



Software Architecture
for Mental Health
Self-Management

SAMS: Data and Text Mining for Early Detection of Alzheimer's Disease

November, 2016
Dr Christopher Bull



Aim of talk

- What is SAMS
- Data Capture
 - Problems and solutions to acquiring this type of text/data
- NLP
 - Tools used
 - Existing
 - Bespoke
- Reflections

Who am I?

Dr Christopher Bull

c.bull@lancaster.ac.uk

[@ChrisBull88](https://twitter.com/ChrisBull88)

[Insert dashing photo here]

- 2011 – PhD
- 2014 – SAMS (PDRA)
- 2016 – Mobile Age (PDRA)

-
- Software Engineering
 - Education/Pedagogy
 - Digital Health Technologies

SAMS Overview

Problem

-
- National Dementia Strategy (2009): early ('timely') diagnosis
 - Only about 50% of people with dementia currently receive a diagnosis
 - Diagnosis is often late - moderate or severe stages

What is Alzheimer's Disease?

- Alzheimer's is the most common cause of dementia (estimated 60%-80% of cases)
 - Dementia “*describes symptoms that occur when the brain is affected by certain diseases or conditions*”
- Symptoms include:
 - memory loss
 - difficulties with:
 - thinking
 - problem-solving
 - language
- Ultimately fatal

Source: Alzheimer's Society

SAMS



Goal:

Explore Technology-dependent proxy markers
Of Alzheimer's Disease



Aims:

- Non intrusive capture of computer use
- Mine the data for trends and patterns
- Infer longitudinal changes in cognitive health



Team



Professor Pete Sawyer School of Computing and Communications, Lancaster University

Dr Paul Rayson School of Computing and Communications, Lancaster University

Dr Christopher Bull School of Computing and Communications, Lancaster University

Professor Alistair Sutcliffe School of Computing and Communications, Lancaster University

Professor Alistair Burns National Clinical Director for Dementia in England, Institute of Brain, Behaviour and Mental Health, University of Manchester

Dr Iracema Leroi Institute of Brain, Behaviour and Mental Health, University of Manchester

Gemma Stringer Institute of Brain, Behaviour and Mental Health, University of Manchester

Dr Samuel Couth Institute of Brain, Behaviour and Mental Health, University of Manchester

