Combining data mining and text mining for detection of early stage dementia: the SAMS framework

Christopher Bull*, Dommy Asfiandy†, Ann Gledson†, Joseph Mellor†, Samuel Couth‡, Gemma Stringer‡, Paul Rayson*, Alistair Sutcliffe*, John Keane†, Xiaojun Zeng†, Alistair Burns‡, Iracema Leroi‡, Clive Ballard§, Pete Sawyer*

*School of Computing and Communications, Lancaster University, UK
†School of Computer Science, University of Manchester, UK
‡Institute of Brain, Behaviour and Mental Health, University of Manchester, UK
§Wolfson Centre for Age-Related Diseases, King's College London, UK

Abstract

In this paper, we describe the open-source SAMS framework whose novelty lies in bringing together both data collection (keystrokes, mouse movements, application pathways) and text collection (email, documents, diaries) and analysis methodologies. The aim of SAMS is to provide a non-invasive method for large scale collection, secure storage, retrieval and analysis of an individual's computer usage for the detection of cognitive decline, and to infer whether this decline is consistent with the early stages of dementia. The framework will allow evaluation and study by medical professionals in which data and textual features can be linked to deficits in cognitive domains that are characteristic of dementia. Having described requirements gathering and ethical concerns in previous papers, here we focus on the implementation of the data and text collection components.

Keywords: dementia, corpus linguistics, natural language processing, data mining

1. Introduction

Dementia is a condition that currently affects around one in six people at the age of 80. Increasing life expectancy means that the number of people who develop dementia will increase. Taking the UK as an example, the number of people living with the condition is predicted to increase from the current figure of 850,000 to over two million by 2051 (Knapp et al., 2007).

Although most forms of dementia such as Alzheimer's Disease are currently irreversible and some are ultimately fatal, obtaining an early diagnosis can help maintain quality of life by treating debilitating side effects, such as depression. Moreover, when improved therapies do eventually become available, it is likely that they will have to be administered before the damage to the brain becomes so severe as to render the therapy ineffective. Currently, diagnosis of dementia or of its harbinger, Mild Cognitive Impairment (MCI), is usually performed using paper-based cognitive tests such as the Montreal Cognitive Assessment (MoCA (Nasreddine et al., 2005)). These are designed to be administered in a clinical setting such as a memory clinic but this can be stressful for the subject and yield poor ecological validity. Worse, many subjects do not refer themselves for a health check until the disease is well advanced. There is therefore a strong interest in developing new techniques for detecting cognitive decline that do not suffer from these disadvantages.

Our work seeks to check for deficits in the same cognitive domains (memory, executive function, motor control and so on) that are tested by the paper tests, using everyday computer tasks as proxies for tasks in the tests (Jimison et al., 2006). The work is based on the simple idea that if someone is finding it increasingly hard to use their computer, then it might be because of change in cognitive function. Many older adults use a computer for (e.g.) home banking, shopping, and keeping in touch with family, so there is an opportunity to exploit the penetration of technology into seniors' homes by developing a non-invasive software tool that helps develop awareness of the users' cognitive health. In our work so far on the SAMS ("Software Architecture for Mental health Self management") project[1], we have focussed on practical problems of how to collect requirements for our monitoring software in order to achieve better acceptance by its end users, as well as the important related ethical concerns for the project (Sutcliffe et al., 2014; Sawyer et al., 2015; Stringer et al., 2015). In this paper, we describe the next stage of the development process. We provide an overview of the SAMS framework for data and text collection created in accordance with these requirements and cross cutting concerns. We also describe a preliminary analysis of initial data mining results.

2. Related Work

To date, little work appears to have been done in the data mining community on analysing sequential patient/user activities to detect the clinical indicators of disease. Seelye et al. (2015) use multiple regression and correlation on mouse movement data from 42 healthy and 20 participants with mild cognitive impairments (MCI) in order to observe that computer mouse movements are a potential indicator of MCI. Much research on mining healthcare data to detect links between health conditions uses association rule mining. This is the technique used by Shin et al. (2010) to mine diagnostic data for patients with essential hypertension and results demonstrate an association between essential hypertension, non-insulin dependent diabetes mellitus, and cerebral infarction. Ohsaki and Sato (2002) use pattern-based time-series data mining for "real medical data that are sequential, numerical and ill-defined", resulting in

[1] http://ucrel.lancaster.ac.uk/sams/

pattern combination rules for testing on chronic hepatitis medical test results. Outside the healthcare domain, sequential pattern mining is used on sequential data representing user activities; for example Pachidi et al. (2014) use this technique to analyse user clickstreams, the clicks made by computer users in order to analyse their use of software.

