



Wang, X. (2016) *A creative computing approach to poetry as data*. PhD thesis, Bath Spa University.

ResearchSPAce

<http://researchspace.bathspa.ac.uk/>

Your access and use of this document is based on your acceptance of the ResearchSPAce Metadata and Data Policies, as well as applicable law:-

<https://researchspace.bathspa.ac.uk/policies.html>

Unless you accept the terms of these Policies in full, you do not have permission to download this document.

This cover sheet may not be removed from the document.

Please scroll down to view the document.

A CREATIVE COMPUTING APPROACH TO POETRY AS DATA

Xuan Wang

A thesis submitted in partial fulfilment of the
requirements of Bath Spa University
for the degree of Doctor of Philosophy

Centre for Creative Computing
School of Humanities and Cultural Industries
Bath Spa University

December 2016

To my grandma, Guizhen Gao
my parents, Chunyan Lu and Dawei Wang
my husband, Zhuangzhuang Dai
my uncles, Liming Lu and Jian Lu
for their love and support

Copyright

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. I have exercised reasonable care to ensure that my thesis is original, and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other intellectual property right.

Declaration

I declare that the work described in this thesis was originally carried out by me during the period of registration for the degree of Doctor of Philosophy at De Montfort University and Bath Spa University, UK, from October 2010 to December 2016. It is submitted for the degree of Doctor of Philosophy at Bath Spa University. Apart from the degree that this thesis is currently applying for, no other academic degree or award was applied for by me based on this work.

Publication Lists

[1] Hui Guan, Xuan Wang and Hongji Yang, “A Framework for Security Driven Software Evolution”, *the 20th International Conference on Automation and Computing (ICAC)*, Cranfield University, IEEE, United Kingdom, 2014, pp. 194-199.

[2] Xuan Wang and Hongji Yang, “Applying Semantic Web Techniques to Poem Analysis”, *the 21st International Conference on Automation and Computing (ICAC)*, University of Strathclyde, IEEE, United Kingdom, 2015, pp. 65-70.

[3] Xuan Wang and Hongji Yang, “User Query Optimisation: A Creative Computing Approach”, *Communications in Computer and Information Science*, Springer, Singapore, vol. 675, 2016, pp. 68-78.

[4] Xuan Wang, Lu Zhang and Hongji Yang, “Design of A Creative Encryption Approach Based On Pataphysics”, *the Third International Conference on Trustworthy Systems and Their Applications (TSA)*, Wuhan University, IEEE, China, 2016, pp. 123-126.

[5] Xiwen Wu, Chen Li, Xuan Wang and Hongji Yang, “A Creative Approach to Reducing Ambiguity in Scenario-based Software Architecture Analysis”, *International Journal of Automation and Computing*, Institute of Automation, Chinese Academy of Sciences, China, January, 2018, pp. 1-13.

Acknowledgements

Without the support and aid from many people, I would not have completed my PhD study. Here I would like to take the opportunity to thank all of them.

Firstly, I wish to express my most profound thanks to my first supervisor, Professor Hongji Yang, for his excellent guidance, invaluable advice, patience and encouragement throughout my PhD study. He always provided me with many useful comments and suggestions for the improvement of the thesis. I learned a great deal from him and I appreciate him for his leading role fostering my academic, professional and personal growth.

My heartfelt thanks go to my second supervisor, Professor Andrew Hugill, for his deep insights and invaluable suggestions. I am grateful for his support and encouragement on my PhD study.

My thanks must go to Dr. Chen Li, for providing me with many helpful suggestions. I really appreciate his academic guidance.

In addition, I would like to thank the Graduate School Office of Bath Spa University for their kind help and outstanding management.

Finally, I wish to express thanks to my dear family: my grandma, my parents, my uncles and my husband, Zhuangzhuang Dai, for their unconditional love, encouragement, patience and support over the past years. This thesis is dedicated to them.

Abstract

With the rapid advent of emerging new services such as cloud computing, mobile technology, and social media, more and more people prefer posting their literary creations such as poems, on the Internet instead of in traditional papers. The era of Digital Humanities has truly arrived.

With ever-growing concerns regarding literary data, ways to utilise and manage them has become a major concern. Many researchers have worked on that and proposed different solutions. However, owing to new challenges and creative requirements, traditional methods need adjustments. For example, most poetry data collection methods, such as surveys, are based on single target searching; that is, only relying upon keywords and themes. Thus, the result can be monotonous. Moreover, the accuracy of algorithms for poetry data analysis is no longer the only benchmark. The underlying meaning of poetry data has drawn people's attention. Meanwhile, traditional poetry data presentation methods need to be enhanced to reflect diversity and media richness.

The aim of this research is to present a Creative Computing approach to poetry data collection, analysis and presentation. The thesis demonstrates the feasibility and details the proposed methods in the following phases.

Firstly, poetry data is being creatively regarded as an object with mass, volume and resistance, from an interdisciplinary perspective. A novel data relevancy rule is proposed to retrieve the closely-related data of an input, which is adapted from the Newton's law of universal gravitation in physics. In this way, a broadened variety of data is being searched using web crawler based on multi-purpose rules. Then, the search results are filtered on the basis of buoyancy phenomenon and Ohm's law.

Secondly, with reference to chemical principles this research carries out innovative poetry data analysis based on the notion that chemical reactions always bring in brand new outcomes, despite having exactly the same elements. The mood, theme and

personal reflection, after going through a piece of literature, presented difficulties for traditional data analysis. In this work, they have been investigated relying on acidity estimation, organic abstraction and oxidation-reduction reactions.

Lastly, presenting the poetry analysis results through creative visualisation has been studied thanks to the elegant mathematics expressions of curves and shapes which are believed to effectively convey underlying emotions of poetry data. To illustrate this idea, a rainbow of variable spectrum and diverse types of trajectories are proposed as background and rolling titles, respectively.

In summary, the proposed approach carries out manipulations on traditional poetry data processing based on models and algorithms of Creative Computing. The proposed approach was evaluated by a selected case study, where a prototype system was built for poetry analysis. Conclusions are drawn and future research is also discussed. Initial experiment results show this work contributes to an effective and Creative Computing approach to poetry data manipulation. This research has potential applications to academic research of texts, to making word recommendations for users to better comprehend literature such as poetry, a novel or drama. Furthermore, it sees the possibility of inspiring creative thinking for human art creation.

List of Figures

Figure 5.1	Newton's Law of Universal Gravitation [101].....	52
Figure 5.2	Architecture of Focused Crawling	55
Figure 5.3	Poem Weighing Method.....	56
Figure 5.4	the Pseudo Code of the Focused Crawler Algorithm.....	57
Figure 5.5	Archimedes' Buoyant Force Principle	59
Figure 5.6	Resistors Connected in Parallel [43]	61
Figure 6.1	Glucose Molecules [94]	68
Figure 6.2	Organic Histogram for 'Rough Winds do Shake the Darling Buds of a Date'	70
Figure 7.1	Mood Presentation of 'Rough Winds do Shake the Darling Buds of a Date'	77
Figure 7.2	Mood Curve of the Poem with pH=8	78
Figure 7.3	Rainbow Plot of 'Rough Winds do Shake the Darling Buds of a Date'	81
Figure 7.4	'Victory' for Confident.....	82
Figure 7.5	'Rising' for Cheerful	82
Figure 7.6	'Sinusoidal' for Ironic	83
Figure 7.7	'Hyperbola' for Yearning	83
Figure 7.8	'Exponential' for Passionate.....	83
Figure 7.9	'Logarithmic' for Grateful.....	84
Figure 7.10	'Helix' for Sense of Lost	84
Figure 7.11	'Descending' for Miserable	85
Figure 7.12	'Downwards Triangle' for Hate.....	85
Figure 7.13	'Rhombus' for Regretful.....	86
Figure 7.14	'Square' for Encouraging	86
Figure 7.15	'Circle' for Delighted	86
Figure 8.1	User Interface of Poetrypedia.....	89
Figure 8.2	Second User Interface of Poetrypedia	90
Figure 8.3	Design Details of Exploration Module.....	93