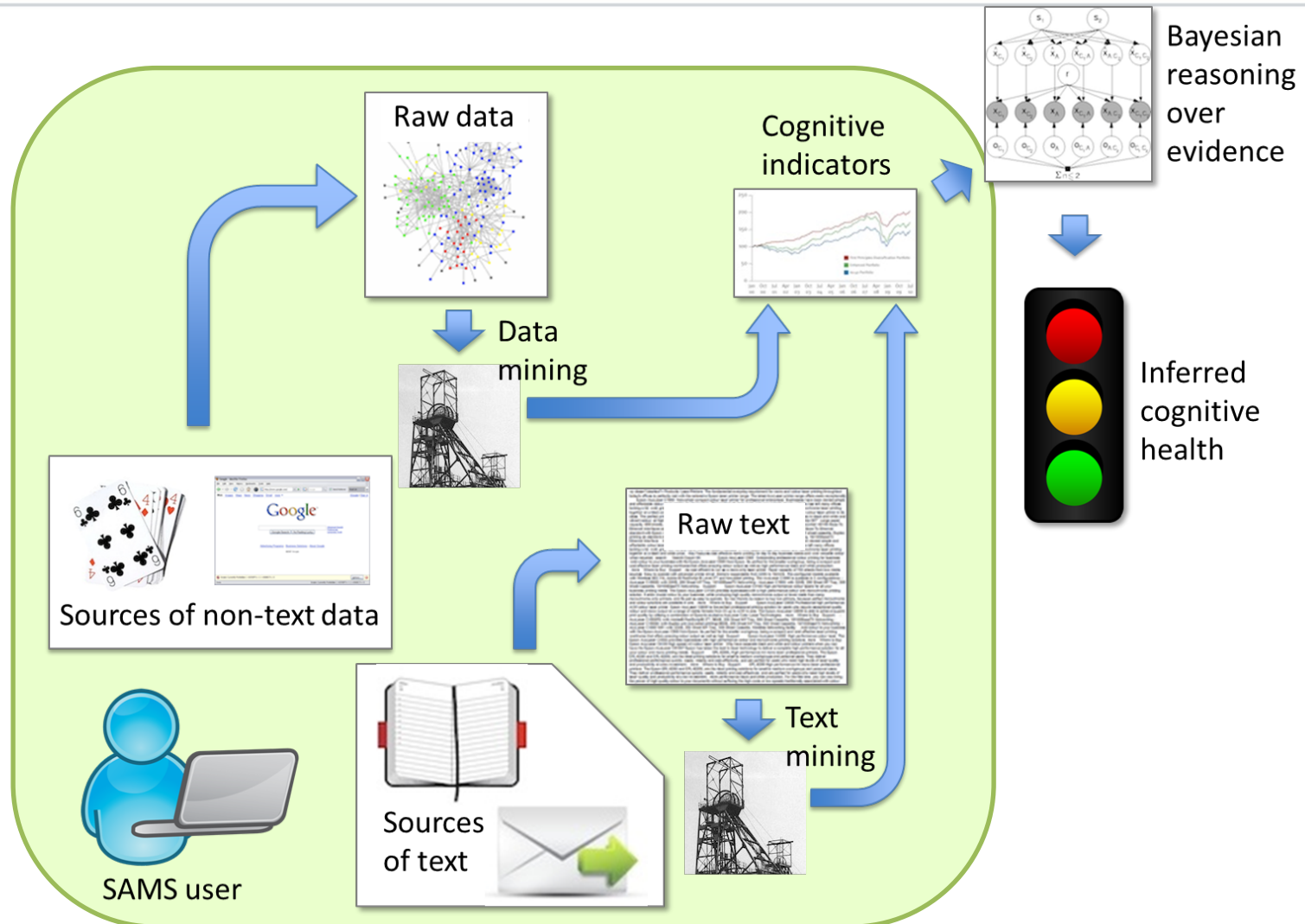
Professor John Keane School of Computer Science, University of Manchester

Dr Ann Gledson School of Computer Science, University of Manchester

Professor Clive Ballard Wolfson Centre for Age-Related Diseases, King's College London



Data Flows



Current Status

-
- Project funding ended September 2016
 - On-going analysis

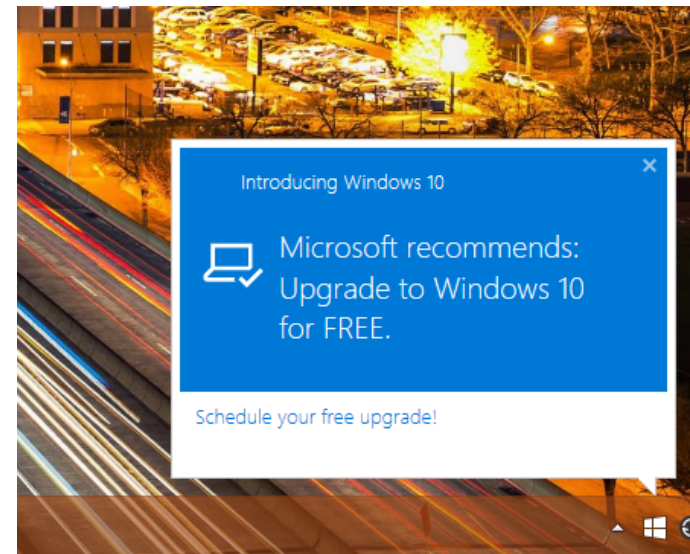
My Role in SAMS ...and Data Collection

My Role

-
- Data capture software
 - Software Design/implementation
 - SAMS Manager
 - Browser extensions
 - Maintenance (obviously)
 - Text Mining
 - Text extraction (reconstruction)
 - Reusing existing NLP pipeline (Wmatrix; UCREL)
 - Implementing extensions to pipeline for specific heuristics
 - General Project Support (Team & Participants)
 - Consider challenges

