

Towards a Catalog of Spreadsheet Smells

Jácome Cunha¹, João P.Fernandes^{1,2}, Hugo Ribeiro¹,

João Saraiva¹

¹ HASLab / INESC TEC & Universidade do Minho,
Portugal

² Universidade do Porto, Portugal

ICCSA 2012

Why do Spreadsheets matter?

Omnipresent
Easy-to-use
Multi-purpose
Flexible

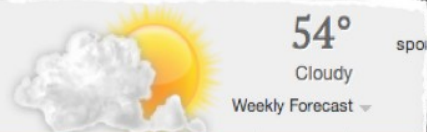
95% of all U.S. firms

90% of all analysts in industry

50% of all SSS are the basis for decisions

Still...

BARABOO NEWS REPUBLIC



News Sports Opinion Obituaries Galleries Jobs CarSoup Homes Rental

W. Baraboo to pay more for borrowed money than believed

Recommend Tweet 5 +1 0

December 09, 2011 8:17 pm • By Brian D. B

Due to a calculating error by their financial advisors, West Baraboo officials learned Thursday they will be paying about \$400,000 more over the lifetime of their most recent 10-year borrowing plan than originally projected.

During its regular December meeting, the West Baraboo Village Board looked back over last month's decision to sell \$1.1 million in general obligation bonds to cover a variety of village projects, said Village Clerk Mary Klingenmeyer. The review was required after the board received a letter from its financial advisory firm, Ehlers of Brookfield.

Ehlers advisor James Mann said "operator error" resulted in a spreadsheet underestimating the total cost of the 10-year bond.

The Salt Lake Tribune | News

Tuesday, April 17, 2012 Last Updated: 01:01 am

News Sports Blogs Opinion Money Lifestyle Entertainment Obituaries Jobs Homes

Utah | Nation + World | Neighborhood | Politics | Justice | Polygamy | LDS Church | Education | Weather | UtahsRight

Salt Lake Tribune Weekly Ad Specials | sears



Utah education officials make big mistake

Utah education officials don't likely be hurt by the resignations of two top finance

First Published Apr 11 2012 11:54 am • Last Updated Apr 11 2012 11:22 pm

A miscalculation at the State Office of Education has led to a \$25 million mistake in Utah's education budget for next school year — and the resignation of two top finance officials.

Education leaders, however, say they don't expect the potential shortfall to hurt schools or districts. State leaders are considering solutions ranging from using education money expected to be left over at the end of this school year to calling a special legislative session.



Photos

Join the Discussion
Post a Comment

"We committed to fund [enrollment] growth and this is an important part of growth," said Senate Budget Chairman Lyle Hillyard on Wednesday. "We would hope to get it fixed, and I think that's going to be our first priority."

The \$25 million represents less than 1 percent of the state's overall \$3 billion-plus education budget.

The problem was that the state office essentially underestimated the number of students expected in schools next school year. The correct number will cost the state \$25

million more than anticipated.

State Superintendent Larry Shumway attributed the mistake to "a faulty reference" in a spreadsheet. He emphasized that no money was misappropriated. He called the mistake "significant" but "manageable."

Economy losses of \$10 billion/year!

And, there are Security Issues...

Jornal de Notícias



INÍCIO | MULTIMÉDIA | BLOGUES | TENDÊNCIAS | OPINIÃO | DOSSIÉS | CIDADÃO REPÓRTER | SERVIÇOS

Últimas | Política | Sociedade | Segurança | Economia | País | Mundo | Desporto | Cultura | Gente

Inquérito da PGR atinge jornalistas

Publicado em 2006-02-16

CARLOS VARELA *

Like Be the first of your friends to like this.

Share 0 Tweet 0 LinkedIn Share 0 +1

Magistrados judiciais e elementos da Polícia Judiciária entraram ontem de rompante na redacção do jornal "24 Horas", em Lisboa, numa operação-surpresa para apreensão de material que o Ministério Público considera associado ao caso do "envelope 9" e ao processo "Casa Pia", que deu origem a uma notícia publicada há mais de um mês naquele diário. A PJ realizou também uma busca, em Portalegre, à casa do jornalista free-lancer co-autor do trabalho.

Na notícia, publicada a 13 de Janeiro, o jornal dava conta da existência de um conjunto de registos telefónicos de altas figuras do Estado, entre as quais Jorge Sampaio e Souto Moura, que surge agregado ao processo Casa Pia, com base num ficheiro Excel enviado pela PT, englobado no famoso "envelope 9".

