



THE UNIVERSITY  
*of* ADELAIDE



# 2015 Honours Project Cracking the Voynich Code

Honours Project ID: 31

Supervisors: Derek Abbott, Brian Ng and Maryam Ebrahimpour

Students: Andrew McInnes and Lifei Wang

# Outline

- History, background and objective
- Detail of Voynich manuscript
  - Layout, content
- Analysing and decoding
  - Technique, method
- Proposed Approach
  - Phases
- Project management
  - Risk, schedule, work breakdown and budget
- Engineering Connections and Applications
- Conclusion





# History & Background

## Voynich Manuscript

# Voynich Manuscript



Reproduced from Internet Archive <https://archive.org/details/TheVoynichManuscript>



# Wilfred Voynich

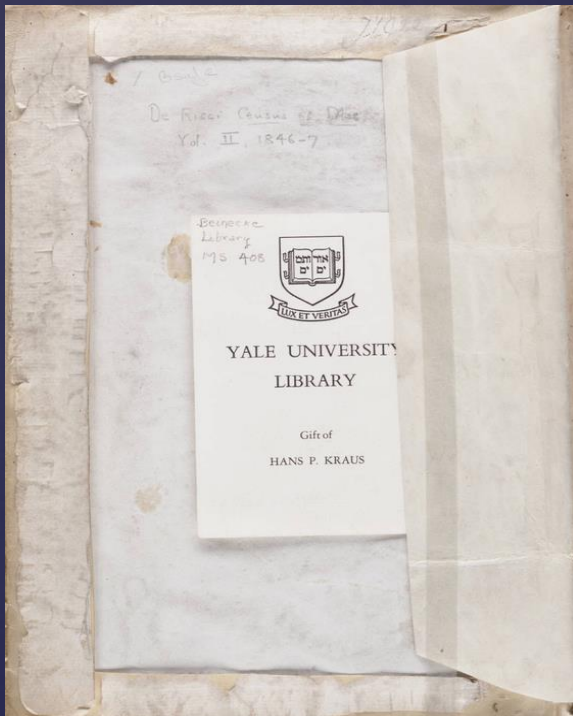
- Born in Telsze, a town in Russian Empire on 31 October 1865
- Revolutionary, Antiquarian and Book Dealer
- He found the manuscript in an ancient castle in Southern Europe in 1912
- He took the Voynich Manuscript to London in 1912, and later in 1914 to the United States.



Rene Zandbergen. (2014). The Voynich MS - General Introduction. Accessed on 22/03/15 from <http://www.voynich.nu/intro.html>

# Beinecke Rare Book and Manuscript Library

- Hans P. Kraus bought the Voynich Manuscript in 1961
- It was donated to Yale University in 1965.
- Housed at the Beinecke Rare Book and Manuscript Library
- Official register number MS 408.



Label of Voynich manuscript in Yale University



Beinecke Rare Book and Manuscript Library

# Objective

- Separate the alphabet from other tokens of the Voynich manuscript.
- Compare linguistic features of the Voynich Manuscript and other languages.
- Determine whether the language in the Voynich Manuscript is cipher, codes, natural language, constructed language or hoax.
- Determine possible authors of the Voynich Manuscript.





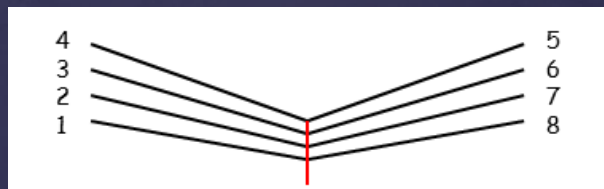
# Details

## Voynich Manuscript

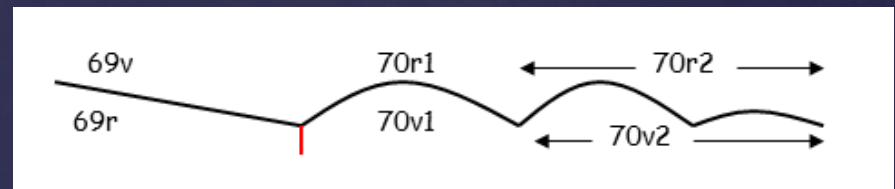


# Layout of the Manuscript

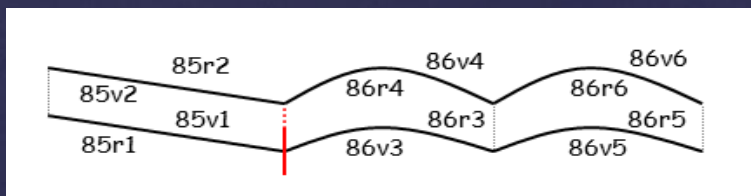
- The manuscript is made up many folios, numbered from f1 to f116
- Each folio consists two pages, labelled r and v.
- There are gaps between the folio number , which indicate missing folios.



