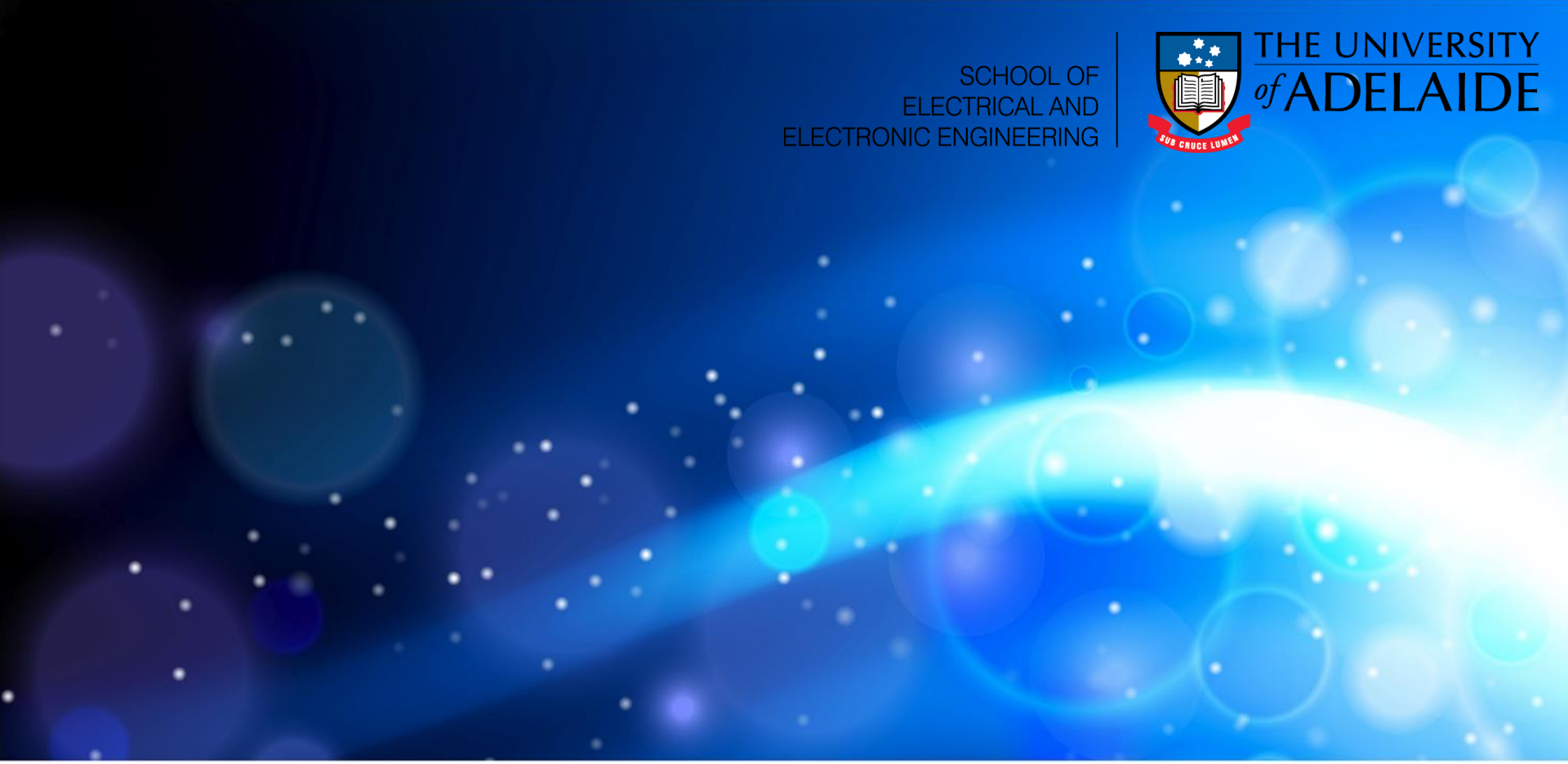
SCHOOL OF
ELECTRICAL AND
ELECTRONIC ENGINEERING

THE UNIVERSITY of ADELAIDE



Students: Peter Roush and Bryce Shi

Supervisors: Derek Abbott, Maryam Ebrahimpour and Brian Ng

Project 44: Cracking the Voynich Code

Reproduced from YouTube /watch?v=FRqs-5tImrI under ‘fair dealing’, The Copyright Act 1968

History and Details

Voynich Manuscript

Wilfred Michael Voynich

Reproduced from Wikimedia Commons Michał Wojnicz c. 1885

- Born in Kovno, Lithuania on October 31, 1865
- Part of the Polish nationalist movement
- Captured an imprisoned in 1885 for 2 years, escaped to London
- Married to Ethel Boole in 1902, later Ethel Voynich
- Business manager selling revolutionary books
- In 1912 he found the Voynich Manuscript buried in a chest at a castle in Italy