List of Figures

Figure 8.4	User Interface of Exploration Module.....	93
Figure 8.5	Design Details of Appreciation Module.....	94
Figure 8.6	User Interface of Appreciation Module.....	95
Figure 8.7	Design Details of Help-to-Write Module.....	96
Figure 8.8	User Interface of Help-to-Write Module.....	97
Figure 8.9	Design Details of Dissemination Module.....	97
Figure 8.10	User Interface of Dissemination Module.....	98
Figure 8.11	User Input of Analysis by Aspects.....	99
Figure 8.12	User Output of Analysis by Aspects.....	100
Figure 8.13	User Input of Possible Creative Connections.....	102
Figure 8.14	Songs List of Possible Creative Connections.....	104
Figure 8.15	Top Seven Songs of Possible Creative Connections.....	104
Figure 8.16	User Input of Poetry Presentation.....	105
Figure 8.17	User Output of Poetry Presentation.....	106
Figure 8.18	User Input of Poem Generation.....	107
Figure 8.19	User Output of Poem Generation.....	107
Figure 8.20	User Input of Word Recommendation.....	108
Figure 8.21	User Output of Word Recommendation.....	109
Figure 8.22	User Input of Poem Fill-in-the-Blank.....	110
Figure 8.23	User Chosen of Poem Fill-in-the-Blank.....	110
Figure 8.24	User Output of Poem Fill-in-the-Blank.....	111

List of Tables

Table 5.1	Word Classes	52
Table 5.2	Word Classes Values	60
Table 6.1	<i>pH</i> Values of Gorgeous Cherry Blossom	66
Table 6.2	Six Themes for Poems and Their Organic Elements	68
Table 6.3	Elements Count of 'Rough Winds do Shake the Darling Buds of a Date'	69
Table 6.4	Themes of Poems and Their Oxidation and Reduction	71
Table 6.5	Themes and Their Analysis Results	72
Table 7.1	Colours and Related Meanings [34].....	79
Table 7.2	Six Colours for Themes	80
Table 7.3	Twelve Kinds of Personal Reflections and Their Curves	81

List of Acronyms

DPSS	Deep-prior Searching Strategy
BPSS	Broad-prior Searching Strategy
RPSS	Relevance-prior Searching Strategy
KNN	K-Nearest Neighbours
WGR	Word Gravitation Rule
FFR	Flotation Filter Rule
CCR	Conductivity Calculator Rule
NR	Neutralisation Rule
OSR	Organic Substance Rule
ROR	Reduction-Oxidation Rule
MPR	Mood Presentation Rule
TPR	Theme Presentation Rule
PRPR	Personal Reflection Presentation Rule

Contents

Copyright	I
Declaration.....	I
Publication Lists.....	II
Acknowledgements.....	III
Abstract.....	IV
List of Figures	VI
List of Tables.....	VIII
List of Acronyms.....	IX
Contents	X
Chapter 1 Introduction.....	1
1.1 Overview of Problems and Proposed Research.....	1
1.2 Research Objectives.....	4
1.3 Research Method	4
1.4 Research Questions and Hypotheses	6
1.5 Original Contributions	8
1.6 Success Criteria	8
1.7 Organisation of Thesis	9
Chapter 2 Poetry Data Management and Creative Computing.....	12
2.1 Overview.....	12
2.2 Poetry Data Management	12
2.2.1 The Era of Digital Humanities	12
2.2.2 Collection, Analysis and Presentation of Poetry Data	13
2.3 Creative Computing.....	19
2.3.1 Requirements of Creativity	19
2.3.2 Concepts and Forms of Creative Computing.....	20

2.4 Summary.....	24
Chapter 3 Related Work on Poetry Data Manipulation	25
3.1 Overview.....	25
3.2 Poetry Data Collection Methods.....	26
3.2.1 Common Data Collection Methods.....	26
3.2.2 Poetry Data Collection Web Sources	27
3.2.3 Web Data Collection — Web Crawler	28
3.3 Poetry Data Analysis Methods.....	31
3.3.1 Common Data Analysis Methods	32
3.3.2 Literature Analysis Methods	33
3.3.3 Poetry Analysis Methods	35
3.4 Poetry Data Presentation Methods.....	36
3.4.1 Common Data Presentation Methods	36
3.4.2 Poetry Data Presentation Methods.....	37
3.5 Relevant Laws of Physics, Principles in Chemistry and Mathematics Formulae.....	38
3.5.1 Laws of Physics	38
3.5.2 Principles in Chemistry	39
3.5.3 Mathematics Formulae.....	39
3.6 Summary.....	40
Chapter 4 A Creative Computing Approach to Poetry Data Collection, Analysis and Presentation.....	42
4.1 Overview.....	42
4.2 Framework of Creative Computing Approach to Poetry Data Manipulation.....	44
4.3 Creative Poetry Data Manipulation Phases	46
4.3.1 Multi-Purpose Poetry Data Collection Phase.....	46
4.3.2 Multi-Dimension Poetry Data Analysis Phase.....	47
4.3.3 Multi-Media Poetry Data Presentation Phase	48
4.4 Summary.....	48
Chapter 5 A Multi-Purpose Poetry Data Collection Method.....	50

5.1 Overview.....	50
5.2 Using Web Crawler for Multi-Purpose Poetry Data Collection	51
5.2.1 Poetry Data Relevancy	51
5.2.2 Word Gravitation Rule (WGR)	52
5.2.3 Web Crawler Implementation	54
5.3 Using Flotation Filter and Conductivity Calculator Rules for Poetry Data Pre-processing.....	57
5.3.1 Common Poetry Data Quality Issues	57
5.3.2 Flotation Filter Rule (FFR)	58
5.3.3 Conductivity Calculator Rule (CCR).....	60
5.4 Summary.....	62
Chapter 6 A Multi-Dimension Poetry Data Analysis Method	64
6.1 Overview.....	64
6.2 Data Sentiment Analysis	65
6.3 Neutralisation Rule for Mood Analysis	66
6.4 Organic Substance Rule for Theme Analysis	68
6.5 Reduction-Oxidation Rule for Personal Reflection Analysis	70
6.6 Summary.....	73
Chapter 7 A Multi-Media Poetry Data Presentation Method	75
7.1 Overview.....	75
7.2 Mood Presentation Rule	76
7.2.1 Mood and Emoticons	76
7.2.2 Rule Definition.....	76
7.2.3 An Example for MPR.....	77
7.3 Theme Presentation Rule	78
7.3.1 Colour and Meaning of Data.....	78
7.3.2 Theme Presentation Rule	80
7.3.3 An Example for TPR.....	80
7.4 Personal Reflection Presentation Rule.....	81

