

# Hybrid, Funky, and Dynamic Intra-Document Analytics

Jens Dittrich

Patrick Bender

Saarland University  
[infosys.cs.uni-saarland.de](http://infosys.cs.uni-saarland.de)

## ABSTRACT

### 1. INTRODUCTION

### 2. PROBLEMS IN EXPERIMENTAL SECTIONS

access to raw data  
access to code

### 3. SIGNIFICANCE OF THE CONTRIBUTION

1. **Graph Reproducibility.** connected workflow from data production to presentation  
measurements embedded in the pdf
2. **Data Preservation.** data published together with the paper  
what is the “paper” anyway? Shouldn’t we revise our notion of “paper”?
3. **In-Document Data Analysis.** slice and dice; drill-down, roll-up, pivots
4. **Alternate Views on the Data.** log vs linear; table vs graph;  
absolute runtimes vs normalized runtimes

applicable to any research working with data, not just the db community  
long time preservation of the underlying data files  
allows other researchers to easily run different analyses on the same data

### 4. RELATED WORK

SIGMOD reproducibility initiatives and papers (two, right?)  
upcoming Dagstuhl perspectives workshop  
experiments and analysis track and its problems: access to data, access to code  
CIDR 2015 talk

BTW 2015 talk with link to youtube video  
dynamic docs in R  
Pweave <http://mpastell.com/pweave/index.html>

### 5. THE PDBF FRAMEWORK

PATRICK  
core compilation steps (with graphs could use the ones I did for my BTW talk)  
different output format options  
pros and cons of those output formats

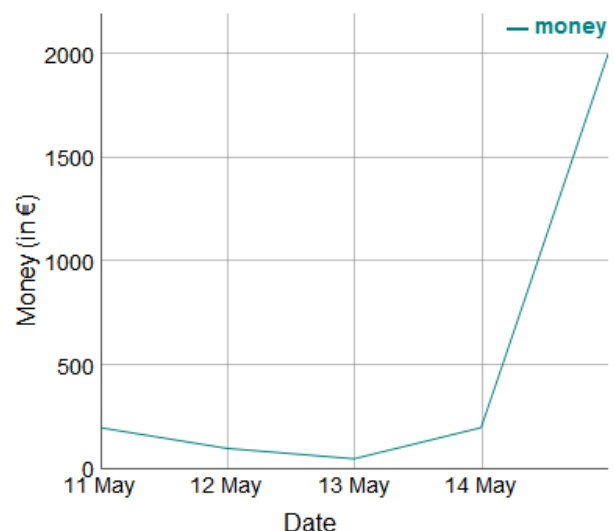
### 6. THE DEMO

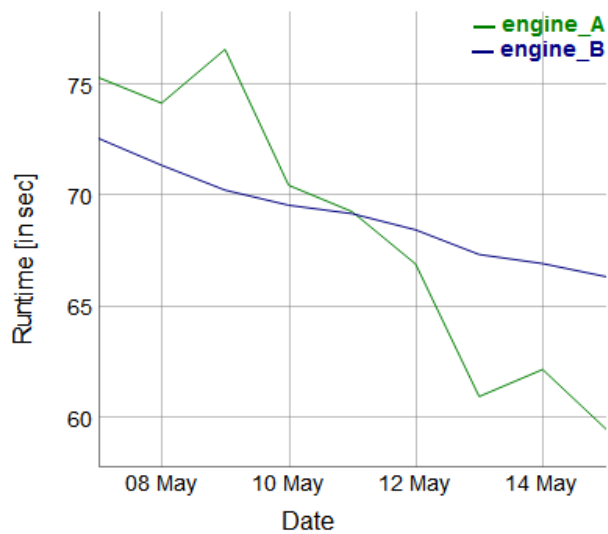
We invite the reader to change the suffix “.pdf” of this submission to “.html”. Your browser will open and you will be able to see the dynamic features. Notice that the dynamic content is completely offline and runs entirely in your Web browser’s sandbox.

PATRICK

#### 6.1 Feature 1...

Switch between data representations for better visualization  
Chart, Pivot Table, Table





### 6.3 Feature 3...

Analyze data in browser Remove filtering functions from graphs  
Look which method performed best under certain condition

### 6.4 Else

## 7. CONCLUSIONS

publish the code?  
website?  
should we mention security aspects of funky?

## 8. REFERENCES

### 6.2 Feature 2...

Get raw data, source code  
raw data can be exported as csv with header  
maybe other file attachments?