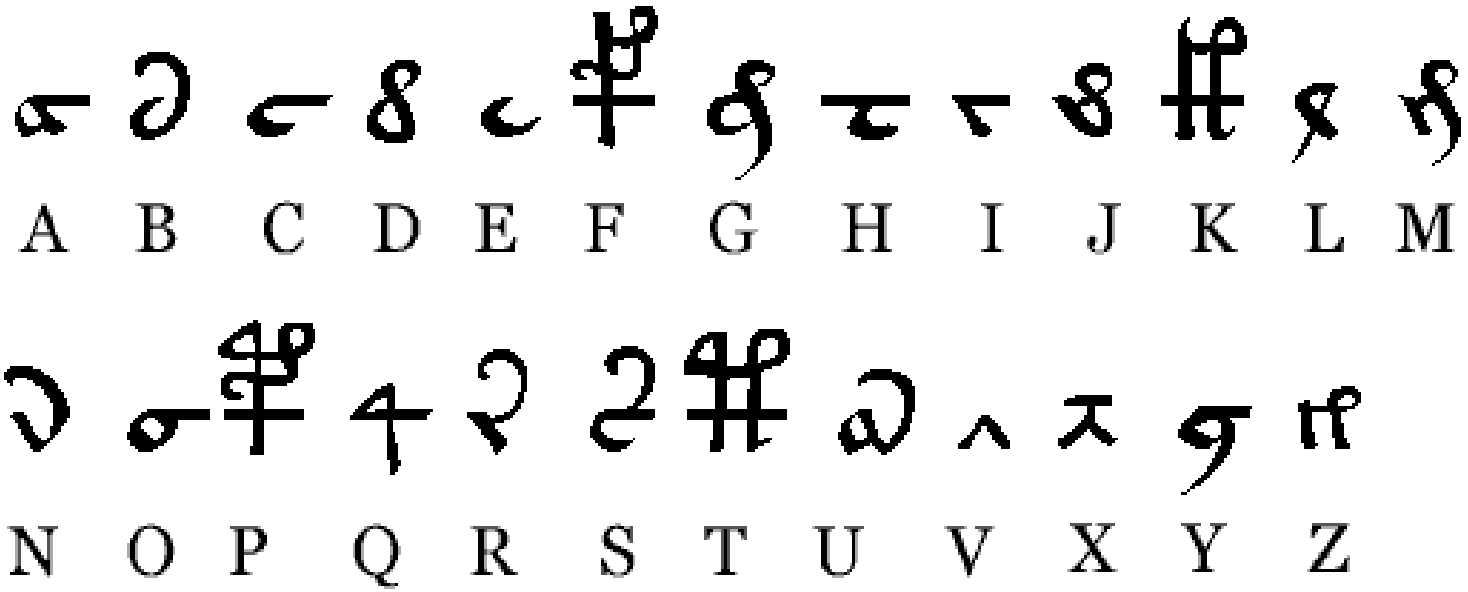
Anne Fremantle, "The Russian Best-seller", *History Today*, Vol. 25 Issue 9 (September 1975), p. 629-637

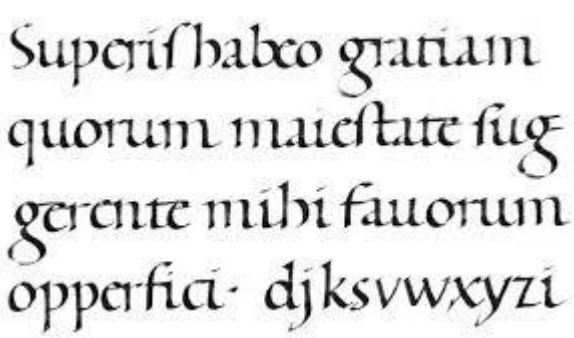
Voynich Manuscript

- **Part 1** (Botanical/Herbal)
  folios 1r - 66v
- **Part 2** (Astronomical)
  folios 67r – 73v
- **Part 3** (Biological)
  folios 75r – 84v
- **Part 4** (Cosmological)
  folios 85r – 86v
- **Part 5** (Pharmaceutical)
  folios 87r – 102v
- **Part 6** (Recipes)
  folios 103r – 116v

Detailed chemical analysis can be found at Yale:
http://beinecke.library.yale.edu/sites/default/files/voynich_analysis.pdf