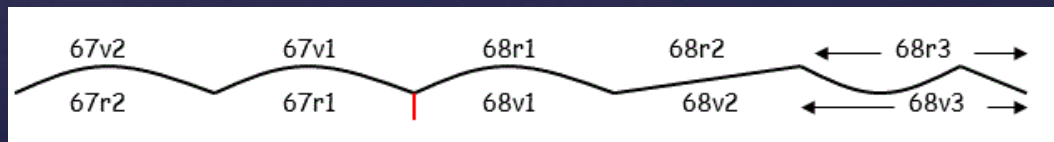
Folio 1-66, 73-84, 108-116



Folio 69-72, 85-90, 93-94, 99-102



Folio 85-86



Folio 67-68

Rene Zandbergen. (2014). The Voynich MS - General Introduction. Accessed on 22/03/15 from <http://www.voynich.nu/intro.html>

# Sections

- Herbal section

- Few text with herbal pictures in each page
- Folios 1r - 66v



- Astronomical section

- Contain pictures of suns, moons, stars and astronomy
- Folios 67r - 73v



Reproduced from Internet Archive <https://archive.org/details/TheVoynichManuscript>



# Sections

- Biological section
  - Text with figures, most of the figures are naked women
  - Folios 75r - 84v



- Cosmological section
  - Has foldouts with circular diagrams
  - Folios 85r – 86v



Reproduced from Internet Archive <https://archive.org/details/TheVoynichManuscript>

# Sections

- Pharmaceutical section

- A few text paragraphs with many isolated plant parts
- Folios 87r - 102v



- Recipes section

- Full pages of text paragraphs
- Folios 103r - 116v







# Analysis & Decoding

## Voynich Manuscript

# Analysis of the Illustrations

- The illustrations could be the breakthrough point to analysis the Voynich manuscript.
- Illustrations in folio 102 r2 (pharmaceutical section), There is a frog in the top right corner.

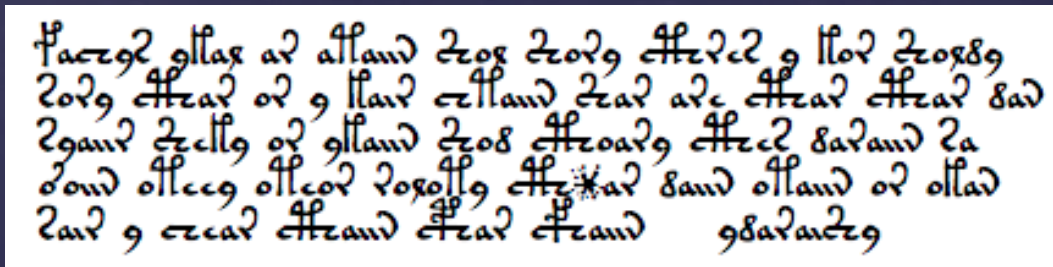


Rene Zandbergen. (2014). The Voynich MS - General Introduction. Accessed on 22/03/15 from <http://www.voynich.nu/intro.html>

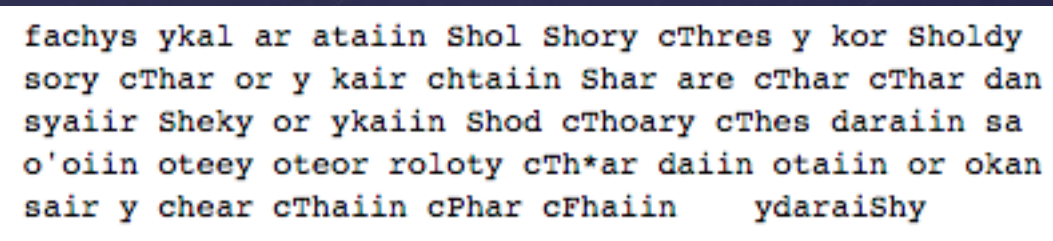


# Interlinear Archive

- Convert text to alphabet
- There are 16 transcriptions of Voynich manuscript
- For example, if we use European Voynich Alphabet (EVA)

A snippet of the original Voynich manuscript, showing several lines of text written in a highly stylized, cursive script with various ligatures and unique characters.

Original Voynich manuscript

The same text from the original manuscript, but transcribed using the European Voynich Alphabet (EVA) True Type font. The characters are now standard Latin letters, making the text legible.

Convert using the EVA True Type font

Rene Zandbergen. (2014). The Voynich MS – Analysis of the text. Accessed on 22/03/15 <http://www.voynich.nu/analysis.html>

# Language Comparison

- The UN Declaration of Human Rights has been translated to over 400 languages
- Compare the Voynich manuscript language with other known languages
- Feature:
  - Total word, word frequency, token frequency, word length, etc.