Turning to related work in the text mining or natural language processing (NLP) area, there is a growing body of research and interest in health-related research in recent years. In addition to this year's RaPID-2016 workshop hosted at LREC, there have been three "Computational Linguistics and Clinical Psychology" workshops held annually at ACL or NAACL since 2014[2], six "International Workshops on Health Text Mining and Information Analysis" held at various locations since 2008[3], and a NIPS 2015 Workshop on Machine Learning in Healthcare[4].

A number of papers have focussed on the notion of "idea density", approximated as the number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words, and its decline in old age and Alzheimer's disease (Snowdon et al., 1996; Kemper et al., 2001; Brown et al., 2008). This research used data from what became known as the longitudinal "Nun Study": a collection of autobiographies from the School Sisters of Notre Dame, written when they became nuns (18–32 years old), and cognitive tests much later in life (75–95 years old), and the CPIDR (Computerized Propositional Idea Density Rater) software which implements the metric. Real clinical study data that has been released for replication studies is hard to come by, no doubt due to medical ethics restrictions, so NLP researchers have tended to look elsewhere in their work. Garrard et al. (2005) considers three publications by British writer Iris Murdoch who continued to write novels even after she developed Alzheimer's disease. Garrard et al. (2005) analysed Murdoch's first published work (1954), her last (1995) along with another from 1978, in order to investigate language change using various measures include lexical diversity. Le et al. (2011) and Hirst and Feng (2012) extend this by including a large number of measures and more comparative data: 20 novels for Iris Murdoch, 15 novels for Agatha Christie, and 15 novels for PD James (as control). Using an SVM classifier, they deduce that Agatha Christie also probably developed dementia towards the end of her life.

The Western Collaborative Group Study (WCGS) proves to be a rich source of data for Jarrold et al. (2010), as it provides transcriptions of 15-minute interviews from a 40+ year wide ranging longitudinal study. They use a combination of part-of-speech tagging software, Linguistic Inquiry Word Count (LIWC) (Tausczik and Pennebaker, 2010) and CPIDR to contribute to key measures in a predictive model for Alzheimer's Disease, cognitive impairment and clinical depression. Similar methods were used by Jarrold et al. (2014) on samples from the Western Aphasia Battery to determine dementia subtypes. Finally, a more promising publicly available dataset is the DementiaBank. This is used by Orimaye et al. (2015), who apply machine learning to a combination of skip-gram features, and by Fraser et al. (2015) who employ a much larger (370) set of features to train a machine learning classifier to distinguish participants with Alzheimer's from healthy controls. The DementiaBank[5] clinical dataset consists of interview transcripts of MCI and control participants describing the Cookie-Theft picture component of the Boston Diagnostic Aphasia Examination. Compared to all these elicited interview datasets, the type of text that we are collecting via the SAMS non-invasive approach is significantly different. In contrast to other studies, SAMS will analyse text captured from everyday activities, i.e. email and dairies.

3. SAMS Framework: data/text collection

The SAMS framework is designed to record low-level events (i.e. mouse and keyboard), as well as higher-level contextual information about the Operating System and applications (e.g. drag/drop events, window resizing).

The framework faced a number of challenges, but one of the primary challenges derived from our aim to deploy it on real users' home computers and to collect data as they used their computers to do everyday things. Resources did not permit us to develop a SAMS product line configurable for every type, brand and version of desktop computer, operating system, web browser and desktop application. Guided by information about home computer usage and configurations that we elicited from a superset of the older adults we recruited as SAMS study participants, we took the pragmatic decision to develop SAMS to work on Windows 7, 8, and 10, the MS Office 2007 and later suites of desktop application software, and the Internet Explorer 11 and Chrome web browsers.

