SAMS
Software Architecture for Mental Health Self-Management

Lancaster University

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

November, 2016

Dr Christopher Bull

Aim of talk

Lancaster University

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

Who am I?

Lancaster University

SAMS

Dr Christopher Bull

c.bull@lancaster.ac.uk

@ChrisBull88

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

- Software Engineering
- Education/Pedagogy
- Digital Health Technologies

[Insert dashing photo here]

High Wire
Innovation for the Digital Economy
An EPSRC Doctoral Training Centre

School of Computing and Communications

Lancaster University

SAMS

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

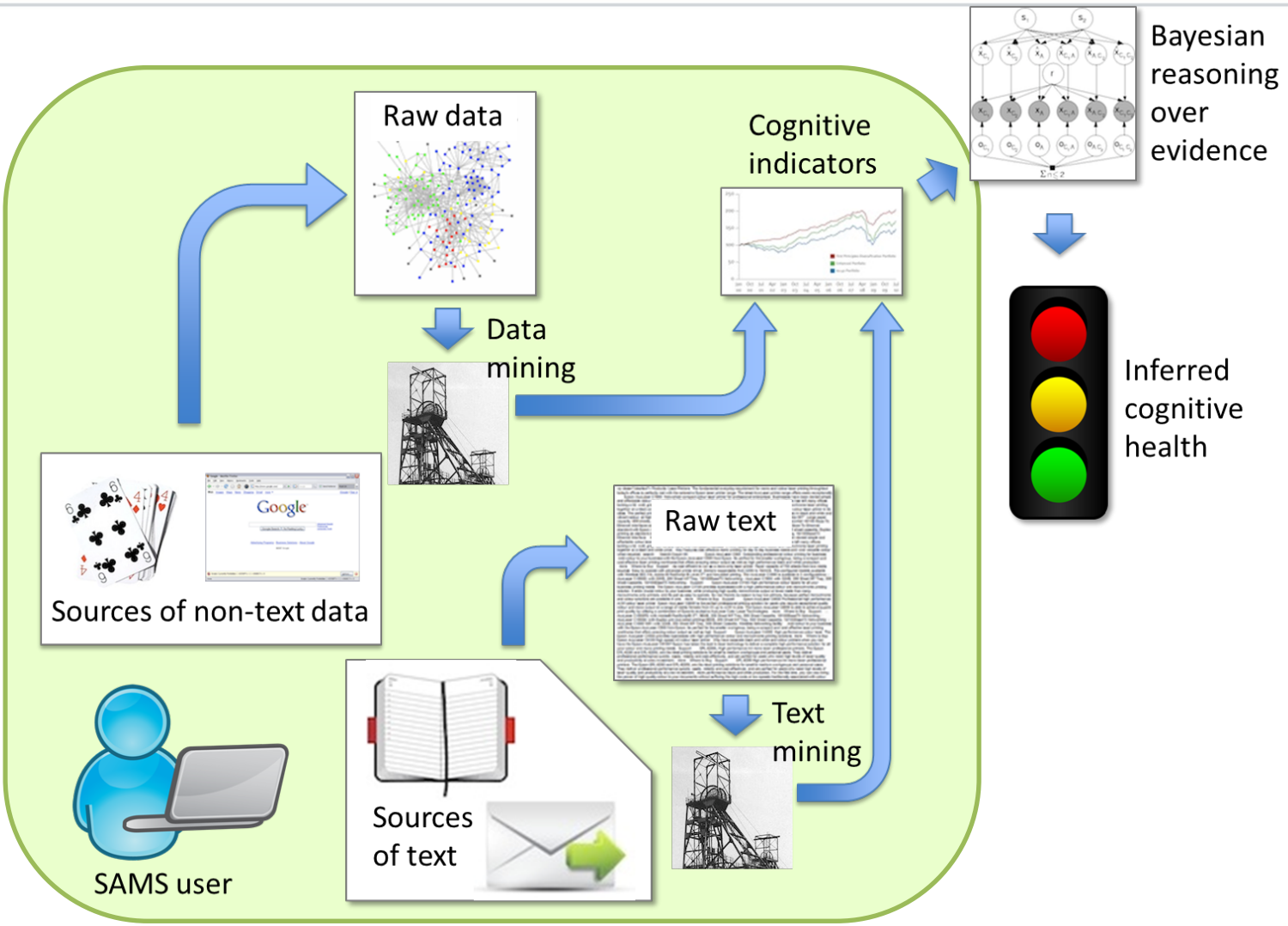
Aims:

- 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


Raw data
Cognitive indicators
Bayesian reasoning over evidence
Data mining
Inferred cognitive health
Raw text
Sources of non-text data
Google
Text mining
Sources of text
SAMS user

Current Status

- Project funding ended September 2016
- On-going analysis

My Role in SAMS

…and Data Collection

My Role

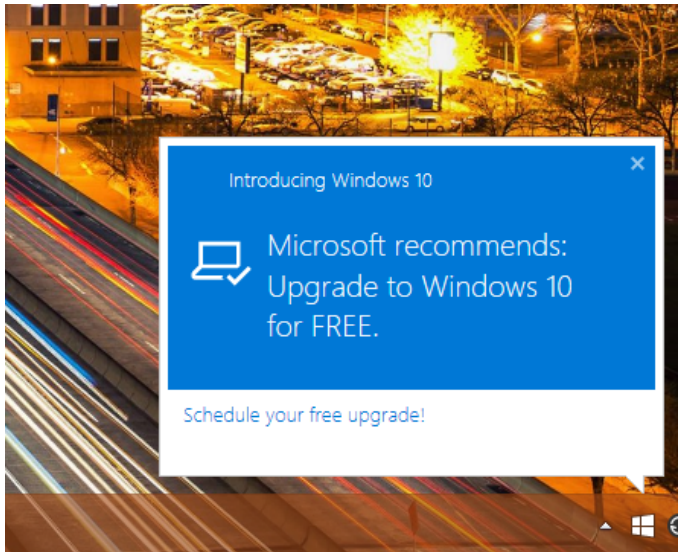
Lancaster University

SAMS

- 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 →



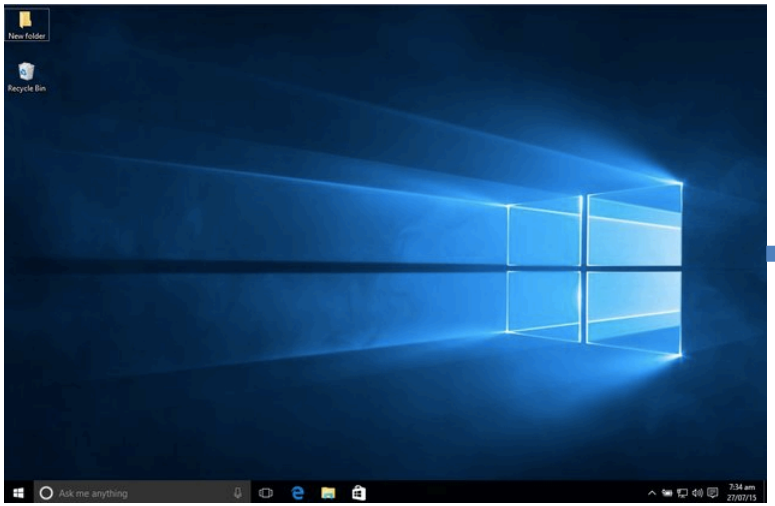
Abstract Architecture (Data Collection)