7.4.1 Rule Definition.....	81
7.4.2 Twelve Kinds of Curves for Personal Reflection.....	82
7.4.3 An Example for PRPR	86
7.5 Summary.....	87
Chapter 8 Case Studies	88
8.1 Overview.....	88
8.2 Design Details and User Interfaces of Poetrypedia	89
8.3 Case Studies Summary	98
8.3.1 Function of Analysis by Aspects in Poetrypedia.....	98
8.3.2 Function of Possible Creative Connections in Poetrypedia.....	101
8.3.3 Function of Poetry Presentation in Poetrypedia.....	105
8.3.4 Function of Poem Generation in Poetrypedia	106
8.3.5 Function of Word Recommendation in Poetrypedia	108
8.3.6 Function of Poem Fill-in-the-Blank in Poetrypedia.....	109
8.4 Summary.....	111
Chapter 9 Conclusions and Future Work.....	113
9.1 Summary.....	113
9.2 Evaluation	114
9.3 Contributions	115
9.4 Future Work	116
References.....	118
Appendix A. List of the Most Commonly Used English Words With pH and Theme Value (Top 2000 Words in Descending Order)	127
Appendix B Selected C# Code of Possible Creative Connection Function in Poetrypedia	198
Appendix C Matlab Code of Rainbow Spectrum	206
Appendix D Matlab Code of Curves and Shapes	207

Chapter 1

Introduction

Objectives

- To observe the problems of poetry data management
 - To introduce Creative Computing and proposed poetry data manipulation approaches
 - To explain the research objectives and research methods
 - To discuss research questions and develop research hypotheses
 - To highlight original contributions and define the measure of success
 - To outline the organisation of the thesis
-

1.1 Overview of Problems and Proposed Research

With the rapid advent of emerging new information services such as blogs, Facebook and Weibo, more and more people prefer posting their literary creations, such as poems or novels, on the Internet instead of in traditional paper form. The digital era of literature is with us. In this era, poetry or novels have been transferred from a form of traditional literature to complex data resources, and how to manage and utilise this kind of data has drawn much attention, lately. As is known, there are four primary literary forms including drama, prose, poetry and the novel. Poetry is usually the shortest and the most vigorous. Therefore, the proposed literary data manipulation approach was put forward to regard poetry as a representative, and to be generalised further.

Poetry is a creative expression of language which is full of imagination and beauty. The ambiguity of poetry increases the difficulty in interpretation and appreciation. It becomes a hotspot when using computer science to manipulate a large amount of

poetry data. Forms of poetry data manipulation instances include statistical methods, e.g., style and meter [44], or generating poetry of greater readability using genetic algorithms [59], etc. Although these approaches provide a foundation for poetry data - analysis, they are restricted in their range of applications, lack systematic structure, and, mostly importantly, lack creativity in terms of collection, and analysis, as well as presentation.

Creativity, owing to its mysterious power, has also been studied for a long time and in various domains. As an interdisciplinary research area, Creative Computing has gradually expanded into various related research directions [92, 48, 109, and 85], such as: creative design, creative requirement engineering, and creativity analysis and evaluation. According to research at that time, Creative Computing could be seen as the combined product of computing and other technologies. It can be utilised to generate creative products: for instance, creative software, which combines computing and other technologies to provide creative services [110, 85, and 71]. Creativity is not only the target for technology, but is also the foundation of human creation [13].

Creative Computing has been widely used in many fields. In Boden's opinion, creativity is the ability to create something new, surprising, and valuable [12]. As for the meaning of "new", there are two different degrees: new to a specific person (P-Creativity) and new to the whole world (H-Creativity) [12]. Similar to P-Creativity and H-Creativity corresponding to the significance of "new", there are also three types of creativity with respect to the meaning of "surprising", which can be employed in data manipulation to realise its creativity features: combinational creativity, transformational creativity and exploratory creativity [2, 95, and 107]. Based on different forms of Creative Computing, using it to process data allows for the capability of increasing the diversity and meaningfulness of analysis, of extending the scope and depth of analysis, and of producing creative results to inspire creative minds. Moreover, the online data storage is currently too simple. Methods of accessing and analysing these data are limited and lack full comprehension. Data presentation also needs to be made more interesting and

inspirational. Hence, data itself may rely on Creative Computing to trigger a better understanding of its contents.

Based on the method of Creative Computing, this research discusses the process and possibility of presenting a creative data manipulation approach in poetry. In order to examine this approach, this research innovatively borrowed laws of physics, chemistry and geometry to collect, analyse and present poetry data, respectively. Poetry data can be abstracted as objects, compounds or numbers. Theorems in Physics can be adopted to provide quantitative analysis on data collection. Here a poem of interest is regarded as an object in which the significance of each word is measured by mass, and the emotion of the poem is calculated using acid-based properties. The poem is subsequently presented depending on the results of the calculations, with the aid of chromatics and geometry. Principles in Chemistry can be used to provide qualitative data analysis. Shapes and curves in Mathematics are adopted to present the analysis results accordingly. Through this method, the diversity, richness and creativity of poetry data manipulation is enhanced.

In order to validate this approach, a Poetrypedia system that tests all of the above methods upon a carefully chosen collection of poetry data is presented. Poetrypedia is a comprehensive multi-functional poetry system, which not only helps users explore all the relevant information about poems, but also assists learning, comprehension and composition. Based on the proposed multi-disciplinary analysing approach, a series of functions to aid users to appreciate poems are developed. For example, the Possible Creative Connections module can provide a list of potentially relevant songs after being given a poem. The Poetry Presentation module presents a poem based on a specific geometrical shape or curve. Poem Fill-in-the-Blank is a mini-game aiming to help users memorise a poem. Poetrypedia not only explores the diversity and depth of poetry data collection, analysis and presentation, but also broadens the horizon for users to think and create.

1.2 Research Objectives

Currently, research concentrating on the creativity of data processing on literature is lacking in existing data studies. Particularly, there is little work focussing on designing or developing a method, a framework or an approach for poetry data manipulation to provide means of creative poetry data processing.

To address the above issue, this thesis narrows down the study to present a creative poetry data manipulation approach. In particular, it proposes a Creative Computing approach with designed algorithms for data collection, data analysis and data presentation of poetry. It relies greatly on creativity features in data processing to support and inspire users. Therefore, it requires a new computing method that is different from traditional ones to provide essential creativity and to promote poetry data processing. In the proposed data manipulation approach, there are three phases, Multi-Purpose Data Collection, Multi-Dimension Data Analysis and Multi-Media Data Presentation. Algorithms and rules are picked and designed to support each phase. The reasoning rules are designed based on Creative Computing, the laws of Physics, principles of Chemistry and Mathematics formulae.

1.3 Research Method

This section describes the research method applied in this work. Generally, research is a process of studying knowledge or the understanding of phenomena, and there are various ways of classifying research methods [66, 24, 51]. Four dominant philosophical methods are categorised in Creswell's book [23], which are positivism, constructivism, critical theory, and pragmatism. The research field in this work belongs to software engineering and the aim of this study is to provide a novel approach for poetry data manipulation. Therefore, the research method selected and applied is positivism.

Positivism is known as a philosophical theory that was founded by the French philosopher Auguste Comte. It is widely accepted that positivism is of great value to scientific research; even Stephen Hawking is one of the advocators of positivism.

“Any sound scientific theory, whether of time or of any other concept, should in my opinion be based on the most workable philosophy of science: the positivist approach put forward by Karl Popper and others [41].”

— *Stephen Hawking*

The Universe in a Nutshell

Positivism believes that reality is stable and observable from an objective point of view [24, 54]. It is applied to increase confidence in the proposed approach through the empirical observation of the implementation of proposed poetry data manipulation methods.

There are usually four steps comprising the research method of Positivism:

1. The first step is observation. State-of-the-art poetry data management methods are inspected in Chapter 2 and 3. The limitations and defects of these methods are addressed and concluded.
2. The second step is establishing hypothesis. In Chapter 4, a creative computing approach is proposed to collect, analyse and present poetry data. A series of creative rules are developed in Chapter 5, 6 and 7 based on principles of Physics, Chemistry and Mathematics.
3. The third step is experimentation. A Poetrypedia system is designed based on the proposed approach which is demonstrated in Chapter 8.
4. The fourth step is validation. Survey and feedback are to be gathered and analysed after the Poetrypedia is published, which would provide validation for the proposed approach.