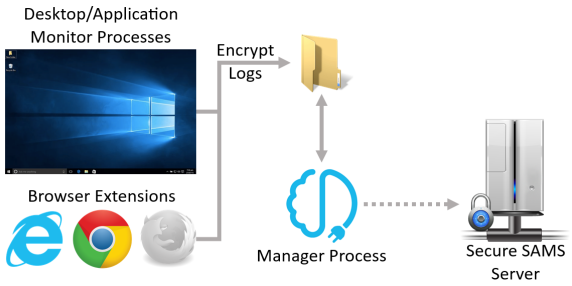

Figure 1: Abstract architecture of SAMS framework

The part of the SAMS framework dedicated to data and text collection is split into several components: the primary desktop logger, web browser logging extensions, and a manager component, see Figure 1. The desktop and web browser loggers are both responsible for data collection and text collection. The browser extensions are required in addition to the desktop logger due to web content in a browser being inaccessible to the desktop logger; a browser extension can have direct access to webpages content. The logs generated by these components are immediately secured using asymmetric encryption. The manager component

[2] http://clpsych.org/

[3] http://louhi2015.limsi.fr

[4] https://sites.google.com/site/nipsmlhc15/

[5] https://talkbank.org/DementiaBank/

is responsible for all user interface elements, for starting, stopping, and pausing the loggers, uploading the encrypted logs to the SAMS server, and updating the SAMS software.

3.1. Desktop Logger Component

The desktop logger records user activities at three levels, as shown in Table 1; level 1: keyboard and mouse, level 2: operating system (e.g. desktop activities), and level 3: application level. All windows events deemed potentially useful for detecting the clinical indicators of dementia are recorded, with the view to further analysis to determine those that are most pertinent. The events that are logged are detailed throughout this section. Activities are captured as a list of time-stamped events using a variety of technologies. Mouse/keyboard level detection utilises an imported .NET library[6]. At the operating system level, native C# .NET libraries[7] [8] are used to detect file system events (files changed, created and renamed) and changes to the clipboard. Microsoft UI Automation events[9] are used to record events such as opening/closing/minimizing/maximizing windows, changes in focus, menus opened/closed and elements selected by the user. At the application level, the Office Primary Interop Assemblies[10] and the Internet Explorer automation object [11] are used to detect events from Microsoft Word, Outlook and Internet Explorer, the three applications considered most relevant for monitoring activities of older adult users.

Further 'high level' events have been developed for the SAMS framework, derived from the low level data events described above. A mouse monitor has been created to read original mouse events, too abundant to be efficiently recorded and too low-level to be of use for later analysis, and aggregates these into mouse drags and mouse 'phases' (time periods between clicks or half second intervals), obtaining more useful information such as time, distance, and screen areas crossed. Similarly, key up and down events are paired and the code and duration are recorded. At the operating system level, mouse drag events are classified where possible into 'move', 'move into', 'resize' and 'scroll' events based on what is known about simultaneous low-level events (for example icon/window position or size changes, scroll and file system events). In addition, UI Automation[12] is used to maintain a map of the desktop including all window and icon positions. This map is used to derive higher level mouse move events, capturing moves into and out of icon or windows and to augment mouse event data with information such as the underlying icon/window name, position and display level.

3.2. Web Browser Extensions

The browser extensions provide the SAMS framework access to webpage content which otherwise is a blackbox to the desktop logger. When a web browser has focus we elect to halt keypress logging in the desktop loggers, allowing the browser extensions to take over that responsibility. The desktop loggers continue to log all other events. This helps avoid the collection of sensitive information such as passwords, as the browser extensions can easily distinguish between password and normal text fields.

The browser extensions work by injecting Javascript (JS) into all webpages. Websites that are loaded with https (secure), and not http, are not monitored; the assumption here is that https webpages are considered private and will likely contain sensitive information or have it entered into them by the participant (e.g. bank details on shopping websites). The injected JS parses the websites DOM, adding numerous event listeners to a wide variety of text and non-text elements detailed in Table 2. The events indicate user interactions and collect text. They can then be analysed later to determine behaviours. When these events fire they are logged to an encrypted file on the user's computer. The Manager component periodically picks up these files, as well as all other SAMS logs, and sends them to the SAMS server. Dynamic webpages, those that create DOM elements after the page has loaded, have a 'Mutation Observer'[13] listen for when new elements are attached to the DOM and adds the event listeners at runtime.

The text-related events that are collected within the browser extensions record a higher fidelity of meta information as well. Upon each participants' interaction with a text element, see 'Text Elements' in Table 2, the selection range (the index of highlighted text) is also recorded. This allows for key presses to easily be reconstructed as full bodies of text later, rather than just individual characters in the log files, and also provides an additional future analysis vector: analysing text editing processes. For example, log entries will indicate if a participant highlights some text and then replaces it.

