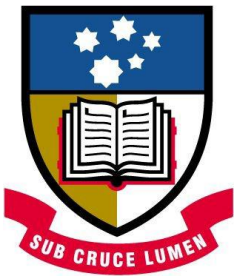

THE UNIVERSITY OF ADELAIDE

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

ADELAIDE, SOUTH AUSTRALIA, 5000

141: Cracking the Voynich manuscript code (The first draft)

Ruihang Feng

ELEC ENG 7076 A/B MASTERS PROJECT

Date submitted: 22 April 2016

Supervisor: Prof. Derek Abbott

Adviser: Dr. Brian Ng

Signature of Supervisor:

Abstract

The Voynich manuscript, written by a kind of unknown language, is a famous mystery in the field of linguistic. Over the years, many researchers had tried to crack this manuscript, but no one had succeeded. In this thesis, the project of cracking the Voynich manuscript will be introduced from five aspects:

- Introduction.
- Related work (the history of the Voynich manuscript research).
- Requirement.
- Proposed method.
- Project management.

Content

1. Introduction

1.1 Background

The Voynich manuscript is a document written in unknown alphabets that was found by Wilfrid Voynich (1865-1930) in 1912 [1]. Because of the Voynich manuscript's long history, some pages of manuscript were missing. As the result, there are almost 240 pages remaining [2]. In addition, the folios of the manuscript were numbered from f1 to f116 and each folio involved two pages, r and v.

There are six sections in the Voynich manuscript: herbal, astronomical, cosmological, biological, pharmaceutical and recipes section. Generally, the Voynich manuscript was made up of three parts: text, illustrations and marginal symbols.

1.2 Aim

The aim of project is using the statistics and comparison to infer that the Voynich manuscript is code, nature languages, constructed languages, cipher code or hoax.

Due to the massive number of words and illustrations in the manuscript, it is unnecessary to solve the whole manuscript in a one year project.

1.3 Motivation

In the field of linguistics, the Voynich manuscript is a representative. Researchers deem that there is a kind of useful information among the mysterious alphabets of manuscript.

In the course of this project, statistics and comparison will be applied to crack the Voynich manuscript. If the manuscript can be cracked successfully, the result of this project will be useful for linguists to compare other unknown languages.

1.4 Significance

There are many guesses about the Voynich manuscript. Because of the manuscript's long history, many historians believe that the mysterious alphabets of the Voynich

manuscript are related to ancient civilizations [3]. If manuscript can be cracked, the Voynich manuscript will be helpful for historians to explore the culture of ancient society.

In addition, the statistical method which will be used in this project is also useful in other fields, such as engineering, finance and architecture. Moreover, comparison is widely used, such as Turn-It-In, Google translate, Grammarly and Bing.

1.5 Technical Background.

The major technique which will be applied in this project is data mining. Data mining is an effective method to search laws among the massive number of data and has a fantastic performance. The two major methods of data mining are statistics and comparison. Statistics is used to count the frequency of the occurrence of some special words. Comparison is served to find out relations between two languages.

In the field of linguistics, European Voynich Alphabet (EVA) is a representative digital transcription of the Voynich manuscript [4]. Therefore, major data will be extracted from EVA in the process of this project.

Moreover, other resources will be considered, such as expressions of some representative ancient languages.

1.6 Knowledge Gaps

Due to the massive number of data in the Voynich manuscript, the project requires skilled data processing technique and software programming capabilities, however, no one in this project team has ever dealt with so many data. Hence members should develop data processing ability and software programming skills.

On the other hand, the project requires particular knowledge about statistics, so members must be adept at sorting data.

1.7 Technical Challenges

Technical challenges of this project involve two aspects.

First of all, it is very difficult to infer which language the author used. The language of the manuscript does not belong to any known languages [5] and even this language may have been extinct. What is more, due to the long history of the Voynich manuscript, some important information is nowhere to be searched, such as exact information about author. In that case, it is difficult to infer which language the author used from the author's nationality. In order to solve the above problem, members must search many different languages as references and compare those languages with the language of the manuscript.

Secondly, references of cracking the Voynich manuscript are limited. Because of unknown language and mysterious illustrations in manuscript, it is difficult to crack the whole manuscript. Although there are very few words have been cracked by researchers, on one can guarantee that the results are right. In the field of linguistics, there are not recognized correct results about cracking the Voynich manuscript. In that case, it is hard to find reliable references. So members must search references from different ways and find out enough accurate references.