How to determine if the approach is creative? This question can be answered from three aspects. Firstly, results of the analysis are compared, using the approach to the literature, database and online resources to identify differences and improvements; secondly, surveys to find out responses based on the analysis are carried out; thirdly, the Poetrypedia system is going to publish on a website to collect extensive feedback in terms of inspiration.

This research aims to inspire creative retrieval via a multi-disciplinary investigation method. Creativity from collision is going to be generated, e.g., by cross-correlating, principles in physics, chemistry, geometry and statistics. Concepts and visual presentations are merged to demonstrate poetry data and information in an innovative way. Poems are deconstructed based on theme, emotion, personal reflection etc., and are reconstructed to bring about new understandings. All the above proposals are implemented and exhibited on a website developed from scratch. This website is believed to be applicable for educational, academic, commercial and public use. And it is not only a good synthesis of methods adopted, but also a milestone of the achievement.

1.4 Research Questions and Hypotheses

Research questions are the core part of the structure of the proposed research. The principle research question in this study has been formulated below.

How can a creative data manipulation approach for poetry be presented that is valuable and inspiring for users?

The research work described in this thesis aims to address this research question effectively. In order to achieve this aim, a set of manageable and tractable sub-questions are defined, addressing the problem in detail.

RQ1: What should a creative poetry data manipulation approach be like?

- What is the structure of a general poetry data manipulation approach?
- How can “being creative” be defined?
- What makes the proposed poetry data manipulation approach work in providing creative ways?

RQ2: How can relevant poetry data be obtained and managed?

- Which techniques can be used to support poetry data collection?
- What are the specific rules to be followed to get proper and useful poetry data?

RQ3: How can the obtained poetry data be analysed and utilised?

- Which techniques can be used to support poetry data analysis?
- What rules can be followed to creative poetry data analysis?

RQ4: How can the poetry data analysis results are presented?

- What rules are necessary in the process of presenting the poetry data analysis results?

RQ5: What kinds of tools is capable supporting the development of the proposed poetry data manipulation?

In order to explore these research questions, a series of research hypotheses are developed. The fundamental hypothesis of this thesis is presented in below.

Creative Computing can lead a way of establishing an approach to poetry data manipulation in order to provide creative data collection, data analysis and data presentation for poetry to inspire users.

This phase is tested by developing a Creative Computing approach to poetry data manipulation based on Creative Computing methods and other techniques. A set of propositions is derived from the kernel one:

RH1: Creative Computing can be used as the main method to support creative poetry data collection. It is able to test this hypothesis by adopting Creative Computing related methods in data collection and data pre-processing.

RH2: Creative Computing and relevant techniques can support creative poetry data analysis, especially on the meaning of poetry data. Designing relevant rules and algorithms (e.g., reasoning rules) for data analysis by using Creative Computing and relevant techniques is a way to test this hypothesis.

RH3: Creative poetry data presentation can be provided by Creative Computing and relevant techniques. It can be tested by designing rules by applying Creative Computing and relevant techniques.

Furthermore, appropriate concepts, mechanisms, techniques, and rules are proposed to resolve each of the sub-questions. To achieve that, this research will take advantage of advances in the field of Creative Computing and creativity techniques.

1.5 Original Contributions

A creative poetry data manipulation approach is proposed and validated in the context of Creative Computing. In this research, reasoning rules, which are supposed to support data collection, data analysis and data presentation play a crucial role in aiding the creative approach in poetry. The following are explanations of all the expected original contributions:

- C1: A creative data manipulation approach is proposed and validated based on Creative Computing and relevant techniques, which contains three kernel phases: Multi-Purpose Data Collection, Multi-Dimension Data Analysis and Multi-Media Data Presentation.
- C2: A group of collection rules is designed to support data collection mainly based on the laws of Physics.
- C3: Reasoning rules are proposed and developed based on principles of Chemistry for creative data analysis.
- C4: Reasoning rules are presented to propose creative data presentation according to mathematics formulae.
- C5: A designed Poetrypedia system is implemented to illustrate the proposed approach. A series of functions are presented in details.

Overall, this research provides a Creative Computing method for poetry data manipulation which is able to provide creative data processing to assist and inspire human creation.

1.6 Success Criteria

This thesis aims to provide a creative computing approach to data collection, analysis and presentation for poetry. The overall measure of success must show a possibility of the proposed approach working to provide creative data manipulation via case studies. The following criteria are followed to evaluate the success of the research study that the thesis is expected to offer:

- SC1: an innovative and applicable approach to providing creative poetry data manipulation.
- SC2: a set of creative reasoning rules with the theoretical foundation for data collection, data analysis and data presentation for poetry.
- SC3: a number of case studies carried out to demonstrate the feasibility and generality of the proposed creative approach to poetry data manipulation.

1.7 Organisation of Thesis

The structure and contents of this thesis are organised as follow.

Chapter 1, Introduction, presents the motivation of this thesis, lists the research objectives, research methods, questions, hypotheses, and success criteria, declares its original contributions, and outlines the layout of this thesis.

Chapter 2, Poetry Data Management and Creative Computing, goes through the research background of digital humanities, and discusses the related studies of processes and previous works on poetry data management. This chapter investigates the current state of the collection, analysis and presentation of literary data, especially poetry data, and addresses challenges. In order to overcome these challenges, this chapter looks into the reason why Creative Computing is used as well as its strength and weaknesses, which provides a theoretical foundation for a creative approach to poetry data manipulation.

Chapter 3, Related Work on Poetry Data Manipulation, focuses on the state-of-the-art techniques of how poetry data are being collected, analysed and presented. Based on these studies, this work absorbs and refines their advantages. Particularly in the aspect of creativity, creative poetry data management is proposed by combining knowledge in science and engineering, such as the laws of physics, principles in chemistry and mathematics formulae.

Chapter 4, A Creative Computing Approach to Poetry Data Collection, Analysis and Presentation, proposes an approach to creative poetry data manipulation, which is specified individually in multi-purpose data collection,

multi-dimension data analysis and multi-media data presentation. An overall process is represented as well as an explanation to the approach in detail so as to trigger system development.

Chapter 5, A Multi-Purpose Poetry Data Collection Method, presents a creative multi-purpose poetry data collection method. Poetry data is creatively regarded as an object with mass, volume and resistance from an interdisciplinary perspective. A novel data relevancy rule is proposed to retrieve the closely related data of an input, which is adapted from the Newton's law of universal gravitation in physics. In this way, a broadened variety of data is searched using a web crawler based on multi-purpose rules. Then the search results are filtered on the basis of buoyancy phenomenon and Ohm's law.

Chapter 6, A Multi-Dimension Poetry Data Analysis Method, carries out innovative poetry data analysis based on the notion that chemical reactions always bring in brand new outcomes, while being based on exactly the same elements. The mood, theme and personal reflection after going through a poem used to be challenging for traditional poetry data analysis. In this work, they have been investigated relying on acidity estimation, organic abstraction and oxidation-reduction reactions.

Chapter 7, A Multi-Media Poetry Data Presentation Method, illustrates the presentation of poetry analysis results through creative visualisation has been studied thanks to the elegant mathematics expressions of curves and shapes which are believed to effectively convey the underlying emotions of poetry data. To illustrate this idea, a rainbow of variable spectrum and diverse types of trajectories are proposed as the background and rolling titles, respectively.