Challenges

- Volatility of participant computers
 - Unexpected updates
 - Varying shutdown procedures
 - Various software setups (anti-virus etc.)
- Weak performing computers (and not monopolise valuable resources)
 - Again, various hardware/software setups
- Ethical challenges
 - Privacy/Security
- Novel monitoring approaches
- Internet Explorer *sigh*
- Win 10 roll-out mid project →



Abstract Architecture (Data Collection)

Collecting context,
not just raw data

Desktop/Application
Monitor Processes

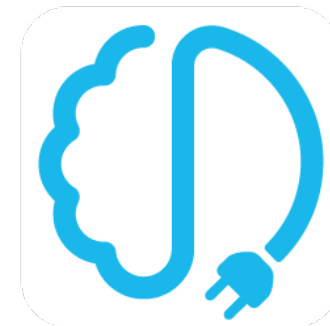


Encrypt Logs

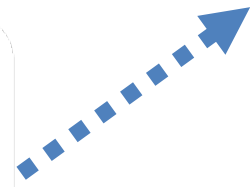


Secure SAMS Server

Browser Extensions



Manager Process



Desktop/Application Monitor Processes

Desktop/App Monitor



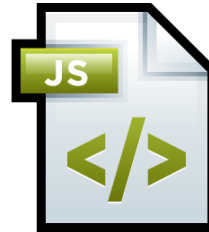
- ▶ C# input event listeners
 - ▶ Variety of Mouse, keyboard.



- ▶ Windows Automation API: UI Automation (UIA)
 - ▶ Observe UI elements (and properties) a user interacts with.
 - ▶ Provides context behind events.

Browser Extensions

Browser Extension



Webpage black/whitelist
(e.g. no https:// unless predefined)



JS DOM parsing
(text fields and interactive elements)



JS event listeners
& context identifier
(Click, Mouse-Move, Focus etc.)



Log message caching (volatile)



Encryption



Write log files



Browser Monitoring - Challenges

- Context to events
- Constantly changing or dynamic DOM

About 2,420,000,000 results (0.20 seconds)

Speedtest.net by Ookla - The
www.speedtest.net/ 
Test your Internet connection bandwidth with our interactive broadband speed test from our servers around the world.
[My Results - Ookla Speedtest Mobile](#)

Test cricket - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Test_cricket
Test cricket is the longest form of the sport of cricket, between national representative teams.
[List of Test cricket records - ICC World Test Championship](#)

In the news





Ashes 2015
Lord's Test
BBC Sport - 1
Australia wicket
second Ashes

New Diabetes Test Could Offer 'Real Time'
Sky News - 2 hours ago

Shane Watson dropped by Australia for
The Guardian - 8 hours ago

[More news for test](#)

BBC - Science & Nature - Human Body and Mind - Spot The ...
www.bbc.co.uk > [BBC Science](#) > [Human Body & Mind](#) > [The Mind](#) 
This experiment is designed to test whether you can spot the difference between a fake smile and a real one; It has 20 questions and should take you 10 minutes ...

Test Format - Pearson
pearsonpte.com/test-format/ 
To complete a PTE Academic test, you will need to attend a secure Pearson test centre. You will use a computer and headset to listen to, read and respond to ...

Choose Add-ons

Speed up browsing by disabling the add-ons that you don't want

These add-ons increase the time it takes to start the browser, open a new tab or navigate to websites by an average of 0.33 seconds. You can also enable or disable add-ons in the Manage Add-ons dialogue.