Lancaster University

SAMS

Collecting context, not just raw data

Desktop/Application Monitor Processes



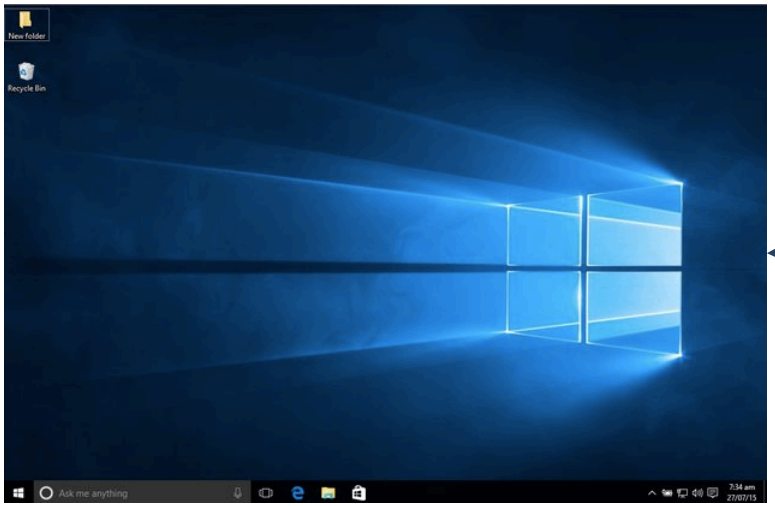
Encrypt Logs

Secure SAMS Server

Browser Extensions

Manager Process

Desktop/Application Monitor Processes



Desktop/App Monitor

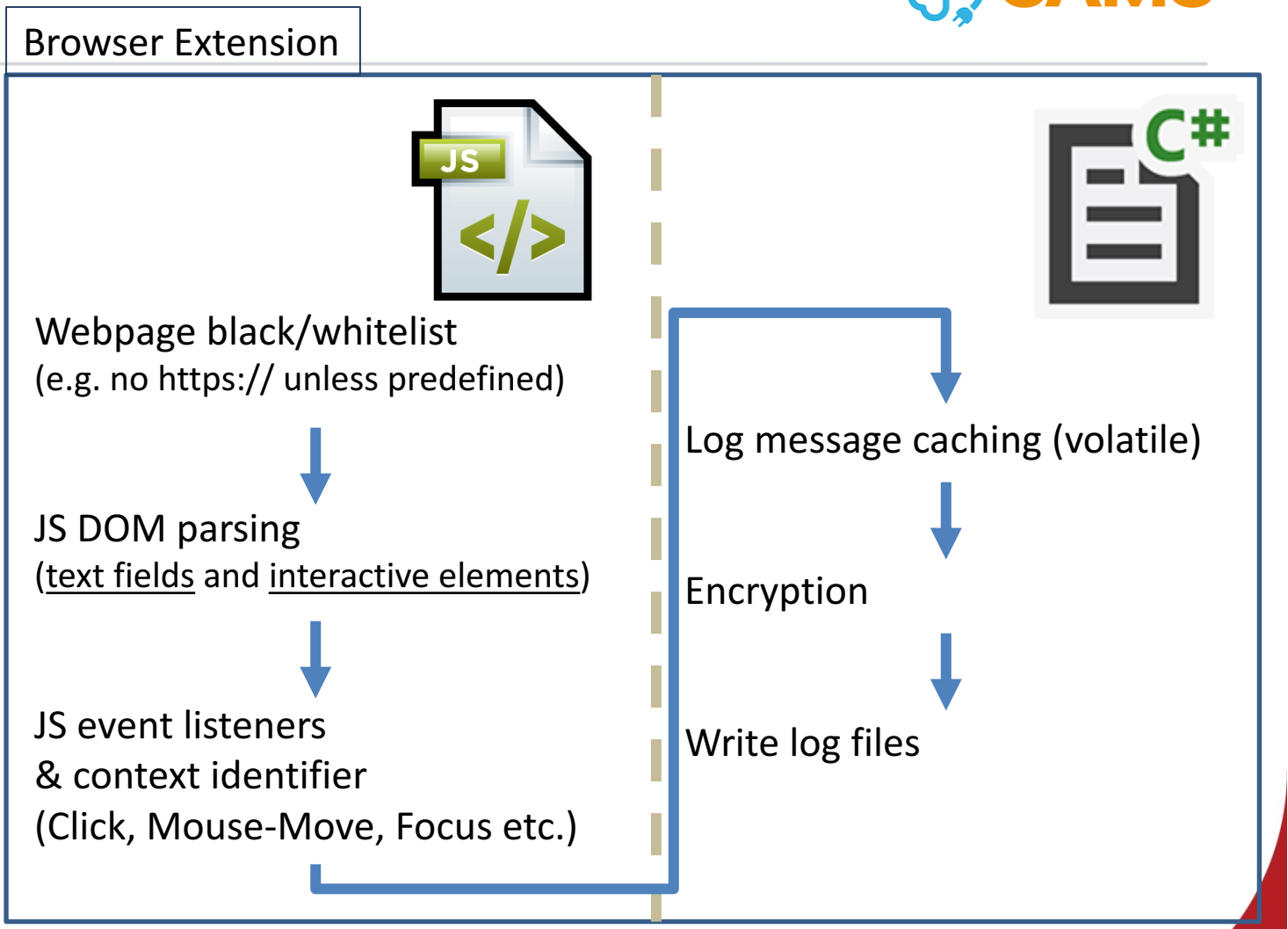
- C# input event listeners
  - Variety of Mouse, keyboard.
- Windows Automation API: UI Automation (UIA)
  - Observe UI elements (and properties) a user interacts with.
  - Provides context behind events.