Chapter 8, Case Studies, shows the developed Poetrypedia system and its six functions. The six functions are developed following the proposed creative poetry data manipulation approach and work in six aspects, i.e., Analysis by Aspects, Possible Creative Connections, Poem Presentation, Poem Generation, Word Recommendation and Poem Fill-in-the-Blank. It explains how the proposed creative poetry data manipulation approach takes effect on various aspects thanks to their

diversities and their differences and validates the feasibility and generality of the proposed approach. Additionally, tools and technologies used for the prototypes are listed with a brief demonstration.

Chapter 9, Conclusions, summarises the thesis, draws conclusions and discusses the future work. Moreover, this section revisits and answers the research questions in order to evaluate the overall success of the proposed approach and other contributions.

Chapter 2 Poetry Data Management and Creative Computing

Objectives

- To introduce the background of digital humanities and literature on data management
 - To present the current state of poetry data management and challenging issues.
 - To review and propose the requirements of creativity and basic concepts of creative computing.
-

2.1 Overview

As the Internet is flourishing, the study of the humanities is also entering an age of digitisation. Using the technologies of computer science to solve problems of the humanities draws increasing attention. This chapter briefly explains the background of digital humanities, and then details the research accomplishments of literature data management. This chapter investigates state of collection, analysis and presentation of literary data, especially poetry data, and addresses challenges. In order to overcome these challenges, this chapter looks into the reason why Creative Computing is used as well as its strength and weaknesses, which provides a theoretical foundation for a creative approach to poetry data manipulation.

2.2 Poetry Data Management

2.2.1 The Era of Digital Humanities

On 17/11/2010, the New York Times published a report, titled ‘Digital Keys for Unlocking the Humanities’ Riches’ [21]. The author put forward a question: ‘what

would the next big idea be?’ His answer was: ‘data’. Since we are living in an era when researchers have access to huge amounts of data as never before, and information technologies offer incredible data processing power, when both join forces, they bring about the so-called Digital Humanities.

A wealth of data is the foundation of digital humanities research. We are living in an age of information explosion in which the digital humanities significantly differs from the past. To obtain large quantities of data is not just the privilege of sponsored researchers. Anyone who has access to the Internet is able to utilise this data, which includes written information, videos, audios, and all other forms of media. This is a remarkable improvement for research in the humanities.

A large amount of data, though, can be problematic. According to Roy Rosenzweig’s statement in 2003 [78], researchers are currently facing the problem of an abundance instead of a lack of data. How can abundant data be dealt with? How much data are significant? How can meaningfulness be obtained from a large amount of data? Is elaborated literal data of the past still useful? If not, what should be done to upgrade it? All these issues are yet to be solved in the field of digital humanities. Researchers are assumed to have different requirements of data retrieval compared to ordinary users. The former is more interested in the underlying messages, and the connections to other content, as well as the overall value of data. For literature data in particular, it is a major concern of literature data management to let users comprehend and interact with data in a creative way, and in depth.

Above all, an incredible data deluge is currently drowning the world. However, paying attention to literature data does not mean understanding it. If people want to make full use of literature data, it is necessary to first comprehend its current management methods and techniques. Considering the representativeness of poems in terms of literary composition, this work focuses on studying poetry data.

2.2.2 Collection, Analysis and Presentation of Poetry Data

Poetry data has a vast range of sources, of which the application requirements and data types are varied, whereas, the basic processing procedures remain consistent. These procedures with the aid of appropriate tools can be defined as extracting and

synthesising broadly isomeric data sources and storing the results according to uniform standards. The process and methods that leverage suitable data analysis techniques analyse the stored data and obtain beneficial information, then felicitously present the results to users are specifically classified as: poetry data collection and pre-processing, poetry data analysis and poetry data presentation.

2.2.2.1 Collecting Poetry Data

One of the most important characteristics of poetry data is diversity, which means the source of data is broad and the types of data are complex. This complex environment gives rise to difficulties in processing the poetry data. Before processing, required data sources must be sampled and collected. Then, the relationships and entities should be extracted. Eventually, the data can be stored after correlation and combination, using a uniformly defined structure. During collection and extraction, data needs to be cleaned to guarantee its quality and reliability. As has been discussed regarding the relationship between the patterns and the data, data generates precedes the patterns, which are rather dynamic.

Data extraction and synthesis is not a brand new technology; it is based on traditional database processes, which have already been investigated. Data collection methods evolve along with the presence of new data sources. According to data synthesis modelling, current data extraction and synthesis methods can be categorised into four types, Materialisation engine, Federation engine or Mediator, Stream engine and Search engine [42].

2.2.2.2 Analysing Poetry Data

In this section, it is going to introduce several research examples of literature data which attempt to reveal a new understanding of how data analysis uses different approaches. These methods may also be applied to the poetry data used in the research.

First of all, the frequency of use presents the number of counts of a word in a text. Taking Shakespeare for example: are all compositions named after Shakespeare his authentic work? It is a controversial question. There are many ways of checking this: e.g., checking titles against his bibliography, finding evidence of someone writing

on his behalf, or analysing the consistency of his style of writing. In this case, it is advised to find out the frequencies of use of representative words.

Is there idiomatic vocabulary among different compositions and chapters? Although they may not directly lead to an answer, they provide evidence for discussion. Before interpreting statistics, regularisation (e.g., stemming) must be carried out to correct the way counting has been carried out. The frequency of use can not only be used to determine style but give rise to further analysis. Since the way of writing or giving a speech exhibits one's pattern of thinking and doing, as well as style, analysing the writing style could also reveal the pattern of thinking of an author.

Researchers have used this method to analyse the inauguration of the presidents of the United States, and presented the results in tag cloud [53]. However, the content of the inaugural speech is too limited for modern computers to reveal the truth. If more reliable information is to be found, the inaugural speech, together with all other scripts of speeches made, should be utilised. For example, the publications of all the newspaper offices in the United States over the last 50 years can be merged for frequency of use analysis to reveal interesting trends. This method can be applied to both current texts and historical data. Can the records of the enthronement of the emperors of the Qing dynasty be analysed? Note that it cannot attain direct result from the frequency of use. It, at most, presents a phenomenon and underlying trends. What this research might infer requires a comprehensive analysis by researchers in the humanities. The other approach of the Digital Humanities is to discover the connections among data relying on data mining and text mining techniques. This approach has a wide variety of applications; e.g., analysing the behaviour pattern of a customer for better service and references [108]. In online book stores, previous shopping records are used to predict purchasing patterns and preferences. There are many connection analysis methods other than data mining. In Data Engineering, data mining includes categorisation, clustering, estimation and prediction, link analysis etc., with various theoretical models. It may not be wise to refer to the methods presented in the textbook in pursuit of answers for topics related to the humanities, because such analysis could be rigid and unreliable.

There is a considerable amount of research on poetry data. The quantitative analysis of poetry dates back to the 1940-s when poet and literary critic Josephine Miles began her extensive work analysing the surface statistics of poetry across time [63, 64]. While Mile's work was influential in establishing a statistical framework for thinking about poetry, it was done largely by hand, and thus is limited in scope and size.

Most recently, Hayward's connectionist model of poetic meter incorporated more sophisticated and varied features in the analysis [42]. For every feature considered, including prosody, meter, and syntax, Hayward hand-assigned numeric scores to each syllable in ten samples of poetry. Analysing these scores allowed him to identify unique patterns for each poet, and to note similarities within each period. However, this analysis required Hayward's personal assessment of the poems, as well as the assignment of feature scores. Since it is unfeasible to apply this method to a large set of poems, Hayward's model also faces limitations of size. One of the most thorough and sophisticated computing analysis of poems to date is the PoetryAnalyser. It is Kaplan and Blei's work on visualised comparison of style in American poetry [8]. Modern statistical and computational tools allowed the authors to integrate more features to analyse a large set of poems in an automated manner. The authors mapped poems from different poets and eras into a vector space based on three types of stylistic elements --orthographic, syntactic, and phonemic -- in order to find stylistic similarities among poems.