2. Related Work (the history of the Voynich manuscript research)

In the past few years, many researchers had tried to crack the Voynich manuscript by using different methods.

Mary E. D'Imperio:

In 1975, Mary E. D'Imperio was introduced to the problem of the Voynich manuscript by John Tiltman [6]. In the following years, she summed up different features of the Voynich manuscript text [7].

Nick Pelling:

Nick Pelling published his book 'The course of the Voynich' at 2006. Based on the illustrations in the rosettes folio of the Voynich manuscript, he believed that the manuscript originated from Milan [8].

William Ralph Bennett:

William Ralph Bennett, a Yale professor, searched the Voynich manuscript with computer. He focused on the research of text by using statistical method. Probably he was the first to note the low entropy of the Voynich manuscript text. As the result, the only language he found with entropy similar to the Voynich manuscript was Hawaiian [9].

John Tiltman:

John Tiltman was a British intelligence specialist. He cracked the text part of the Voynich manuscript with William Friedman. At last, Tiltman and Friedman suggested that the text of manuscript was a kind of artificial (constructed) language [10].

Feely:

Joseph Martin Feely was a Rochester lawyer. In 1943, Feely published a book which involved some solutions of cracking the Voynich manuscript. His solutions showed a viable method to use Latin to replace some words in the manuscript [11].

First study group:

The first study group (FSG) was founded at 1944, dissolved at 1946 [12]. Members of this organization involve:

- Robert A.Caldwell

- G. E. McCracken
- Tomas A. Miller
- Frances Puckett, later Frances Wilbur
- Mark Rhoads
- William M. Seaman

The FSG transcribed most parts of the Voynich manuscript and devised a transcription alphabet [13]. The details of the transcription alphabet are as shown in the A.7 figure7.

3. Requirements.

Although it is not necessary to crack the whole manuscript, there are some basic requirements as following:

- Text investigation: find out linguistic laws from some paragraphs of the Voynich manuscript.
- Illustration research: look for laws from some illustrations.
- Marginal symbols investigation: make a thorough inquiry about laws from marginal symbols.
- Code run smoothly.
- Evaluation for results.
- Make some assumptions which are helpful for the further research.

4. Proposed Method

As shown in the A.1 figure 1, the proposed methods of this project are divided into four phases.

4.1 Phase 1: Text investigation.

There are two parts in this phase: words and digits.

During the process of words research, Matlab will be used as an essential tool. Team members will attempt to search laws from three aspects:

- The total number of words in the Voynich manuscript.
- The words which look like digits from some paragraphs of the manuscript.
- The frequency of special words which appear at the start, middle or end of paragraph.

On the other hand, in the course of digits investigation, team members will search for different kinds of known expressions of digits and make a comparison with the words in the Voynich manuscript. For example, the expression of digits in Roman is as shown in the A.2 figure 2. The word which is as shown in the A.3 figure 3 is extracted from the Voynich manuscript, it is obvious that the form of the word in the A.3 figure 3 is like '*##'. According to the method of comparison mentioned above, this word maybe means seven in Roman.

4.2 Phase 2: Illustration investigation.

An illustration which is extracted from the Voynich manuscript is as shown in the A.4 figure 4.

In this phase, illustrations will be analysed by using Matlab. Generally, there are three aspects which are needed to be dealt with:

- The number of different elements in the illustration.
- The feature of words.
- Relations between different illustrations.

4.3 Phase 3: Marginal symbol research.

A page which contains marginal symbols is as shown in the A.5 figure 5.

This phase also requires proficiency in programming by using Matlab. During the process of this process, there are four major aspects:

- Ordering and quantitative features of the symbols at the margin of the page.
- Linguistic features of the words after marginal symbols.
- The differences between different marginal symbols in one page.
- Relations between two different pages containing marginal symbols.

4.4 Phase 4: Translation.

During the process of this phase, team will try to translate some parts of the Voynich manuscript. The major method is as shown in the A.6 figure 6. There are four important steps:

- Tool: Matlab.
- English linguistic research.
- Structure of words in the Voynich manuscript investigation.
- Some parts of the manuscript translation.

5. Project management.

5.1 Deliverables

As shown in table 1, deliverables involve eleven parts.