** Work of Dr Ann Gledson, Mancs*

Browser Extensions

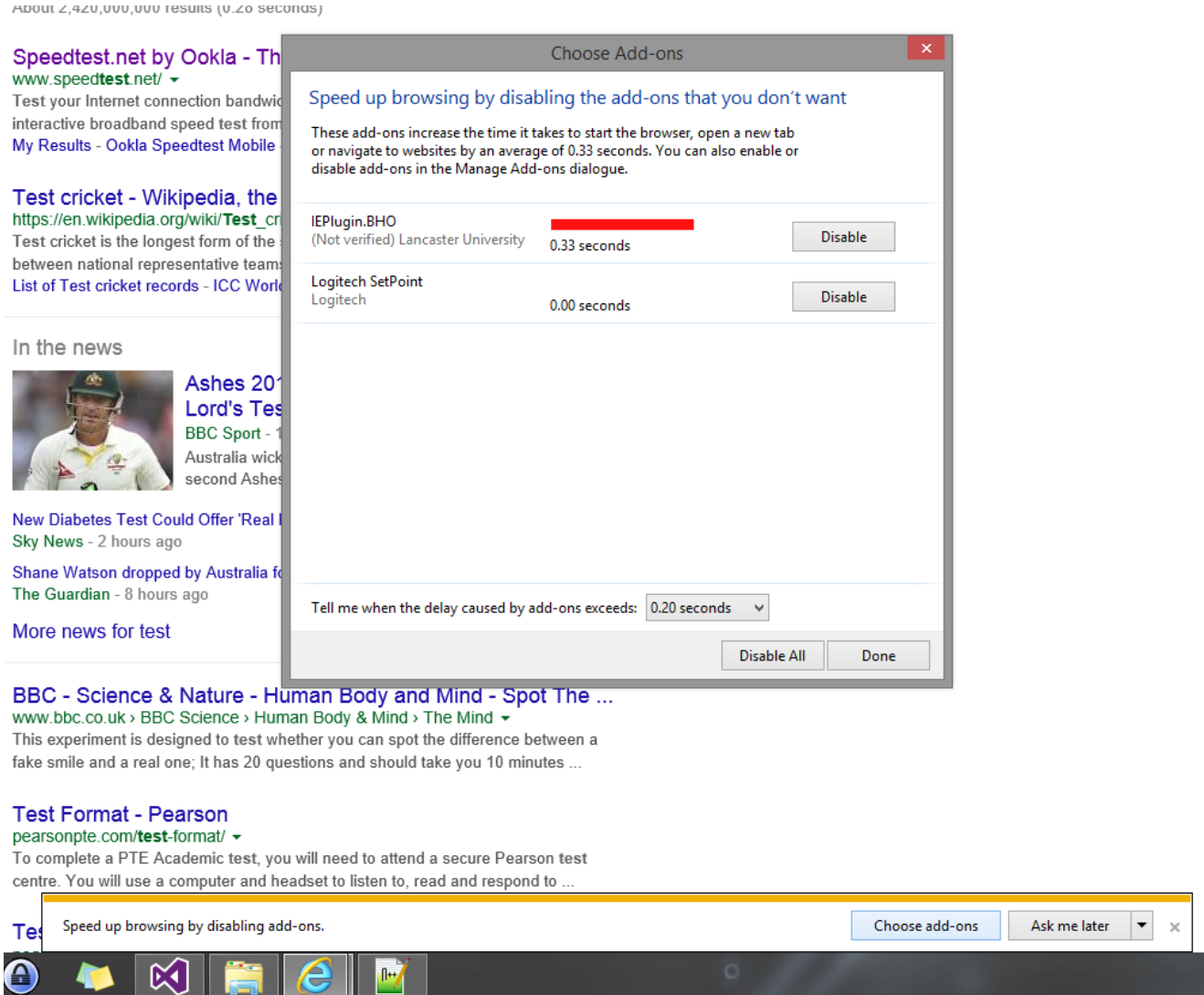
Lancaster University

SAMS



Browser Monitoring - Challenges

- Context to events
- Constantly changing or dynamic DOM



Manager/Uploader

Lancaster University

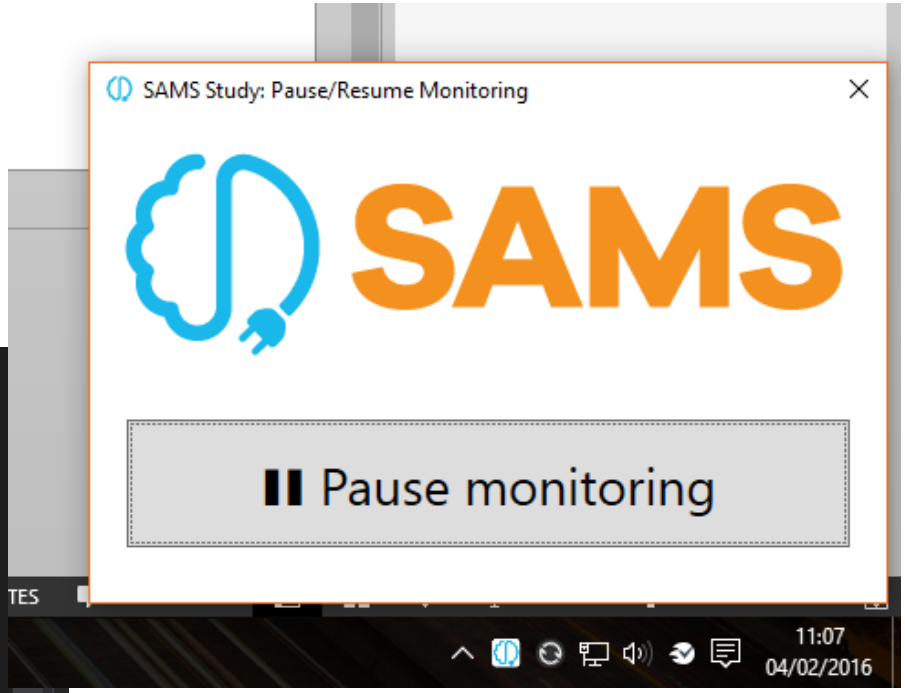
SAMS

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

Manager (2)

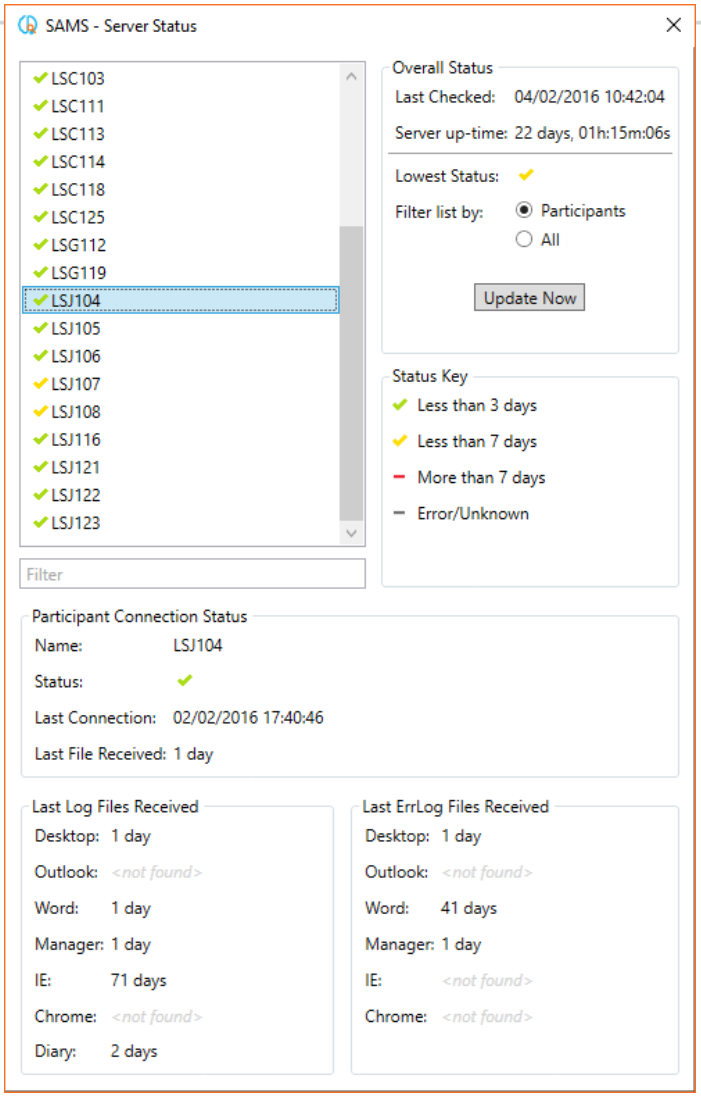
Lancaster University

SAMS



Project Support

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



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.

Psychology and Aging
2001, Vol. 16, No. 2, 227–239

Copyright 2001 by the American Psychological Association, Inc.
0882-7974/01/$5.00 DOI: 10.1037//0882-7974.16.2.227

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)

Lancaster University

SAMS

- “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)

Lancaster University

SAMS

"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

Lancaster University

SAMS

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)

Lancaster University

SAMS

- 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)

Lancaster University

SAMS

- 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

Lancaster University

Thank you

http://ucrel.lancs.ac.uk/sams/
c.bull@lancaster.ac.uk
@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
- **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