Characters

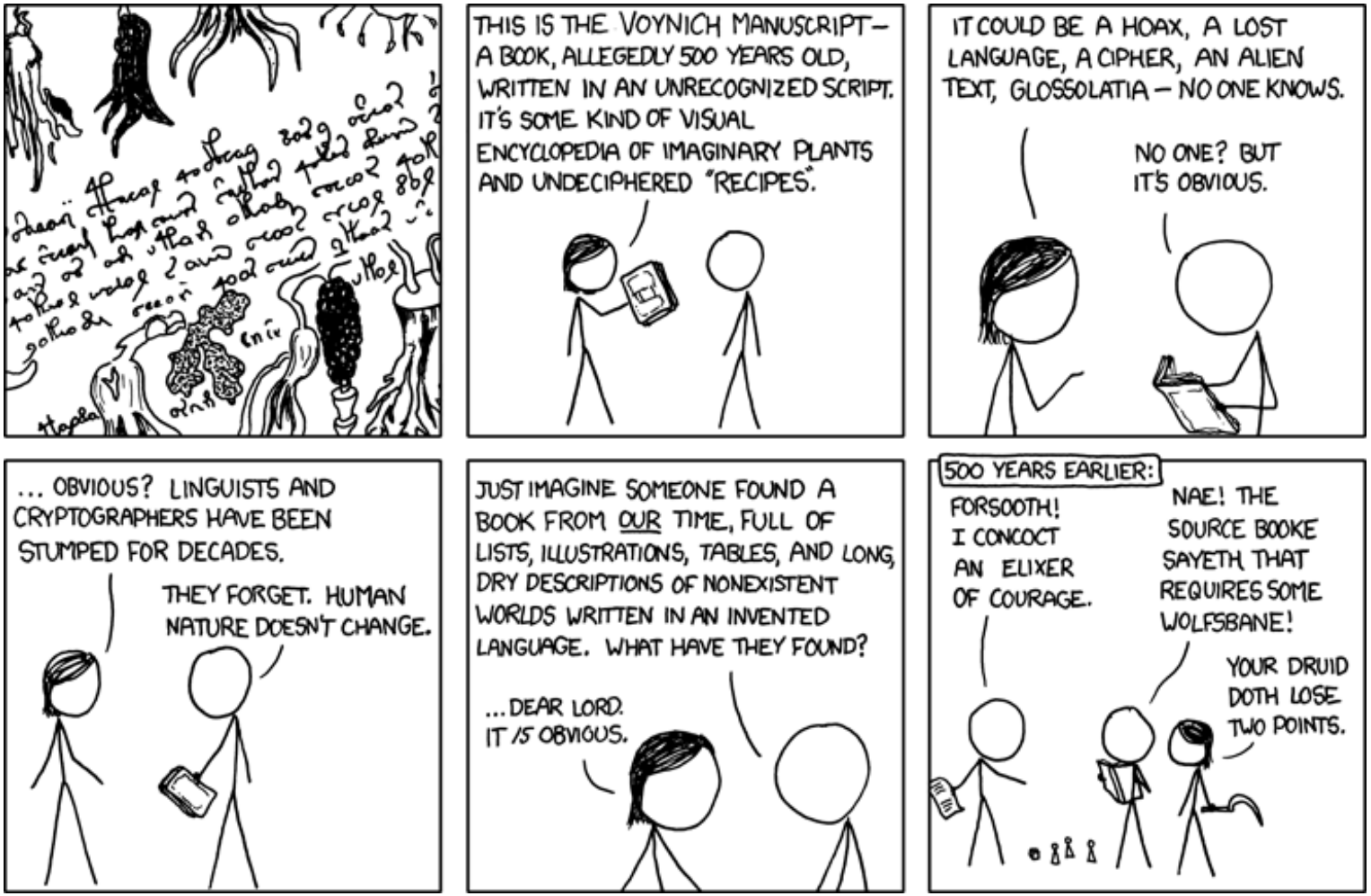




Humanist miniscule writing (left)

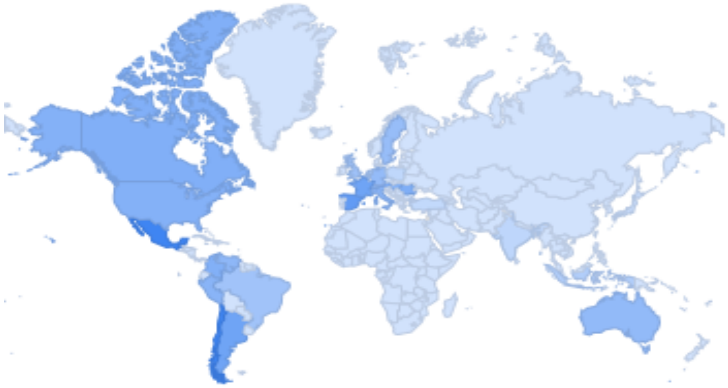
Picture from http://www.afternight.com/runes/a-voynich.gif and dictionarytoday.tumblr.com

The Voynich Mystery



Google Trends

Google trends of the word ‘Voynich’ since 2004 to current



Reproduced from Google Trends using the key ‘voynich’

Region | Town/City

Chile	100
Mexico	100
Spain	76
France	67
Argentina	67
Austria	60
Romania	56

Historical and Modern Methods

Decoding The Manuscript

Traditional methods

- Word substitution
  - Traditional word substitutions
- Key based cipher
  - Theory that the last page f117v holds the key
- Word duplications
  - Repetitions of the same word, strokes
- Glossolalia
  - Speaking in tongues, or channelling outsider art
- Steganography
  - Meaningful information hidden behind inconspicuous details
  - Number of lines
  - Every alternating letter

Picture analysis

Reproduced from Voynich Attacks, https://voynichattacks.wordpress.com/

- Image of *folio 33v* (Botanical/Herbal)
- Left side of the sun flower writing is more distinct than the right
- Possibly left written first, then middle then to the right of the stalk
- Writing based around picture?
- No reason as to why that's the case

Computational comparisons

- Analysis using Landini's interlinear file in EVA (v 1.6e3)
- Several complete and partial transcriptions exist
- **A**: *folios 1r – 25v* (Botanical/Herbal)
- **B**: *folios 75r – 84v* (Biological)
- 2900 unique words in total, ~1600 in **A** and ~1630 in **B**

Words common in Language A, but rare in Language B

Word	# in Language A	# in Language B
	61	1
	20	1
	14	1
	28	2
	13	1
	13	1
	13	1
	13	1
	11	1
	21	2