Deliverable	Deadline
Proposal seminar	4st of April, 2016
Project wiki (introduction)	Semester 1, week 5.
Thesis (1st draft)	22th of April, 2016.
Thesis (2nd draft)	Semester 1, week 12.
Master thesis (final)	Semester 2, week 11.
Expo Poster	Semester 2, week 11.
Project wiki (full)	Semester 2, week 12.
Expo presentation	Semester 2, week 12.
YouTube video	Semester 2, week 12.
USB flash drive (all codes and works)	Semester 2, week 12.
Final seminar	Semester 2, week 13.

Table 1.

5.2 Work breakdown

The details about tasks are as shown in the A.1 figure 1. The key tasks involve two aspects:

- Text investigation (digits).
- Translation.

5.3 Timeline

Timeline of project involves five parts. The specific details are as shown in the table 2.

NO.	Task	Week
	Semester 1	

1	Background research	1
2	Phase 1: Text analysis	6
3	Phase 2: Illustration investigation	8
4	Phase 3: Marginal symbol research	10
	Semester 2	
5	Phase 4: Translation	5-9

Table 2

5.4 Task allocation

Task allocation is divided into five parts:

No	Task	Student
	Semester 1	
1	Background research	Ruihang Feng, Yaxin Hu
2	Phase 1: Text analysis	Ruihang Feng, Yaxin Hu
3	Phase 2: Illustration investigation	Yaxin Hu
4	Phase 3: Marginal research	Ruihang Feng
	Semester 2	
5	Phase 4: Translation	Ruihang Feng, Yaxin Hu

Table 3.

5.5 Management strategy

Team members will be managed through a minimum of two internal meetings every week, and a minimum of one fortnightly meeting with supervisors. In addition, the preparation for each meeting involves three aspects:

- Achievements in the past two weeks.
- Questions about the work of the past two weeks.
- Plan for next two weeks.

After meeting, there are two tasks:

- Meeting content reorganization.

- Code modification.

5.6 Budget

Budget involves four aspects:

- 500 AUS dollars for team members.
- Research need to be carried out further research.
- All programs that need to be used are available on university system.
- All major works can be achieved by using computer.

5.7 Risk analysis

Details of risk analysis are as shown in the table 4.

No.	Risk	Probability	Impact
1	Mismanagement of time	Moderate	High
2	Loss of data or files	Low	High
3	Team member's quit	Low	High
4	Lack of references	High	High
5	Health issues	Moderate	Moderate

Table 4.

5.7.1 Mismanagement of time

Due to other works in daily life, the mismanagement of time may occur. Hence each member should arrange the time in advance to avoid time clash.

5.7.2 Loss of data or files

During the process of project, there may be some accidents, such as code lost or failure of files storage. In order to avoid that kind of situation, team members should buy two or more USB flash drive to store the backup files.

5.7.3 Team member's quit

In order to avoid this case, team members should keep frequent contact with each other.

5.7.4 Lack of references

As the mentioned before, the references of the Voynich manuscript are limited. So members should expand the scope of research, such as Bling, Grammarly and other websites.

5.7.5 Health issues

Members should pay attention to regular work and break to prevent health problems.

6. Conclusion

This project is divided into four phases: Text investigation, illustration research, marginal symbol investigation and translation.

In addition, the goals of this project involve three parts:

- Use statistical method Matlab to search the language rules in the Voynich manuscript.
- Search laws from illustration.
- Investigate laws from marginal symbols.

On the other hand, the major works of this project can be achieved by using computer.

7. References

[1] R. Zandbergen (2016). *The Voynich MS-Introduction* [Online]. Available: http://www.voynich.nu/intro.html

[2] Kevin Knight, Sravana Reddy, *What We Know About The Voynich Manuscript* [Online]. Available: http://www.isi.edu/natural-language/people/voynich-11.pdf

[3] Stojko, John, *Letters to God's Eye: The Voynich Manuscript for the first time deciphered and translated into English*. New York: Vantage Press, 1978.

[4] Joachim Dathe, *The EVA-Transcription* [Online].

Available: https://voynich2arabic.wordpress.com/eva-transcription/

[5] Reed Johnson (2013, July 9), *The Unread: The Mystery Of The Voynich Manuscript* [Online].

Available:

http://www.newyorker.com/books/page-turner/the-unread-the-mystery-of-the-voynich-manuscript

[6] R. Zandbergen (2016), *History of research of the Voynich MS* [Online]. Available: http://www.voynich.nu/solvers.html#n01

[7] Mary E. D'Imperio, *An Application of Cluster Analysis and Multiple Scaling to the Question of "Hands" and "Languages" in the Voynich Manuscript*. Washington, DC, 1992.