Other studies on the analysis of poetry focused on quantifying poetic devices such as rhyme and meter [35, 37], or classifying poems based on the poet and style [29, 44]. These studies show that computational methods can reveal interesting statistical properties in poetic language that allow us to better understand and categorise great works of literature. However, there has been very little work on the assessment of poetry.

Data analysis is at the core of poetry data processing as the value of poetry data is generated during the analysis. Processed data from various data sources are stored for data analysis. Depending on different demands data analysis can be carried out

among all or part of the stored data. Traditional data analysis technologies, such as data mining, machine learning, statistical analysis, etc., need adjustment because they encounter new challenges in digital humanities listed as follow:

- The increase in the quantity of data does not necessarily suggest an increase in the value of data, but leads to something even worse: data noise. Since cleaning is compulsory before data analysis, pre-processing such huge amounts of data is highly demanding for hardware, as well as algorithms. And the way in which valuable data is selected for different objectives is important.
- Algorithms need adjustment in the digital humanities era. First of all, timeliness is a key feature for the application of poetry data. The accuracy of algorithms is no longer the only benchmark for poetry data application. In many cases, a trade-off is inevitable between the timeliness and the accuracy: e.g., the algorithm of online machine learning. Secondly, for better adaption to the framework of data processing platforms, adjustments are required for algorithms, as well as them being more extendible. Last but not least, it needs careful selection of algorithm to perform data processing. Over a certain amount of data, an algorithm that effectively digs information from a small portion of data could not be applied to poetry data. In statistics Bonferroni's principle is a typical instance [99].
- It is not difficult to obtain analysis results but evaluation of these results becomes a new challenge in the digital humanities era. Due to the huge amount and diverse formats of data, it can be challenging to understand the distribution of data, which subsequently brings raises obstacles to the design of evaluation measures and indicators.

2.2.2.3 Presenting Poetry Data

Data analysis is at the core of poetry data management, whereas, data presentation is usually what users care about. If the analysis results are not appropriately interpreted, they may be incomprehensible, or even misleading. There are multiple ways of presenting data among which, traditionally, texts are output or displayed directly on the computer terminal.

The era of digitisation brings new possibilities to poems, multi-media is one of them. The Internet offers more sensorial content including both visual and audio. A previous president of the American History Association, Robert Darnton, prefers audio material, in particular. In his latest publication, *Poetry and the Police: Communication Networks in Eighteenth-Century Paris* [25], which collects underground poems criticising the government of Paris in the 18th century, he reproduced related underground songs of that time to make readers feel personally on the scene. As telecommunication technologies develop, such presentations, using visual and audio aids together with text, may become more popular. Not only poems, but all other forms of literature are facing revolutionary methods of presentation: visual, audio and other materials are required to arouse sensorial resonances in the audience.

These options may be good when dealing with a small amount of data. However, the amount of analysis results in the digital humanities era can be huge. Moreover, the correlations among the results are complicated. Traditional presentation methods no longer work effectively. This has led to the adoption of the following two approaches to enhance data presentation:

- Introducing visualisation. Visualisation is one of the most effective means to interpret large amounts of data: it was first adopted by scientific and engineering computing. Visualisation aids to help deliver the results of the analysis in a more comprehensible and acceptable manner than when plain text is used. Common visualisation techniques include: tag cloud, history flow, spatial information flow, etc., the most suitable one being chosen accordingly.
- Giving the user the opportunity to understand and participate in the analysing process. Human-computer interaction is one way to guide user analysis step by step, and this helps the user better comprehend the analysis results. The other way is via the provenance technique, which helps trace the whole process of data analysis [9].

So far, the background and current state of poetry data management and its challenges have been presented. In this thesis, the research aims to utilise Creative

Computing methods to improve the collection, analysis and presentation of poetry data.

2.3 Creative Computing

2.3.1 Requirements of Creativity

Creativity, owing to its mysterious power, has been studied for a long time and in various domains. There have been a variety of conceptions about creativity. It could take the form of ability, personality, process, and even just a state of mind. In terms of ability, people always focus on the training or coaching of different kinds of creative skills. Creative people are often studied to extract personality traits about creativity, such as ambiguity or tolerance. Studying creativity through the creative process, researchers have identified some important stages, such as incubation. Due to the emergence of disciplines like neuroscience in recent years, the mysterious Aha! Moment can also be studied scientifically. Various areas of the brain, and activities associated with this kind of creative state of mind, have been discovered. Furthermore, besides individual creativity, researchers have also carried out studies on how to improve organisational creativity. Issues, such as how to make the best use of employees' creativity, how to facilitate and improve employees' creativity, or how to organise a creative environment for employees to be creative, are key concerns in organisations all over the world [60].

Human creativity urgently needs improvement in modern competition. Globally, the market has been static in oversupply status for a long time. The quantity and even the quality, of certain products is no longer attractive nowadays. People increasingly are more and more concerned with the satisfaction of so-called "spiritual requirements". Objects, activities, and services are becoming more and more interesting, enjoyable, and playful. In the modern world, a new era has arrived—creativity competition [14]. Furthermore, because of the advanced technologies nowadays, infrastructures and facilities do not possess competitive advantages any more. As features like uniqueness

and novelty have become the secret ingredients of great success, the flash of creative ideas or human creativity as a whole, has been the hot pursuit in modern times.

As a result, in order to seize the great opportunities available, companies, organisations, and people all over the world have started to provide various kinds of services for creativity. Companies like OmniSkills, offers training, facilitating, and consulting in creativity. Organisations like Mycoted and Mind Tools are acting as “dictionaries” or “toolkits” for creativity. They are producing various kinds of tools, techniques, methods, books and other information on creativity. As for people teaching creativity, one of the most famous is Edward de Bono. His popular ideas of lateral thinking have gained great success across the world, and helped many people to come up with creative ideas. Even a cliché “think outside the box” has been created to encourage people to change the conventional thinking styles. Besides lateral thinking, there are also many other creative thinking methods, like divergent thinking. Tools like brainstorming, mind mapping and free writing are often used to aid people to perform divergent thinking.

2.3.2 Concepts and Forms of Creative Computing

2.3.2.1 Definition of Creative Computing

When it comes to Creative Computing, there has not been a commonly accepted definition. Starting in the late 1990-s, computing researchers have been paying attention to creativity. As an interdisciplinary research area, Creative Computing has gradually come into various related research directions [92, 48, 109, 85], such as: creative design, creative requirement engineering, creativity analysis and evaluation. According to the research of that time, Creative Computing could be seen as the combinational product of computing and other technologies, which can be utilised to generate creative products: for instance, creative software, which combines computing and other technologies to provide creative services [110, 85, 71]. Creativity is not only the target for technology, but is also the foundation of human creation [13].

In recent years, researchers have focussed on producing new, innovative and valuable products with the help of Creative Computing [103, 104], which seeks to

reconcile the objective precision of computer science with the subjective ambiguity of human creation [103]. For instance, a new web engine, called the Syzygy Surfer, has been developed by Hendler and Hugill [45]. It aims to operate a more creative search service than traditional ones.

There are generally three levels of meanings of Creative Computing:

- Functionally, Creative Computing aims to facilitate human creativity.
- Literally, the meaning of “Creative Computing” mainly relies on the significance of the adjective “creative”. However, there are various meanings ascribed to what “creative” might be. In synthesis, as related to different things (e.g. product, people, process, etc.) the word “creative” can also have different meanings. For example, when referring to a product, the word “creative” can stand for “new” and “valuable”. A “creative” person can be marked as one who is capable of creating things. A “creative process” (e.g., music composition) is often deemed as containing some particular activities, like the incubation of ideas [82, 47].