IEPlugin.BHO (Not verified) Lancaster University	0.33 seconds	<input type="button" value="Disable"/>
Logitech SetPoint Logitech	0.00 seconds	<input type="button" value="Disable"/>

Tell me when the delay caused by add-ons exceeds:

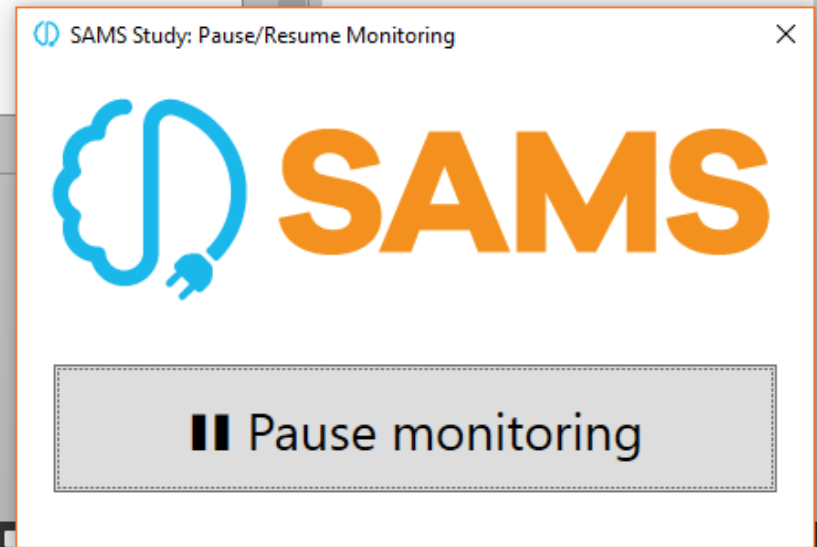
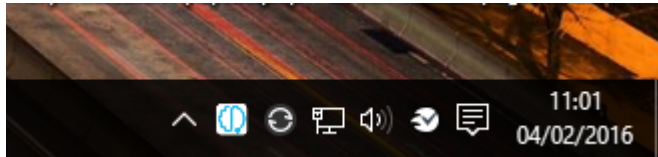
Speed up browsing by disabling add-ons.



Manager/Uploader

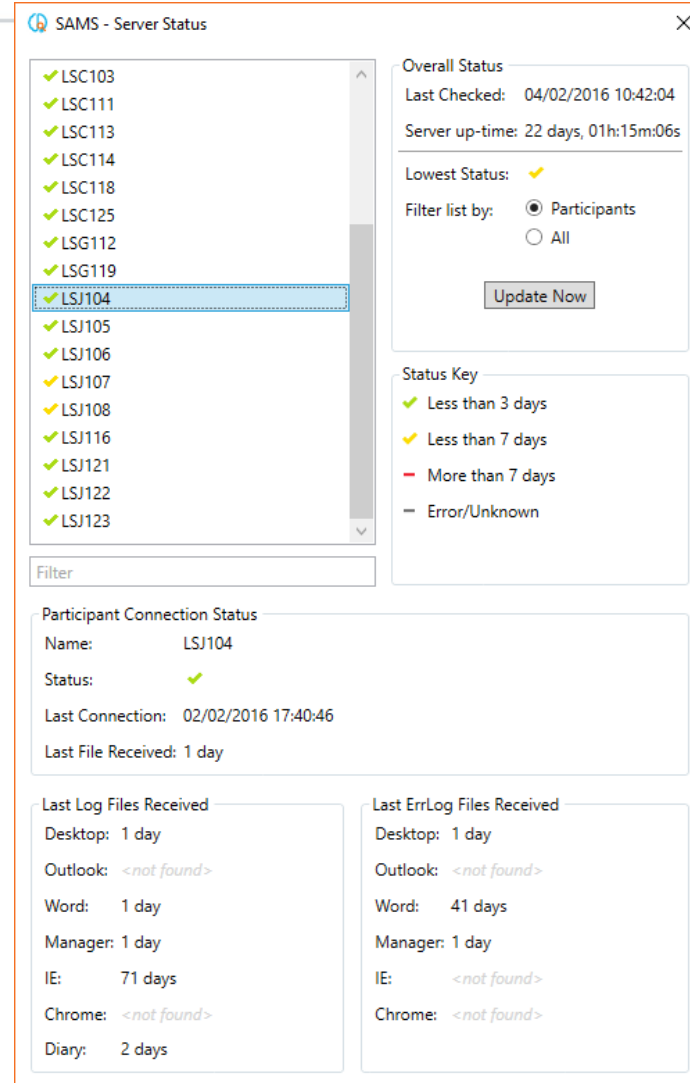
-
- Process management
 - Server communication
 - Remote updating
 - Log message caching and encryption