[8] Pelling, Nicholas, *The curse of the Voynich; the secret history of the world's most mysterious manuscript*, Compelling Press, Surbiton, 2006.

[9] Bennett, William Ralph, *Scientific and Engineering Problem Solving with the Computer*. Englewood Cliffs: Prentice-Hall, 1976.

[10] Tiltman, John, "The Voynich Manuscript, The Most Mysterious Manuscript in the World". NSA Technical Journal 12 (July 1967), pp.41-85.

[11] Feely, Joseph M, *Roger Bacon's Cipher: The Right Key Found*, Rochester, 1943.

[12] D'Imperio, Mary E, *The Voynich Manuscript - an elegant enigma*, Aegean Park Press, 1978.

[13] R. Zandbergen (2016), *History of research of the Voynich MS* [Online]. Available:

http://www.voynich.nu/solvers.html#n43

8. Appendix

8.1 A.1.Proposed Method

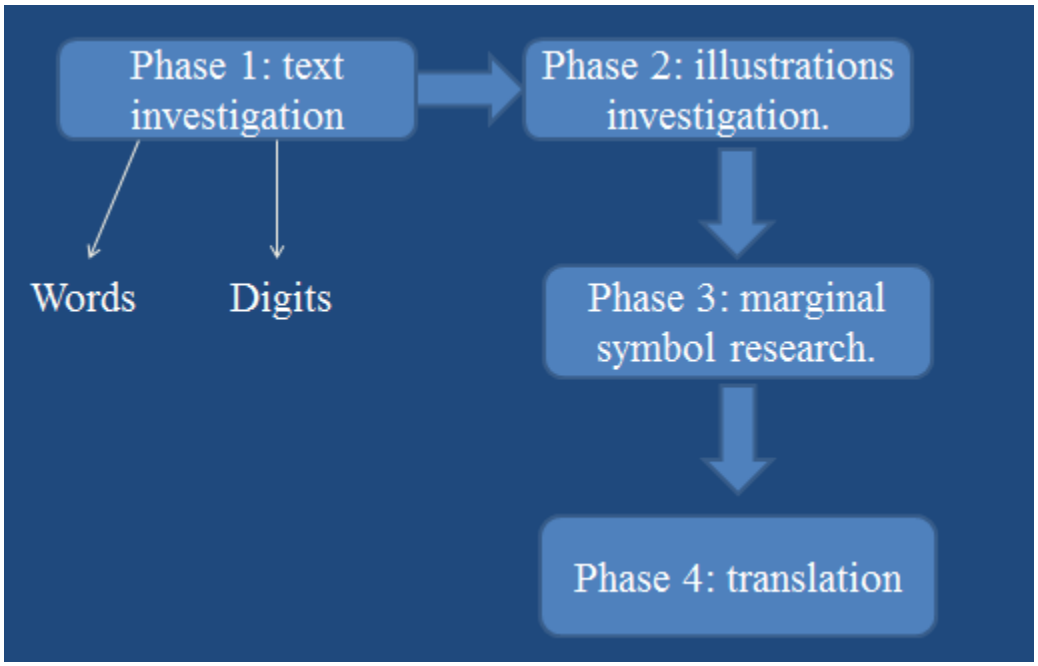

Figure 1.

8.2 A.2. Roman Numeral

Number	Roman Numeral
1	I
2	II
3	III
4	IV
5	V
6	VI
7	VII
8	VIII
9	IX
10	X

Figure 2.

8.3 A.3. Words from the Voynich manuscript.

Figure 3.

8.4 A.4. Illustration from the Voynich manuscript.

Figure 4.

