SmartCheck System

And We Have another one!

I really would like to have a repository for look for well define challenges and its numbers (like POS tag problems)....

...We need a nice and friendly user interface

Gerbil !

Just one more for you!

**Ok, you have your experimental data already...
It's hard to starting over and publish by using
the MyExp Ontology, is it?**

**So, we need a way to import the legacy data
and export as an ontology!**

MyExporter

and we have more ideas...

But we can not tell you right now... :-p

Project Proposal

Conclusions

The idea consists in provides:

- **Ontology for ML and Statistical Experiments**
- **Integration with a Benchmark System (Repository)**
- **Tools to validation and check procedures**
- **Tools to import and export data**

Team

AKSW (Leipzig Universität)

- **Diego Esteves**
- **Dimitris Kontokostas**
- **Jens Lehmann**
- **Sebastian Hellmann**
- **Ricardo Usbeck**
- **Diego Moussallem**

IME (Instituto Militar de Engenharia)

- **Maria Claudia Cavalcanti**
- **Julio Cesar Duarte**
- **Marcus Albert**
- **Antonio Pecli**

UFT (Universidade Federal do Tocantins)

- **Tanilson Dias dos Santos**