Manager (2)



Project Support

- Participant Status Checker
 - For clinical & Tech teams
 - +Android app
- Phone support
 - Clinical Team
 - Participants
- Participant visits (Installs)



The screenshot shows the 'SAMS - Server Status' application window. It features a list of server IDs on the left, a summary panel on the right, and detailed status information at the bottom.

Server List:

- ✓ LSC103
- ✓ LSC111
- ✓ LSC113
- ✓ LSC114
- ✓ LSC118
- ✓ LSC125
- ✓ LSG112
- ✓ LSG119
- ✓ LSJ104 (highlighted)
- ✓ LSJ105
- ✓ LSJ106
- ✓ LSJ107
- ✓ LSJ108
- ✓ LSJ116
- ✓ LSJ121
- ✓ LSJ122
- ✓ LSJ123

Overall Status:

- Last Checked: 04/02/2016 10:42:04
- Server up-time: 22 days, 01h:15m:06s
- Lowest Status: ✓
- Filter list by: Participants, All
- Update Now button

Status Key:

- ✓ Less than 3 days
- ✓ Less than 7 days
- More than 7 days
- Error/Unknown

Participant Connection Status:

- Name: LSJ104
- Status: ✓
- Last Connection: 02/02/2016 17:40:46
- Last File Received: 1 day

Last Log Files Received:

- Desktop: 1 day
- Outlook: <not found>
- Word: 1 day
- Manager: 1 day
- IE: 71 days
- Chrome: <not found>
- Diary: 2 days

Last ErrLog Files Received:

- Desktop: 1 day
- Outlook: <not found>
- Word: 41 days
- Manager: 1 day
- IE: <not found>
- Chrome: <not found>

Existing Study(s)

Nun Study:

- Measures obtained from autobiographies
- written over a 60-year span (age 22 to 83).

	No dementia	Dementia
Grammatical complexity	-mean 4.78 -declined .04 units per year	-mean 3.86 -declined .03 units per year.
Idea density	-mean 5.35 propositions per 10 words - declined .03 units per year	-mean 4.34 propositions per 10 words -declined .02 units per year.

Language Decline Across the Life Span: Findings From the Nun Study

Susan Kemper
 University of Kansas

Lydia H. Greiner
 University of Kentucky

Janet G. Marquis, Katherine Prenovost, and Tracy L. Mitzner
 University of Kansas

Propositional Idea Density (P-density)

- “Idea density [...] is the number of expressed propositions divided by the number of words. In terms of semantics, idea density is a measure of the extent to which the speaker is making assertions (or asking questions) rather than just referring to entities”
 - “Automatic measurement of propositional idea density from part-of-speech tagging” (Brown et al, 2008)
- Existing Implementation
 - CPIDR (Computerized Propositional Idea Density Rater)
 - (pronounced “spider”)
 - only tool to automate this*

* At time of starting SAMS

Kusari (Toolchain manager)

“Toolchain and data dependency manager for use with conventional NLP toolchains”


Dr Steve Wattam

<https://delta.lancs.ac.uk/Steve/kusari>

<https://delta.lancs.ac.uk/Steve/kusari-links>

Toolchain

Spelling Variation ↓	VARD ucrel.lancs.ac.uk/vard/ Java
Part Of Speech Tagger ↓	CLAWS ucrel.lancs.ac.uk/claws/ C
Semantic Tagger ↓	USAS ucrel.lancs.ac.uk/usas/ C
Frequency Lists ↓	Tmatrix ucrel.lancs.ac.uk/wmatrix/ C
SAMS software	SNOWCAT delta.lancs.ac.uk/SAMS/SNOWCAT Java