8.5 A.5. Marginal symbols from the manuscript.

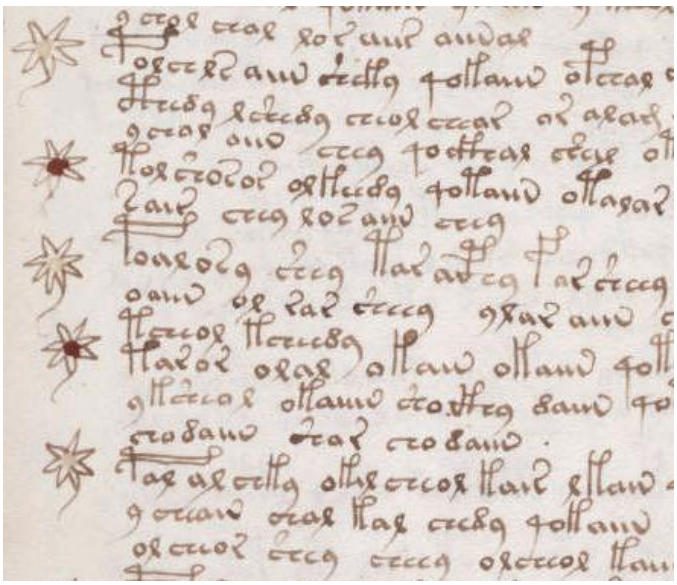

Figure 5.

8.6 A.6. Translation.

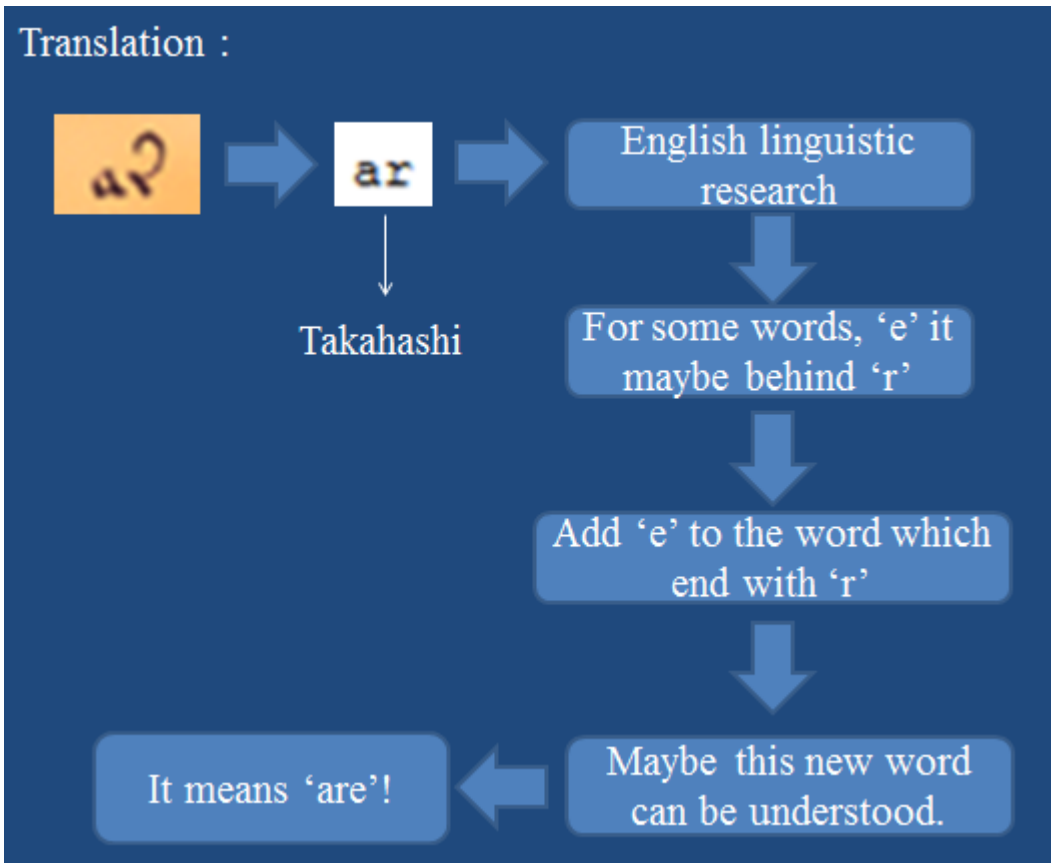

Figure 6.

8.7 A.7. FSG

Char	Bennett	FSG	Currier		Char	Bennett	FSG	Currier
	D	4	4			I	I	I
	O	O	O			IL	IE	G
	S	8	8			IIL	IIE	H
	G	G	9			IIIL	IIIE	1
	Z	2	2			IQ	IR	T
	L	E	E			IIQ	IIR	U
	Q	R	R			IIIQ	IIIR	0
	CT	T	S			U	L	D
	ET	S	Z			N	N(*)	N
	H	H	P			M	M(*)	M

Figure 7.