FERRAMENTAS

- Comentar
- Imprimir
- Distribuir
- Enviar
- A+ Aumentar
- A- Diminuir

ESTATÍSTICAS

Patrocínio



Many more horror stories...

EuSPRIG ORIGINAL HORROR STORIES

The stories are listed in reverse chronological order.

089) Which way around do you calculate a percentage change?

<http://www.startribune.com/587/story/367953.html>

Making a political plus of math error, Star Tribune, 12 Apr 06

DFLer Rebecca Otto is accusing Auditor Pat Anderson, a Republican, of sloppy work. In a column reporting the percentage changes in unreserved fund balances from 2003 to 2004, instead of dividing the difference by the 2003 figure, the auditor's office divided it by the 2004 figure. Deputy Auditor Tony Sutton said "The researcher who worked on that report just made a mistake in the formula in the spreadsheet. He feels bad about it."

Risk: *Giving political opponents an opportunity to comment adversely on the state auditor*

Avoidance: *Check not just the data but the formulas too.*

Advice: [Register Now! for the EuSpRIG 2012 Conference](#)

088) Scoring an own goal

http://www.channelregister.co.uk/2006/03/10/ogc_spreadsheet_snafu/

The Register (UK) OGC spreadsheet madness 10 Mar 06

The Office of Government Commerce is blaming a spreadsheet error for a foul-up over accrediting suppliers for its new Catalyst procurement programme. In a letter to suppliers, they confessed "Unfortunately, [we are] not, as we had hoped, in a position to accept your tender at this time. This is because an error in the original evaluation spreadsheet has been identified, necessitating rescoring of all tenders for this project... this error has now been corrected and this has caused a small number of changes to the original award decision."

Risk: *inaccurate choice of vendors for large government contracts, legal challenges, bad press*

• EuSpriG Original Horror Stories

• <http://www.eusprig.org/stories.htm>

SpreadSheets as SW Languages

In fact Spreadsheets lack advanced mechanisms found in modern programming languages, namely:

- *Type and Modular Systems*
- *Abstractions*
- *Debuggers/Profilers*
- *Best Practice Rules...*

This paper

- We propose a catalog of bad smells for spreadsheets
- We implement the catalog as the SmellSheet Detective Tool
- We refine the original catalog by executing our tool with the large EUSES spreadsheet repository and by (manually) evaluate the results.

Code Smell

Code Smell, or **Bad Smell**, *as proposed by Martin Fowler*, is a concrete evidence that a piece of sw **may have a problem**.

(usually, a smell is not an error!)

SpreadSheet Smells: Catalog Definition

Based on our personal experiences, we propose an initial catalog of spreadsheet bad smells