SNOWCAT

Sams a**N**alysis of **O**utput from **W**matrix for the **C**ognitive **A**ssessment of **T**ext

- Input
 - Tmatrix (FQLs)
 - USAS (Sem)
- Output
 - CSV of metrics

SNOWCAT: Sample Output (1/2)

- Total Words (MWE), 26278
- Total Words, 27787
- Vocabulary size (MWE), 3533
- Vocabulary size, 3444
- **Type:Token (ratio; MWE), 0.134**
- **Type:Token (ratio), 0.124**
- **Type:Token (normalised ratio), 0.403**
- Words occurring once (MWE), 1842
- Adjective (total; MWE), 1288
- Adjective (ratio; MWE), 0.049
- Noun (total; MWE), 4280
- Noun (ratio; MWE), 0.163
- ...

SNOWCAT:

Sample Output (2/2)

- Pronoun (total; MWE), 2672
- Pronoun (ratio; MWE), 0.102
- Verb (total; MWE), 6135
- Verb (ratio; MWE), 0.233
- Content words (total; MWE), 13757
- Content words (ratio; MWE), 0.524
- Filler words (total; MWE), 183
- Filler words (ratio; MWE), 0.007
- Noun:Verb (ratio; MWE), 0.698
- Mean Length of Utterance, 27.653
- VARD Variant (total), 69
- VARD Variant (ratio), 0.003
- **Propositional Idea Density, 0.565**

Early (unpublished) Results

- Validate P-Density (comparison to CPIDR tool)
- Uses novelist study to explore usefulness of SNOWCAT metrics
- [Show spreadsheet of early (unpublished) results]

Charts

[Charts removed from this public copy of the presentation, due to the data being pre-publication]

What's next?

-
- Continue NLP analysis
 - Correlate Data and Text Mining analyses
 - ...SAMS 2.0

Lessons Learnt

- Ethical process
 - Affects fundamental design decisions
- Complexity of data collection outside of “lab setting”
- Validating other studies/claims important

Thank you



<http://ucrel.lancs.ac.uk/sams/>
c.bull@lancaster.ac.uk
[@ChrisBull88](https://twitter.com/ChrisBull88)

November, 2016
Dr Christopher Bull

Publications

ucrel.lancs.ac.uk/sams/papers.php

- **Combining data mining and text mining for detection of early stage dementia: the SAMS framework.**
[Bull, C.](#), Asfiandy, D., Gledson, A., Mellor, J., Couth, S., Stringer, G., Rayson, P., Sutcliffe, A., Keane, J., Zeng, X., Burns, A., Leroi, I., Ballard, C., & Sawyer, P. (2016). In *LREC-2016 Workshop: RaPID-2016* [[proceedings](#); [slides](#)]
- **From Click to Cognition: Detecting cognitive decline through daily computer use.**
 Stringer, G., Sawyer, P., Sutcliffe, A., & Leroi, I. (2015). In D. Bruno (Ed.), *The Preservation of Memory: Theory and Practice for Clinical and Non-Clinical Populations* (pp. 93-103). Hove, UK: Psychology Press. [[online preview](#)]
- **Dementia and Social Sustainability: Challenges for Software Engineering.**
 Sawyer, P., Sutcliffe, A., Rayson, P., & Bull, C. (2015). In *37th International Conference on Software Engineering (ICSE '15)* (pp. 527-530). Florence, Italy: IEEE. DOI: [10.1109/ICSE.2015.188](https://doi.org/10.1109/ICSE.2015.188)
- **Discovering affect-laden requirements to achieve system acceptance.**
 Sutcliffe, A., Rayson, P., Bull, C., & Sawyer, P. (2014). In *22nd IEEE International Requirements Engineering Conference (RE'14)*. (pp. 173-182). IEEE. DOI: [10.1109/RE.2014.6912259](https://doi.org/10.1109/RE.2014.6912259)