The browser extensions developed for the SAMS framework focus on the Internet Explorer and Chrome web browsers. In addition to the initial information elicited from the the superset of SAMS participants, Internet Explorer was chosen because it comes pre-installed on Windows computers, and therefore likely to be used by people who favour default setups, and Chrome because it is the most popular browser in 2014/15[14].

All of the main web browsers used on Windows computers

[6] Application and Global Mouse and Keyboard Hooks .Net Library in C#: `http://globalmousekeyhook.codeplex.com/`

[7] FileSystemWatcher Class: `https://msdn.microsoft.com/en-us/library/system.io.filesystemwatcher(v=vs.110).aspx`

[8] Clipboard (.NET): `https://msdn.microsoft.com/en-us/library/windows/desktop/ms648709(v=vs.85).aspx`

[9] Microsoft UI Automation events: `https://msdn.microsoft.com/en-us/library/windows/desktop/ee671221(v=vs.85).aspx`

[10] Office Primary Interop Assemblies: `https://msdn.microsoft.com/en-us/library/15s06t57.aspx`

[11] InternetExplorer object: `https://msdn.microsoft.com/en-us/library/aa752084(v=vs.85).aspx`

[12] Microsoft UI Automation: `https://msdn.microsoft.com/en-us/library/windows/desktop/ee684009(v=vs.85).aspx`

[13] JS Mutation Observer: `https://www.w3.org/TR/dom/#mutation-observers`

[14] Web browser statistics: `http://www.w3schools.com/browsers/browsers_stats.asp`

Table 1: Desktop Logger's captured events.

Level	Sub-level	Description
Level 1	Keyboard	KEYBOARD UP
	Mouse	DESKTOP MOUSE WHEEL MOVE
		MOUSE DOUBLE CLICKED
		MOUSE DRAG PHASE
		MOUSE PHASE COMPLETED (time and mouse movement between clicks)
		MOUSE UP
		MOUSE MOVES IN/OUT OF DESKTOP WINDOWS OR ICONS
Level 2	Clipboard	CLIPBOARD UPDATED
	Drag	DESKTOP DRAG (start/end times and positions etc.)
	File system events	FILE CHANGED
		FILE CREATED / FSW FILE DELETED
		FILE RENAMED
	User interface system events	ELEMENT ADDED/REMOVED FROM/TO A SELECTION
		ELEMENT SELECTED BY USER
		FOCUS CHANGED
		MENU OPENED
		USER INTERFACE OBJECT INVOKED
		WINDOW OPENED/CLOSED
		WINDOW MAXIMIZED/MINIMIZED/TO NORMAL
Level 3	Internet Explorer	OPEN/CLOSE IE WINDOW OR TAB
	Outlook	CHANGE EMAILS SELECTED
		MOVE EMAIL MESSAGE
		START/QUIT OUTLOOK
		READ/REPLY EMAIL
		SEND EMAIL
		SWITCH FOLDERS
	Word	CHANGE TEXT SELECTED
		OPEN/CLOSE/SAVE/SWITCH DOC

(IE, Chrome, Firefox) were found to be capable of allowing their extensions to write files to the user's computer, and therefore enable logging alongside a desktop application counterpart. Microsoft Edge is not included in the SAMS framework because, at the time of writing, extension support for that browser is not yet available.

4. Preliminary Results

A controlled experiment has been completed comparing a healthy control group with a MCI/mild dementia group using a set task composed of GUI-Windows operations, Email-Outlook use, Word processing and Internet searching. Both groups experienced the same conditions in the experiment, and in the longitudinal study recordings are not intrusive and users will not be distracted by the monitoring software. Full consent for the study was given by all participants, following the ethics standards of Manchester University. In these preliminary results, we focus on the data mining aspects only, and will report text mining results in future papers.

The logger outputs time stamped records at the msec level for each user and system generated events at two levels: general from Microsoft UIA tool and SAMS augmented detail of event identities. Event identities are recorded faithfully from all Microsoft browsers but the fidelity of identity varied between web sites with other Internet Browsers.

Preliminary analysis of logs produced by the SAMS tool have shown that even simple frequency analysis of general event types display encouraging trends. For instance, the frequency of individual low-level events associated with mouse movement and keyboard presses have been observed to be different in distribution between healthy and MCI groups.

The difference in distribution amongst the groups for some of these general event-types was found to be significant according to a Mann Whitney U test. Some results can be seen in Table 3. The fact that such differences exist, especially in mouse-movement data, is supported by the work of Seelye et al. (2015).