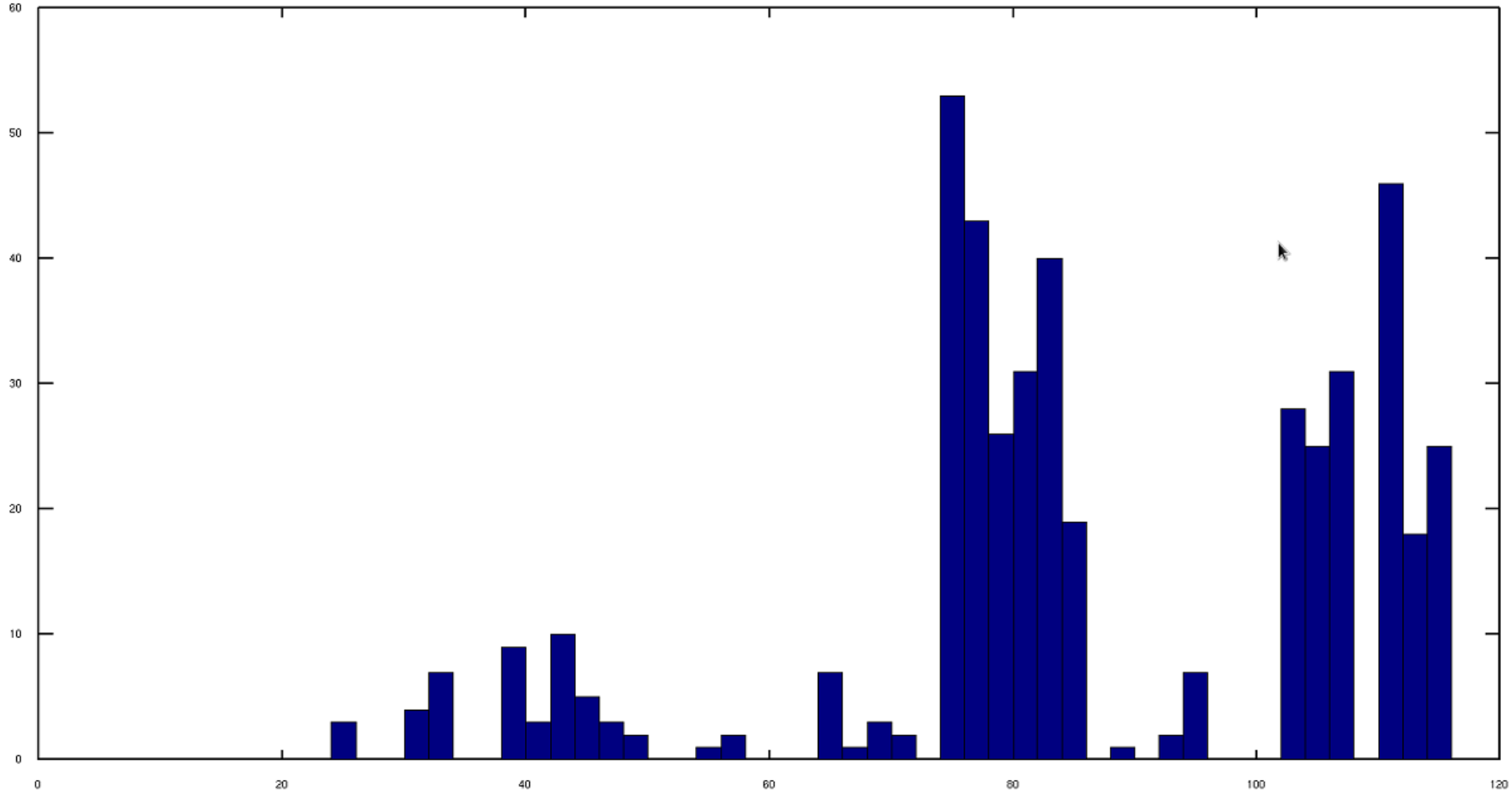
Words common in Language B, but rare in Language A

Word	# in Language B	# in Language A
	159	1
	112	1
	107	1
	50	1
	28	1
	21	1
	81	4
	40	2
	18	1
	31	2

Reproduced from Voynich Attacks, https://voynichattacks.wordpress.com/

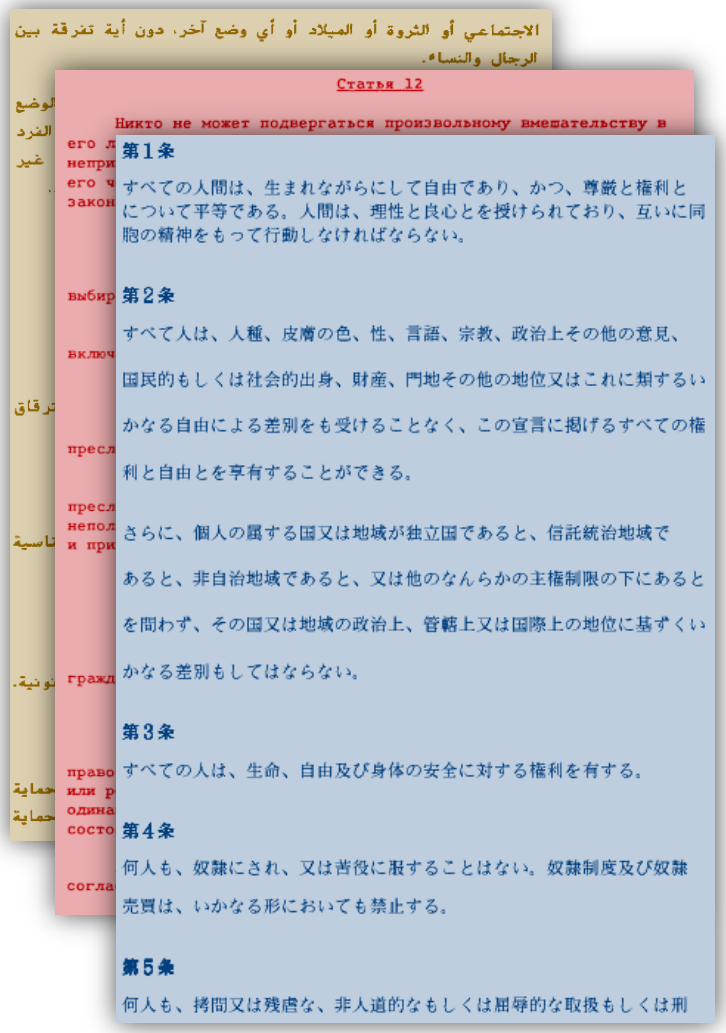
Characterising the Text

Figure 1: Number of Word Occurences Per Page -- "chedy"