Type Smells

- Empty cell
- Patterns

Content Smells

- String distance
- Reference to empty cells

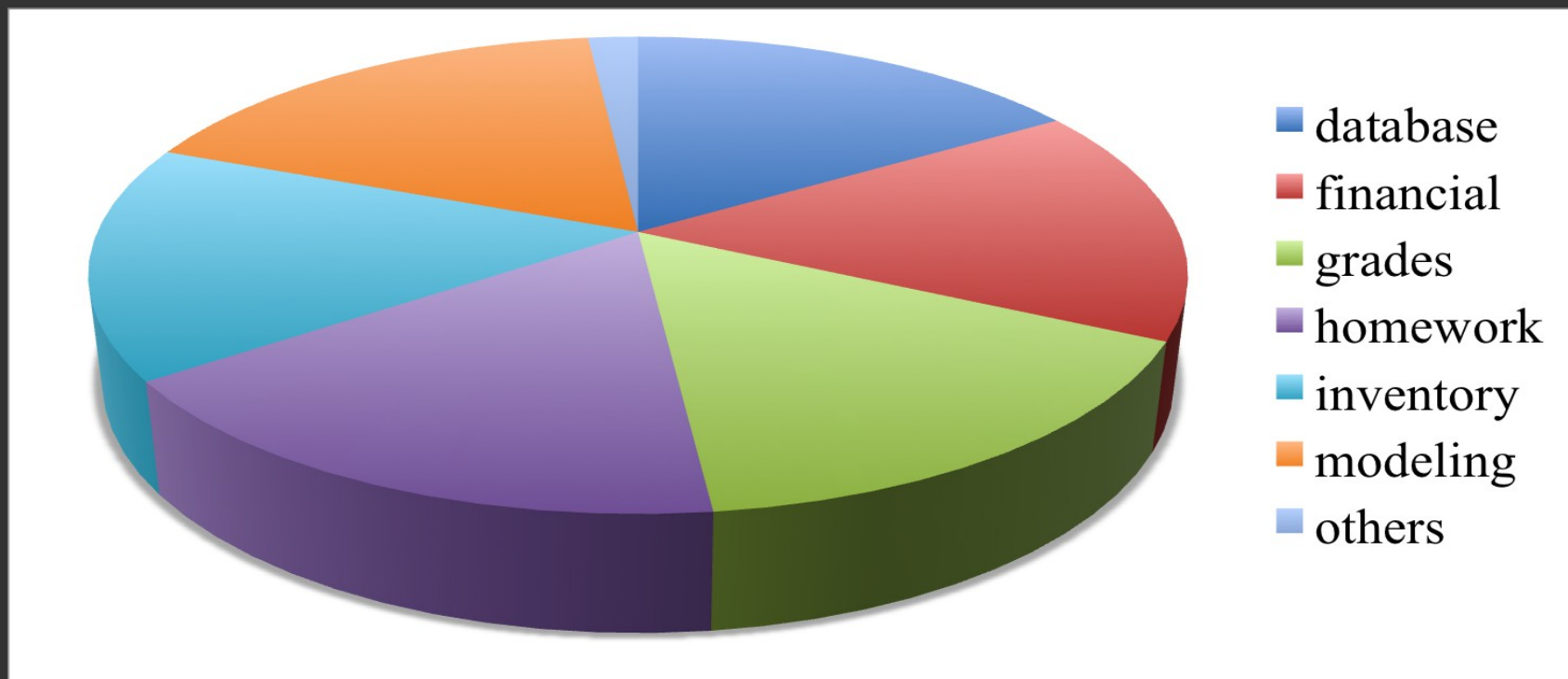
Quasi-Functional Dependencies

Statistical Smells

- Std Deviation

SpreadSheet Smells: Catalog Validation

We validate our initial catalog against real spreadsheets: the EUSES repository (5606 spreadsheets, organized in 6 categories)



SpreadSheet Smells: Catalog Validation

Results obtained from 180 EUSES Spreadsheets

Smell/Category	Database	Financial	Grades	Homework	Inventory	Modeling	Total
Empty cells	53	27	56	42	109	106	393
Patterns	86	62	66	58	111	114	497
Std. Dev.	33	35	134	14	32	7	255
String Dist.	25	3	1357	231	158	658	2442
CFD	12	13	150	2	13	24	204
Ref2empty	0	23	27	0	0	0	50
Total:	209	163	1790	347	423	909	3841

SpreadSheet Smells: Catalog Evaluation

We manually inspect the 180 spreadsheets and empirically classify the computed smells as:

- Low Smells
-
- Medium Smells
-
- High Smells
-
- Not Smells

SpreadSheet Smells: Catalog Evaluation

Smell \ Level	Level				Total
	Low smells	Medium smells	High smells	Not smells	
Empty cells	115	4	0	274	393
Patterns	124	9	0	364	497
Std. Dev. cells	21	0	0	234	255
String Dist.	13	2	7	2420	2442
QFD	64	13	6	121	204
Ref2empty	9	9	8	24	50
Total:	346	37	21	3437	3841

90% of the smelly cells detected were not confirmed...

SpreadSheet Smells: Catalog Refinement

We consider three types of refinement:

- *1- Overlapped Smells*
Empty cells are also detected by patterns...
- *2- New Smells Detected*
Summation cells as constants!!
Layout smells...
- *3- Wrong Smells*
String distance alg. needs improvements
(numeric cells)

SpreadSheet Smells: Catalog Refinement

Smell \ Level	Level				Total
	Low smells	Medium smells	High smells	Not smells	
Empty cells	115	4	0	274	393
Patterns	9	5	0	90	101
Std. Dev. cells	21	0	0	234	255
String Dist.	13	2	7	290	312
QFD	64	13	6	121	204
Ref2empty	9	9	8	24	50
Total:	231	33	21	1033	1315

21.7% of correct smells detected

54 out 231 are Medium or High Smells!

Conclusions

- We presented a detailed study on spreadsheet smell detection.
- We proposed the first catalog of spreadsheet smells.
- We developed the SmellSheet Detective: a tool that implements the proposed catalog.
Available at

<http://ssaapp.di.uminho.pt>