We are now engaged in a longitudinal study, with 32 installations of the SAMS software running unobtrusively on participants' home computers/laptops. Participants have been recruited that conform to a set of selection criteria based on factors such as their age and home computer ownership and use. Our aim is to discover whether the SAMS software can detect cognitive change within individuals during

Table 2: Web events collected.

	HTML Elements	JS Events
All Elements	(all text and non-text elements, listed below, have this superset of event listeners attached)	click, dblclick, mouseover, contextmenu[a], focusin, focusout
Text Elements	<input type="text">, <input type="search">, <textarea>, <* contenteditable>, <* g_editable>	keydown, keyup, keypress, mouseup, cut, copy, paste, dragstart, dragend
Non-Text Elements	<a>, <button>, <* role="button">, <input>[b], <select>, <img>	mousedown[c], keydown[d], keyup[d], keypress[d]

[a] Could indicate a right-click spelling correction.
[b] Includes password fields, avoiding password collection.
[c] Log event before (e.g) button causes page navigation.
[d] Only for collecting 'Enter' or 'Tab' key event.

the course of the study, informed by what we discover from analysis of the controlled experiment. Ground truth is established by clinical cognitive assessments of each participant at the start, mid-point and end of the study period.
Our current analysis strategy is to apply data mining cluster and pattern analysis algorithms to investigate changes within individuals over time and inter-individual variations with known norms for age/gender cohorts of our senior participants (range 65–78 years). Given these reassuring findings, future work includes sequence analysis such as learning Markov models or using SPADE-like algorithms, which have been applied to finding temporal patterns in web-log data (Demiriz, 2002), to discover richer interaction of low-level events over time capable of identifying signs of MCI. Sequence mining will be used to identify atypical user behaviour and errors which might indicate cognitive problems linked to MCI and early dementia. Integration of evidence from data mining activity patterns, sequences of computer operation, and text analysis metrics will be investigated using Bayesian nets to implement a 'diagnostic' model that traces measures derived from data and text mining to cognitive indicators which are associated with MCI. The challenge we face is finding a weak signal indicative of disease in noisy data where variations might be caused by interruptions, changes in user mood, or many environment factors.

5. Conclusion and Future Work

We have developed a novel system architecture that not only logs keyboard, mouse, and contextual environment/application data but also interprets these events as user behaviours. This, combined with text capture from email and diary entries, is input into data and text mining tools, so we can analyse early signs of dementia by combining evidence from many measures across time. The SAMS project is now entering the analysis phase and we await the end of our longitudinal study. We have selected a set of potential text mining features from the related work described in Section 2. These are being implemented around the already existing Wmatrix tag wizard pipeline for part-of-speech and semantic tagging (Rayson, 2008), along with variant detection using VARD (Baron and Rayson, 2008), and the extraction of type and token frequency data at three levels: lexical, grammatical and semantic tags. The SAMS software framework will be available open source from Github[15] and the project website. In future projects, we intend to apply the SAMS architecture for health monitoring in a wide range of domains including mental health as well as dementia.

6. Acknowledgements

The work described in this paper is funded by the Engineering and Physical Sciences Research Council (EPSRC) in the UK, within the Software Architecture for Mental Health Self-Management (SAMS) project, references EP/K015796/1, EP/K015761/1 and EP/K015826/1.

7. Bibliographical References

Baron, A. and Rayson, P. (2008). VARD2: a tool for dealing with spelling variation in historical corpora. Postgraduate Conference in Corpus Linguistics, Aston University, Birmingham, UK.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., and Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40(2):540–545.

Demiriz, A. (2002). webSPADE: a parallel sequence mining algorithm to analyze web log data. In *Proceedings of the International Conference on Data Mining (ICDM '02)*, pages 755–758. IEEE.

Fraser, K. C., Meltzer, J., and Rudzicz, F. (2015). Linguistic features identify Alzheimer's Disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422.

Garrard, P., Maloney, L. M., Hodges, J. R., and Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, 128(2):250–260.

Hirst, G. and Feng, V. W. (2012). Changes in style in authors with Alzheimer's disease. *English Studies*, 93(3):357–370.

Jarrold, W. L., Peintner, B., Yeh, E., Krasnow, R., Javitz, H. S., and Swan, G. E., (2010). *Brain Informatics: International Conference, BI 2010, Toronto, ON, Canada, August 28-30, 2010. Proceedings*, chapter Language Analytics for Assessing Brain Health: Cognitive Impairment, Depression and Pre-symptomatic Alzheimer's Disease, pages 299–307. Springer Berlin Heidelberg, Berlin, Heidelberg.