The UN Declaration of Human Rights

- Translated into over 400 languages
- Modern Rosetta Stone
- Provides easy method of language comparison
- Allows analysis to be language independent



Picture Reproduced From: www.boes.org (Public Domain)

Word Recurrence Interval (WRI)

- WRI is defined as the number of words in between successive occurrences of a keyword

- Keyword being: I

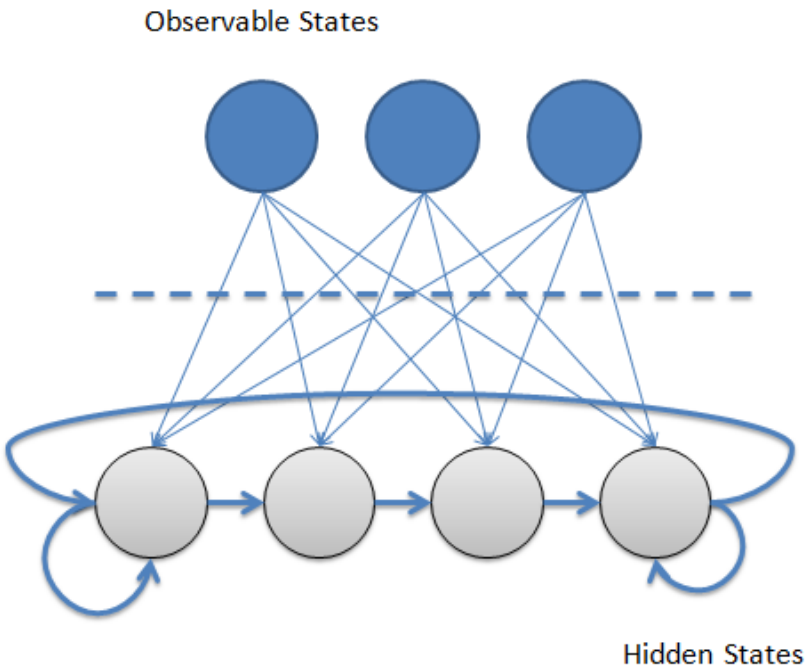
1 2 3 4 5 6 7 8 9 10 11

I have six locks on my door all in a row. When I go out, I lock every other one. I figure no matter how long somebody stands there picking the locks, they are always locking three.

- Word Recurrence interval is: {0, 11, 2, 4}

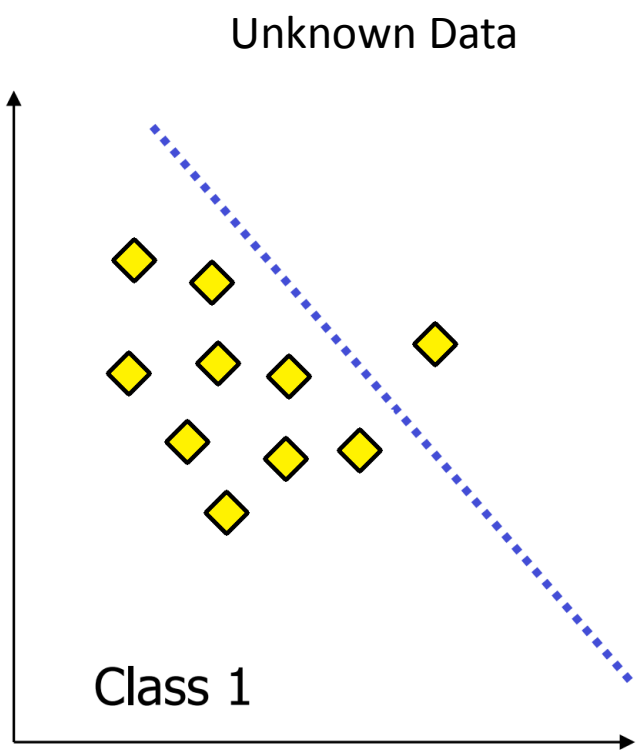
Hidden Markov Model (HMM)

- To infer the most likely sequence of words that produced a given word sequence
- Infer which will be the most likely next state (and thus predicting the next word)
- Calculate the probability that a given sequence of words originated from a book, comparison check
- Use a HMM to explore the syntax of the language



Pictures reproduced from http://www.codeproject.com/Articles/69647/Hidden-Markov-Models-in-C

Support Vector Machine (SVM)