In the realm of Creative Computing, the adjective “creative” means “new” and “valuable”. It has something to do with the significance of computing. The term “computing” can also represent many different things. For example, as a disciplinary concept, computing includes disciplines like computer engineering, software engineering, computer science and information systems. As a relatively general concept, computing may refer to hardware, software and the web. However, as already mentioned, there is not a clear boundary between hardware, software, and the web. For example, the iPhone is often considered a hardware device. However, it also contains excellent software (e.g. iOS) and depends heavily on the web. Therefore, in order to cover all aspects of computing, the term “computing product” has been introduced in Creative Computing, and it is described as being new and valuable.

- Essentially, in order to be creative, computing itself, when studied in the realm of Creative Computing, contains creativity. According to previous analysis, from the common use of computing and the new use of computing to the new

type of computing to facilitate human creativity, the final inspiration for Creative Computing comes from creative computing products such as Apple (e.g., the iPhone). The development of the iPhone has been fused with various kinds of creativity. Although all of the creativities can be traced back to human creativity, they are more commonly represented or reflected by different kinds of creativity, such as: material creativity, colour creativity, etc. Hence, illuminated by continually emerging creative computing products like this, Creative Computing finally finds a way to truly fuse with creativity.

Therefore, it can be seen that the creativity contained in Creative Computing is relatively objective. Creativity can be studied objectively. Although there have been various kinds of difficulties with defining creativity (e.g. vagueness), there are still some particular natural phenomena closely related to it (e.g., sudden insight). Behind these, there may be associated knowledge (e.g., randomness). Hence, based on the idea of knowledge combination, it is possible for Creative Computing to study creativity objectively. Therefore, Creative Computing can be deemed as a new approach to studying creativity. However, it is worth mentioning that the study of Creative Computing on creativity mainly emphasises the knowledge or rules of creativity, rather than the ultimate nature of creativity.

2.3.2.2 Forms of Creativity

In Boden's opinion, creativity is the ability to create something new, surprising, and valuable [12]. As for the meaning of "new", there are two different degrees: new to a specific person (P-Creativity) and new to the whole world (H-Creativity) [12]. P-creativity is, most of the time, considered at the level of children's education. It is thought unnecessary for children to possess a huge amount of knowledge. One of the key concerns is to maintain children's original creativity. This could explain why there are coaching strategies encouraging people to think like a child. However, due to the urgent social requirements of creativity, H-creativity is more requisite and preferred by urgent contemporary competition. There is no doubt that the great success of big digital companies like Apple and Google depend enormously on

H-creativity. For global success, Creative Computing aims to facilitate people's H-creativity.

Creative Computing is also devoted to facilitating the combinational creativity of people. Similar to P-Creativity and H-Creativity being linked to the meaning of “new”, there are three types of creativity with respect to the meaning of “surprising”, which can be employed in big data manipulation to describe its creativity features: combinational creativity, transformational creativity and exploratory creativity [2, 95, 107].

- **Combinational Creativity:** it concerns combining familiar ideas to produce unfamiliar ones, i.e., via building relationships between different domains [11]. For instance, the Apple watch is a creative product that combines the functions of a watch and a smart phone.
- **Transformational Creativity:** transformational creativity, aims to transform the concepts or notions in one domain to another unconnected one [11]. For example, the famous painting “Tete de Femme” (Head of a Woman) of Picasso is a creative change in painting presentation, it borrows the shape of geometry to present the human face.
- **Exploratory Creativity:** this is different from the two creativities above. Exploratory creativity only focuses on one established conceptual space, such as exploring the use of different words to generate new words, even new ones which are not listed in the dictionary, as in cyberspeak.

Based on different forms of creative computing, using creative computing to process data is capable of increasing diversity and meaningfulness in the analysis, of extending the scope and depth of analysis, and of producing creative results to inspire creative minds. Methods of accessing and analysing these data are limited, and lack proper comprehension. Poetry data presentation means also need to be more interesting and inspirational. Data itself may rely on creative computing to trigger a better understanding of its content.

2.4 Summary

In this chapter, the background and state of poetry data and its challenging issues are introduced and reviewed firstly. Then, the requirements of creativity and basic concepts of creative computing are proposed.

- Digital Humanities is investigated as the background of this research. This section comprehensively reviews the cause of digital humanities, the attention that digital humanities have drawn and the research achievements of recent years. Poetry data management issues and the forthcoming challenges are illustrated. The current state of poetry data is reviewed, particularly in relation to this research. Other issues are also discussed.
- Creative computing has been explored in this research. This section discusses and classifies why people today desire creativity, the meanings of creative computing in three aspects and, last but not least, three forms of creative computing: combinational creativity, transformational creativity and exploratory creativity.

Chapter 3

Related Work on Poetry Data Manipulation

Objectives

- To describe current poetry data collection methods
 - To present poetry data analysis methods
 - To propose poetry data presentation methods
 - To introduce applied laws of physics, principles in chemistry and mathematics formulae
-

3.1 Overview

In this chapter, the related work of this research will be described in details. This chapter focusses on the state-of-the-art techniques of how poetry data are being collected, analysed and presented. Firstly, relevant websites of poetry databases are introduced. Secondly, a core retrieving technique, named the web crawler, is expanded upon. In terms of digital humanities, how literature data being analysed is then explained: e.g., author and style analysis and sentiment analysis. Finally, predominant methods of poetry appreciation and presentation are demonstrated. Based on these studies, this work absorbs and refines their advantages. Particularly in the aspect of creativity, creative poetry data management is proposed by combining the knowledge in science and engineering, such as the laws of Physics, principles in Chemistry and Mathematics formulae. This chapter also provides a critical review on the drawbacks of the poetry data manipulation techniques which essentially motivate this research.

3.2 Poetry Data Collection Methods

This section briefly illustrates common quantitative and qualitative data collection methods. Since they broadly lack generalisability, this research evaluates the power of the web crawler, which is then used as a universal raw data collection method. With the help of web crawler, the idea of a multi-purpose poetry data collection method is developed, based on multidisciplinary combinations.

3.2.1 Common Data Collection Methods

Data collection is usually the first step in any scientific research. Nowadays, one can access information from a broad range of sources. There are a variety of data collection methods: e.g., surveys, in-depth interviews, document studies, etc. A survey is a verification means which describes and explains social phenomena and their relationships via practical interviews and data collection [106]. It specialises in gathering first-hand data to describe an intangible entity. Surveys can be categorised into three types: exploration survey, descriptive research survey and explanation survey. Surveys have both advantages and disadvantages. The advantages include sampling part of the population to gain insight and draw conclusions in a relatively short period of time. Assumptions made before a survey can be tested by sampling experiments. However, surveys lack the capability of investigating individual cases or groups in terms of structural content. Moreover, multiple factors, such as ineffective ways of sampling, vague population ranges and an unreasonable selection of samples, are likely to lead to meaningless results.

In depth interviewing is an open, direct and personal means of interviewing, which is executed on a one-on-one basis to reveal some underlying background, attitudes, beliefs and emotions [38]. The merit of in-depth interviewing lies in the independence of the interviewer's opinion. Thus, it is more suitable for understanding complicated or profound issues which cannot be articulated in brief. An apparent drawback of this method is its lack of efficiency.