Jarrold, W., Peintner, B., Wilkins, D., Vergryi, D., Richey, C., Gorno-Tempini, M. L., and Ogar, J. (2014). Aided diagnosis of dementia type through computer-based

[15] https://github.com/UCREL

Table 3: Frequencies of event types per user. The p-value is for a Mann Whitney U test between HC and MCI groups. We can see that the HC group press the keyboard more often. The MCI group double-click much more frequently.

Event type	HC event counts	MCI event counts	p-value
KEYBOARD_UP	546, 619, 458, 926, 445, 508, 406, 683, 849, 244, 482, 280, 718, 628, 350, 441, 599, 460, 543, 439	253, 214, 595, 402, 452, 554, 364, 206, 410, 289, 229	0.0036
MOUSE_DBLCLICKED	0, 0, 12, 0, 8, 0, 0, 10, 0, 0, 8, 0, 0, 0, 0, 0, 0, 0, 0, 4	33, 18, 59, 15, 7, 0, 75, 12, 27, 0, 18	0.0002

analysis of spontaneous speech. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 27–37, Baltimore, Maryland, USA, June. Association for Computational Linguistics.

Jimison, H., Jessey, N., McKanna, J., Zitzelberger, T., and Kaye, J. (2006). Monitoring computer interactions to detect early cognitive impairment in elders. In *1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, 2006. D2H2.*, pages 75–78. IEEE.

Kemper, S., Greiner, L. H., Marquis, J. G., Prenovost, K., and Mitzner, T. L. (2001). Language decline across the life span: Findings from the Nun Study. *Psychology and Aging*, 16(2):227–239.

Knapp, M., Prince, M., Albanese, E., Banerjee, S., Dhanasiri, S., Fernandez, J., Ferri, C., Snell, T., and Stewart, R. (2007). Dementia UK: report to the Alzheimer's Society. *King's College London and London School of Economics and Political Science*.

Le, X., Lancashire, I., Hirst, G., and Jokel, R. (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Literary and Linguistic Computing*, 26(4):435–461.

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., and Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699.

Ohsaki, M. and Sato, Y. (2002). A rule discovery support system for sequential medical data, in the case study of a chronic hepatitis dataset. In *Proceedings of the International Workshop on Active Mining (AM '02) in International Conference on Data Mining (ICDM '02)*, pages 97–102. IEEE.

Orimaye, S. O., Tai, K. Y., Wong, J. S., and Wong, C. P. (2015). Learning linguistic biomarkers for predicting mild cognitive impairment using compound skip-grams. In *Proceedings of the 2015 NIPS Workshop on Machine Learning in Healthcare (MLHC), Montreal, Canada*.

Pachidi, S., Spruit, M., and Van De Weerd, I. (2014). Understanding users' behavior with software operation data mining. *Computers in Human Behavior*, 30:583–594.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4):519–549.

Sawyer, P., Sutcliffe, A., Rayson, P., and Bull, C. (2015). Dementia and social sustainability: challenges for software engineering. In *37th International Conference on Software Engineering (ICSE '15)*, Florence, Italy. IEEE.

Seelye, A., Hagler, S., Mattek, N., Howieson, D. B., Wild, K., Dodge, H. H., and Kaye, J. A. (2015). Computer mouse movement patterns: A potential marker of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(4):472–480.

Shin, A. M., Lee, I. H., Lee, G. H., Park, H. J., Park, H. S., Yoon, K. I., Lee, J. J., and Kim, Y. N. (2010). Diagnostic Analysis of Patients with Essential Hypertension Using Association Rule Mining. *Healthcare Informatics Research*, 16(2):77–81.

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *JAMA*, 275(7):528–532.

Stringer, G., Sawyer, P., Sutcliffe, A., and Leroi, I. (2015). From Click to Cognition. In Davide Bruno, editor, *The Preservation of Memory*, chapter 5, pages 93–103. Psychology Press.

Sutcliffe, A., Rayson, P., Bull, C., and Sawyer, P. (2014). Discovering Affect-Laden Requirements to Achieve System Acceptance. In *Proceedings of the 22nd IEEE International Requirements Engineering Conference (RE'14)*, pages 173–182, Karlskrona, Sweden. IEEE.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.