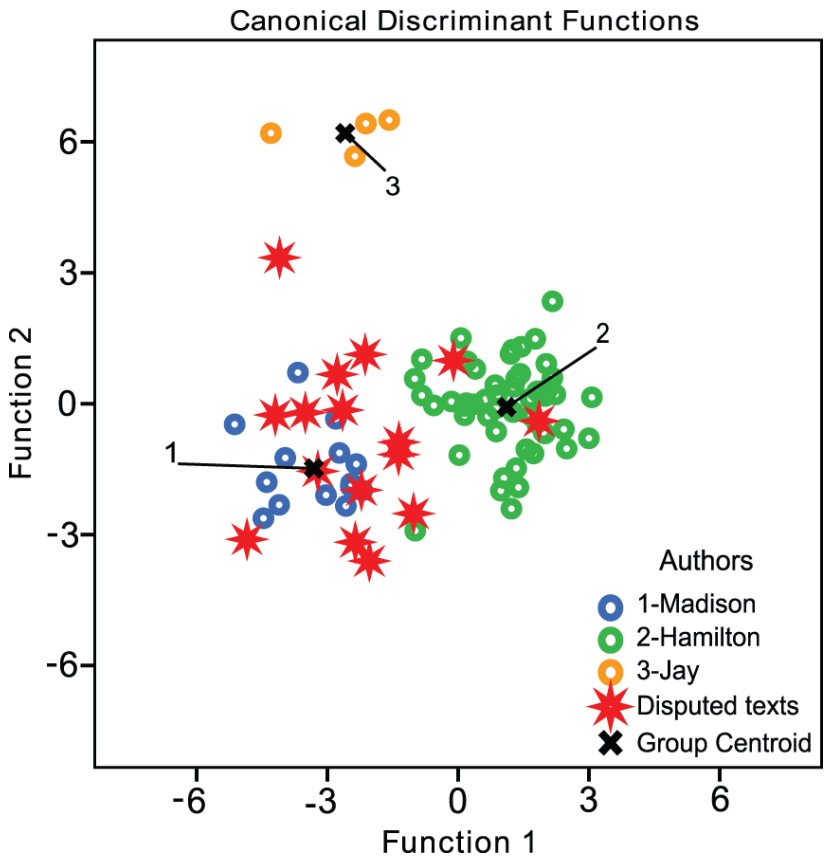


- SVM is a binary classifier
- Defines a decision point from a set of training data which is split into two distinct classes
- Assigns new testing data into one of those classes based on the decision point.
- Can be used for authorship detection

Picture Modified From: Martin Law, 3/1/11, http://www.cise.ufl.edu/class/cis4930sp11dtm/notes/intro_svm_new.pdf
Reference: Ebrahimpour M, Putniņš TJ, Berryman MJ, Allison A, Ng BW-H, et al. (2013) Automated Authorship Attribution Using Advanced Signal Classification Techniques. PLoS ONE 8(2): e54998. doi:10.1371/journal.pone.0054998

Multiple Discriminant Analysis (MDA)

- Reduces differences between data points in order to classify them into N broad clusters
- Using this training data, classifies unknown data into one of the N clusters based on the proximity to the centre of that cluster
- Important to maximise differences between clusters



Reference and Picture from: Ebrahimpour M, Putniņš TJ, Berryman MJ, Allison A, Ng BW-H, et al. (2013) Automated Authorship Attribution Using Advanced Signal Classification Techniques. PLoS ONE 8(2): e54998. doi:10.1371/journal.pone.0054998

THE UNIVERSITY of ADELAIDE

Risks and Management, Budgeting, Timeframes and Deliverables

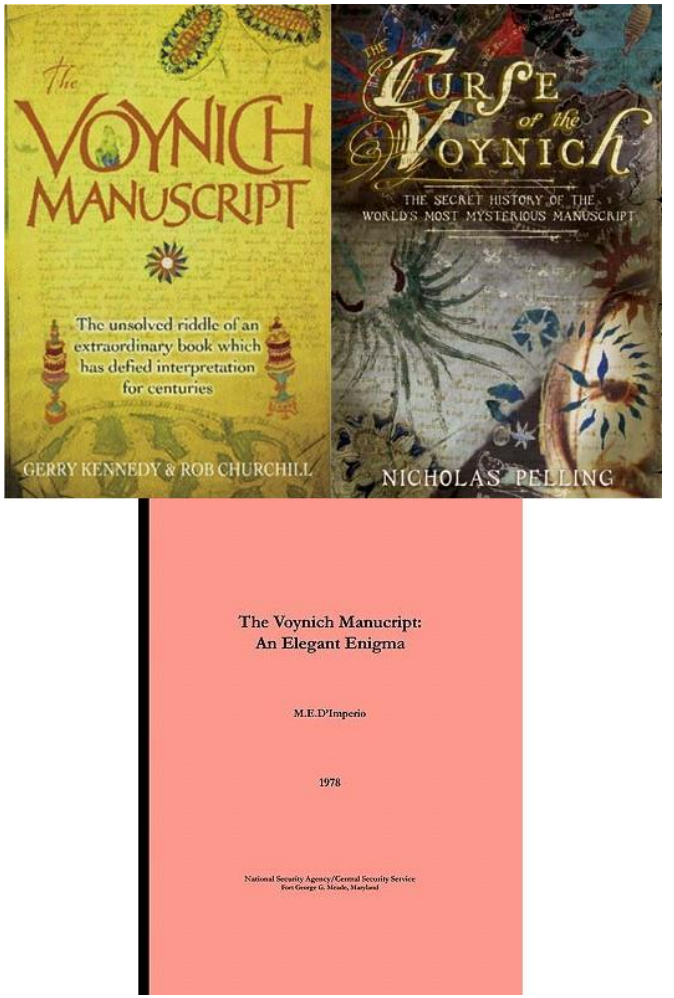
Project Management

Risk Assessment and Management

No.	Risk	Likelihood	Consequence	Risk Level
1	Inaccurate allocation of time and resources to a particular area	Likely	Moderate	High
2	Health issues due to long periods of time sitting and working at a PC	Likely	Moderate	High
3	Files and working copies lost	Rare	Major	Medium
4	Not understanding the project correctly and the processes required	Almost Certain	Negligible	Medium
5	UofA Electrical Engineering server down for unknown reasons	Unlikely	Moderate	Medium
6	Not being able to solve the Voynich Manuscript code	Almost Certain	Negligible	Medium