# Zipf's Law

- In natural language, the frequency of a word is inversely proportional to its ranking of frequency.
- Zipf's law can be used to determine if a given language is a natural language.
- Number of the most frequency word is as twice as the second most frequency word, and as 3 times as the third most frequency word.

Powers, David M W (1998). "Applications and explanations of Zipf's law". Association for Computational Linguistics. pp. 151–160.



# Proposed Approach

The project will be completed in multiple phases



# All Phases - Research

- Looking at how statistics can be used to analyse linguistics
- Focus study on the Voynich Manuscript
- Each phase is based around basic principles of linguistics
- Conducted throughout all phases by both members
- Knowledge to be obtained through the research and through empirical data obtained in each analysis of the phases.

# Phase 1 – Characterization of Text

- Write C++ and MATLAB codes to count the basic features of a given text.
  - Number of words
  - Number of characters
  - Frequency of specific words
  - Frequency of specific characters
  - Tokens that only appear at the start, middle or end
- Comparing the features of Voynich with known languages.



# Phase 2 – English Investigation

- Characterise the English Language
  - Expand the same characterisation code on English
- Investigate how the statistics can be used to extract the alphabet
  - How can we determine what an alphabet symbol is when compared with a punctuation symbol?
- Develop code to extract the English alphabet from text
  - Assume no knowledge of the English alphabet

# Phase 3 – Morphology Investigation

- Morphology deals with the structure of morphemes
- Does the Voynich contain words that are morphemes?
  - alal, alain, alam, ...
- Research and analyse different languages and their representation of morphemes
- Develop code to extract possible morphemes from the Voynich Manuscript



# Phase 4 – Stylometry Investigation

- Deals with ‘linguistic style’ of written language
- Compile corpus from authors of the 15<sup>th</sup> Century
- Average Word Lengths
- Distribution of Word Lengths
- Lexical Richness
- Compare against the Voynich Manuscript

# Phase 5 – Other Ideas

- Keywords and Co-Occurrence (Information Theory)
- Vowel and Consonant Representation
- Zipf's Law
- Word Order
- Follow-on from previous year
  - Hidden Markov Models
  - 15<sup>th</sup> Century Cipher Analysis



# Project Management

Voynich Manuscript



# Risk Management

Risks identified that could affect the project.

Very Low	1 to 2
Low	2 to 3
Moderate	4 to 6
High	7 to 8
Very High	9 to 10

No.	Risk	Likelihood	Consequence	Risk Level
1	<b>Underestimation and/or mismanagement of time and resources</b>	<b>High [8]</b>	<b>High [7]</b>	<b>56</b>
2	Health related issues from long periods at computers	High [7]	Moderate [6]	42
3	Team member illness or injury	Very High [9]	Moderate [4]	36
4	Issues with communication between team and/or supervisors	Low [3]	High [7]	21
5	Loss of software code and/or files	Low [2]	Very High [10]	20

# Schedule



<b>Earliest Finish</b>	<b>Task - Milestone</b>	<b>Latest Finish</b>
Week 4, Semester 1	Phase 1 – Characterization of Test	Week 5, Semester 1
Week 6, Semester 1	Phase 2 – English Investigation	Week 7, Semester 1
Week 9, Semester 1	Phase 3 – Morphology Investigation	Week 11, Semester 1
Week 9, Semester 1	Phase 4 – Stylometry Investigation	Week 11, Semester 1
Week 6, Semester 2	Phase 5 – Other Ideas	Week 8, Semester 2

# Work Breakdown

- Andrew
  - Code for counting the features of the Voynich and comparing with other known languages. (Phase 1)
  - Code for separating the English alphabet from other tokens. (Phase 2)
  - Phase 3 – Morphology Investigation
  - Phase 5 – Collaboration, determine follow-ups from current findings
- Lifei
  - Finding tokens that only appear at the start of words and which are only at the end. (Phase 1)
  - Counting the features of token (Phase 2)
    - token frequency, token recurrence interval, etc.
  - Phase 4 – Stylometry Investigation
  - Phase 5 – Collaboration, determine follow-ups from current findings



# Budget

- Budget of \$500
- Vast majority of work computer-based
- All required programs available on University systems
- May be used to further our research.



# Engineering Connections and Applications

What does this have to do with Engineering?

# Engineering Connections

- Probability and Statistics
- Information Theory
- Data Mining
- Encryption and Decryption
- Computational Linguistics



[http://www.cbionm.com/wp-content/uploads/2012/07/tumblr\\_chryk\\_300dpi.jpg](http://www.cbionm.com/wp-content/uploads/2012/07/tumblr_chryk_300dpi.jpg)  
<https://afontgeek.files.wordpress.com/2012/05/counterparts.jpg>





# Conclusion

# Conclusion

- Objective is not, necessarily, to solve the Voynich Manuscript
  - Obtain knowledge on linguistic features and compare with Voynich
- Majority will be computer-based following the outlined phases
- Research to be done throughout the life of the project





**Do you have any questions?**