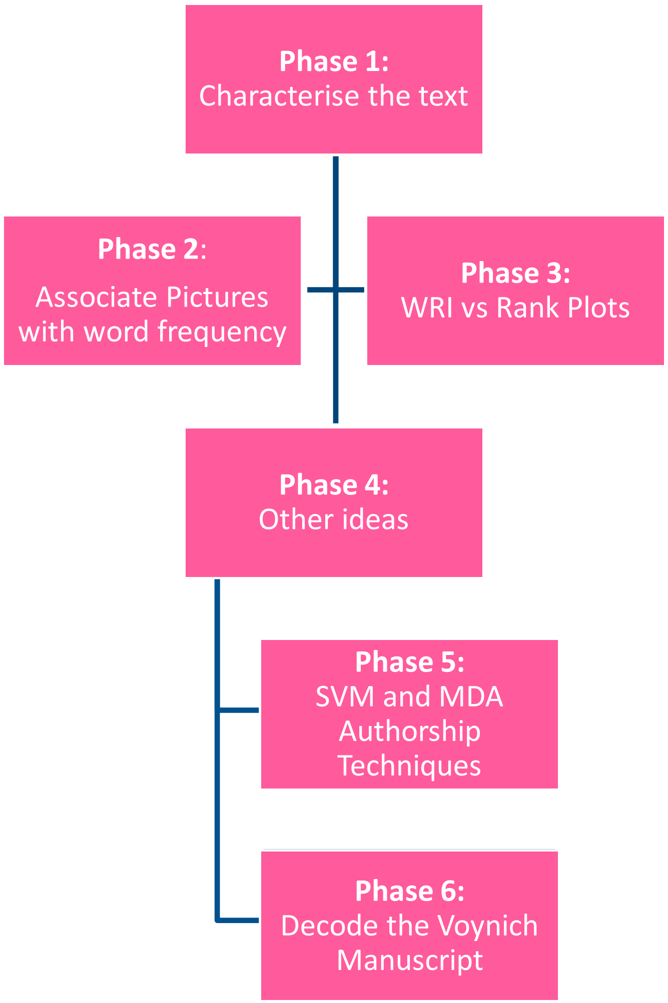
Budget

- Project is primarily based in mathematical methods and computer analysis
- No equipment or components need to be purchased
- Money allocated to three reference books instead for background information and theories
- $348.24



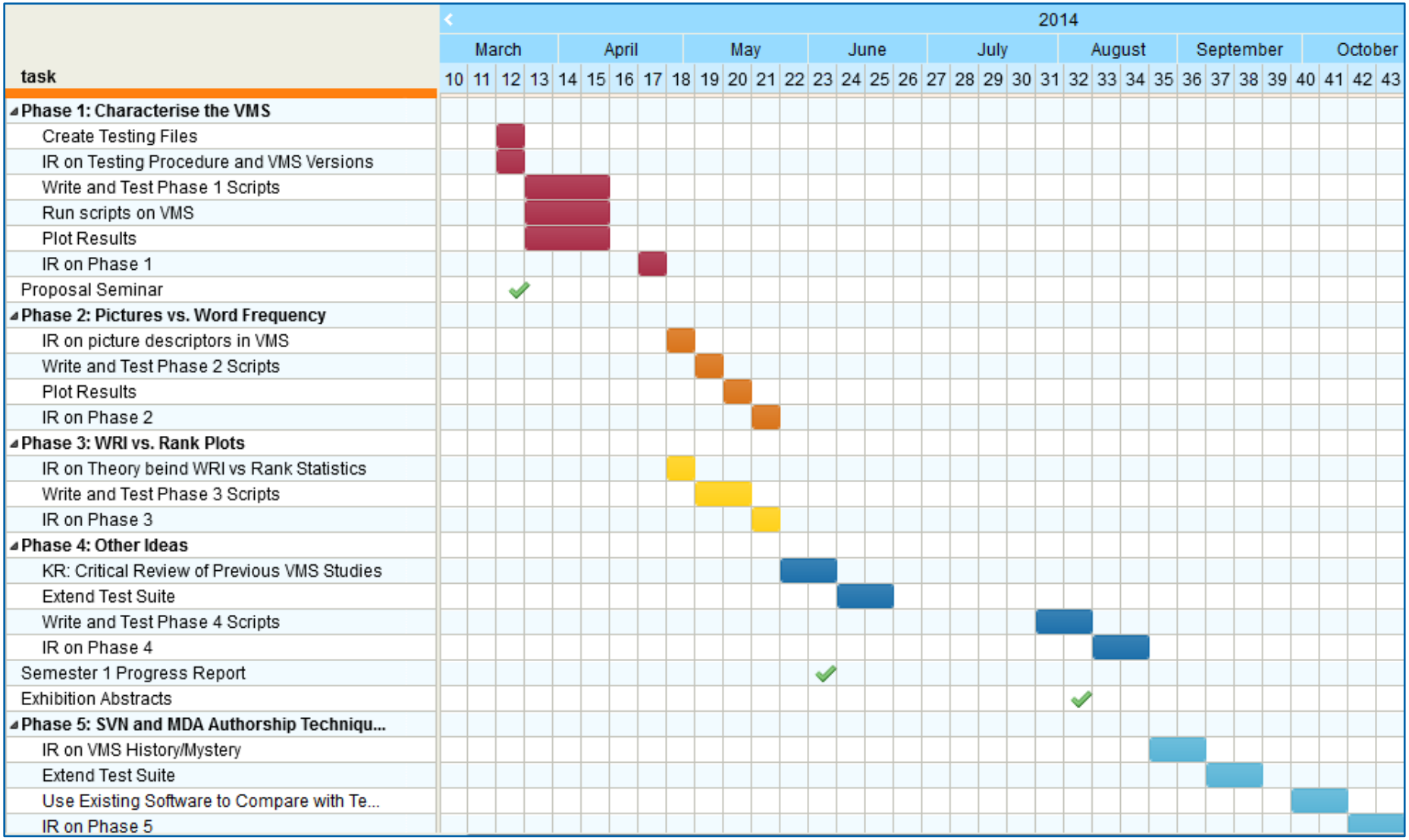
Reproduced from Amazon.com

Proposed Approach

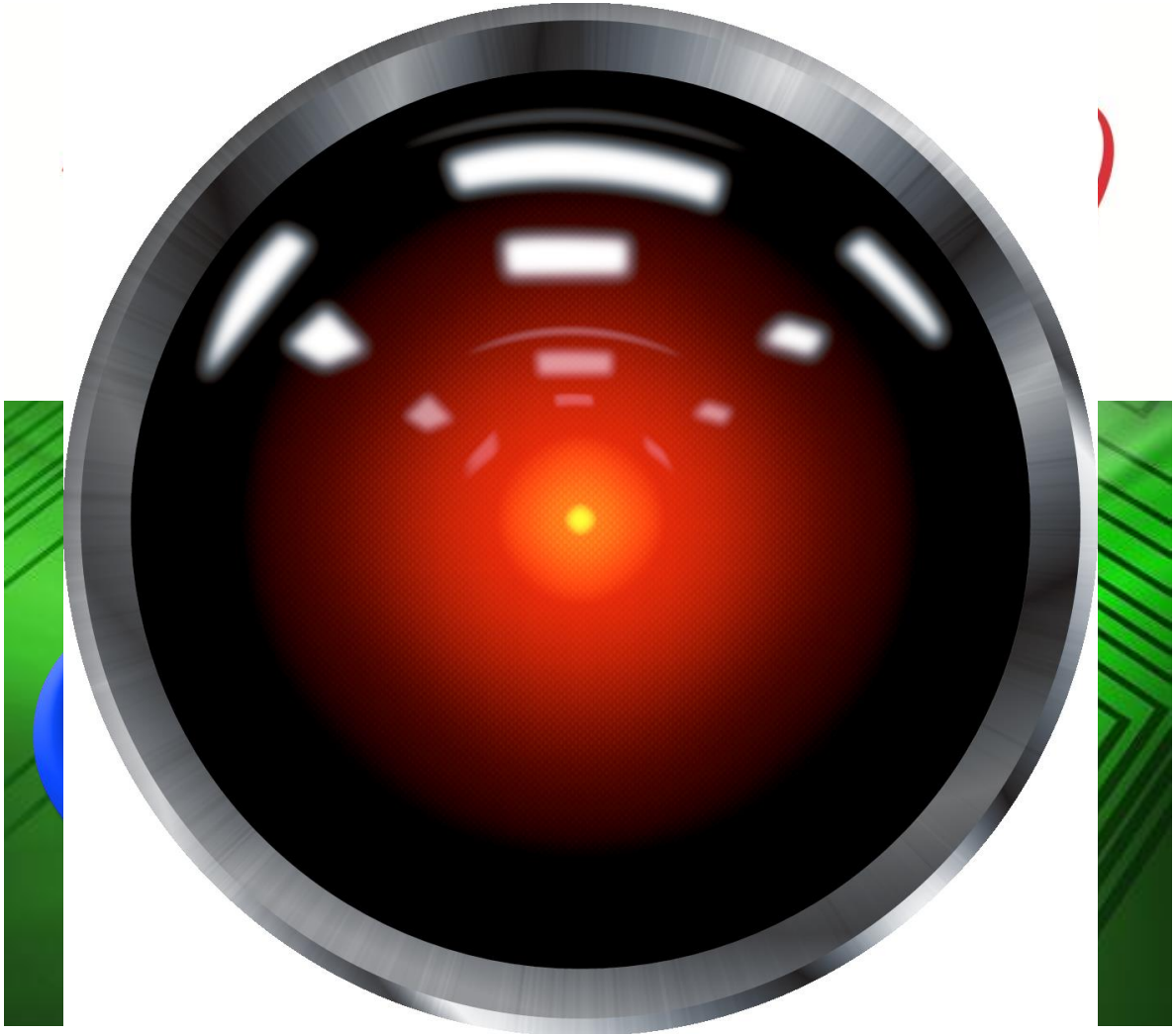


- Rank words in [illegible]
- Convert Voynich Manuscript into an easily searchable format [illegible]
- Write a general descriptor for the each of the [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible] Voynich Manuscript

Gantt Chart


2014
March
April
May
June
July
August
September
October
task
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
Phase 1: Characterise the VMS
Create Testing Files
IR on Testing Procedure and VMS Versions
Write and Test Phase 1 Scripts
Run scripts on VMS
Plot Results
IR on Phase 1
Proposal Seminar
Phase 2: Pictures vs. Word Frequency
IR on picture descriptors in VMS
Write and Test Phase 2 Scripts
Plot Results
IR on Phase 2
Phase 3: WRI vs. Rank Plots
IR on Theory beind WRI vs Rank Statistics
Write and Test Phase 3 Scripts
IR on Phase 3
Phase 4: Other Ideas
KR: Critical Review of Previous VMS Studies
Extend Test Suite
Write and Test Phase 4 Scripts
IR on Phase 4
Semester 1 Progress Report
Exhibition Abstracts
Phase 5: SVN and MDA Authorship Techniqu...
IR on VMS History/Mystery
Extend Test Suite
Use Existing Software to Compare with Te...
IR on Phase 5

Engineering Connections



Reproduced from Various Sources: HAL9000, Google, Turnitin and Fingerprint scanner


THE UNIVERSITY
of ADELAIDE
End of presentation
Questions?