Document Studies refer to obtaining information via collecting, analysing and investigating statistics or reports, such as: library sources, information agents,

product manuals, etc. [56]. This method requires the confirmation of retrieving strategies and the support of computer and search engines. The following are common means of compiling document studies:

1. Gathering: including purchasing, posting and ordering documents;
2. Exchanging: as one of the major methods, exchanging classified information often takes place among companies;
3. Requesting: directly asking for information can be an effective method;
4. Copying: as an unusual means of leveraging information, copying documents, especially rare ones, also applies.

The most outstanding advantage of document studies lies in that it is time and effort efficient; therefore, many small companies adopt this method. However, the information that documents offer can be pointless or obsolete [57].

3.2.2 Poetry Data Collection Web Sources

In order to collect the data about poetry, related web sources provided an efficient pool. There are a large number of poetry websites for data collection:

- Poetry Foundation [72] – This is a website which publishes Poetry magazine. It consists of nearly 10000+ various poems and user can browse by filtering options such as subjects, occasions etc.
- Poem Hunter [73] - The site is a huge storehouse for poems. The user can search for poems by entering keywords.
- Poets [74] - The web includes various poetry programs such as Poets.org. It aims to support the careers of American Poets.
- Poetry Daily [75] – This site brings users a poem every day from books, magazines or journals. It helps people to read the poem more easily.
- Poetry Archive [76] – It is a website focussing on poets. Users can find the poets by searching their name or poem title. It also has some educational sources for children and teachers.

The above websites include various types of data about poets and poems. However, each of them is lacking some information, such as helping the user to write poems

or how to teach students to learn poetry. In order to extensively collect poetry data, the web crawler is used as a major poetry data collection method.

3.2.3 Web Data Collection — Web Crawler

Internet data collection is mainly achieved via the Internet crawler, also referred to as web crawler, spider or robot. An internet crawler is a piece of program or script that automatically obtains information from the World Wide Web with a certain pattern [91]. The information on the Internet is dispersed over hundreds of billions of websites which are separately stored in millions of servers around the world. Usually, web browsers only obtain information by hyperlinks by which they navigate through websites. The crawler is able to combine information from multiple sites and further analyse and excavate them by means of online (websites been downloaded) or offline (websites been saved) activity. If we suppose the World Wide Web is a static put-together of all websites, we hardly need to consume time crawling them all. If all the websites were withdrawn and stored, the job would have been done. But the World Wide Web is actually an ever-changing dynamic entity. As a result, it requires continuous crawling, such as adding, deleting, moving or manipulating information about hyperlinks or websites; in order to keep the applications updated.

There is a wide variety of means of exploiting web crawler. First of all, there is its use with commercial intelligence that enterprises or institutions can leverage to gather information on rivals or chances for cooperation. Then it can be adopted to monitor users' interested websites. When there emerge new messages at specific locations, users can be warned. There is the potential abuse of web crawler. For example, web crawler could be illegally used as a mean to steal personal information, which may be subsequently used to generate phishing, and thus deceive the user. The web crawler is most widely applied in supporting searching engines [30]. Indeed, these web crawlers are the main consumers of broadband Internet. Search engines construct indexes upon the websites located. Popular search machines, such as Google, Yahoo and MSN, have their own crawlers which are

extremely efficient in gathering various websites. However, there are some biased or specialised web crawlers that only download websites containing specific content.

Depending on the strategies by which the crawler grabs websites, web crawlers are generally categorised into three types: deep-prior searching strategy, broad-prior searching strategy and relevance-prior searching strategy [15].

- **Deep-prior Searching Strategy:** DPSS starts from the original URL address of the original website. At first, it obtains corresponding websites of the original URL address, and extracts new URL addresses from which the DPSS selects a new URL as a starting point. The above procedures are executed until new addresses cannot be obtained any longer. Then the other routes are processed similarly. In DPSS, the subset nodes come in precedent to nodes of the same layer. The design of DPSS is rather simple. Meanwhile, the principle of DPSS determines that the crawler will enter deeper levels of websites until no new addresses are returned; but given the huge number of websites among which a great deal of useless ones exist, the crawler of DPSS digging along one path would probably end up in a vain search, or trapped. Moreover, the deep digging of DPSS certainly affects the accuracy and efficiency of grabbing websites. Thus, the setting of digging depth is critical to DPSS. Overall, DPSS requires elaborate depth setting in order to evade trapping issues, and can only be applied on a smaller scale.
- **Broad-prior Searching Strategy:** BPSS, also called extension-prior searching strategy, exhaustively searches by hierarchy. It starts from the URL address and obtains all directly related URL addresses to the origin, upon which BPSS locates new directly related URL addresses. Note that the URL address obtained in advance is processed in precedence to the later URL addresses. Specifically, the next layer will not be processed until all the current layers of URL addresses have been searched. The principle of BPSS was developed from the Breadth First Search Algorithm [56]. BPSS treats all URL addresses on the original website as nodes of a digraph, which are regarded as direct son nodes of their upper layers, and thus, considered better related and prime to grab. BPSS is

primarily applied in cases when extensive and comprehensive captures are required. Sometimes BPSS is combined with website filtering techniques to eliminate irrelevant results. However, in practice the efficiency of this proposal is greatly reduced due to the effort used in downloading many irrelevant websites and analysing them for relevance.

- **Relevance-Prior Searching Strategy:** RPSS refers to those systems that give priority to the queue of crawlers and unvisited websites. The priority is calculated upon the topology features, page features (e.g. the relativity between user query and the original page containing the hyperlink), etc. Restricted crawlers are used to capture a specific type of websites. RPSS selects the most fascinating types of websites from the queue depending on the predicted text classification score. Theme crawlers do not provide adequate trainable samples with markers, but only a small portion with subset website, or some theme description for the interest of the user or community. It does not have text classification, but by comparing visited websites and theme descriptions, the theme crawler effectively browses and captures preferred zones of the Web.

Since DPSS might suffer from trapped issues and to predetermine depth is indispensable, currently BPSS and RPSS are more widely used.

As topic data collection is designed for certain professional groups, the data is only restricted to specific areas. Furthermore, concerning the performance and consumption of exhaustive searching the topic data collection often needs to discuss how to estimate the relevance between hyperlink directions and the topic, and to decide whether the website is worth visiting. It also requires further investigation into how to crawl the Web in order to collect more valuable websites, with less irrelevance.

The working process of the focused crawler appears more complicated than universal crawlers. Firstly, algorithms should be used to get rid of irrelevant URL addresses and only keep those related to the topic of which the hyperlinks are to be added to the unprocessed queue. Then, the target URL address must be obtained through a certain crawling approach. The above analysing and processing

procedures are repeated until terminating conditions are reached. Meanwhile, the crawler will save the websites that have been processed and create indexes in order to raise the search efficiency. The site-saving will also provide effective feedback and guidance for any future crawling strategy.

In comparison to traditional web crawlers, a focused crawler is more complex. The following issues should be attended to

1. How to precisely define the aiming topic;
2. How to effectively analyse and filter the websites and data to be processed;
3. How to design an efficient optimisation method and crawling strategy for the URL queue.

The definition and description of the topic aim determines the analysis algorithm needed, which also provides the foundation for obtaining URL addresses from websites. At the same time, the optimisation algorithms for the website analysis and the URL address queue to be processed, that decides the efficiency of the crawler, together with the crawling strategy for the web crawler play as two significant parts that comprise the crawler. The relationship between the two is indeed very close.

Based on the above review on data collection methods, the aim of this work is to develop a more comprehensive and systematic poetry data collection method. This work proposes a multi-purpose poetry data collection method to acquire related data from different sources with the help of a web crawler. Several examples of data collection and pre-processing are created by combining the laws of physics.

3.3 Poetry Data Analysis Methods

In this section, common data analysis principles and techniques are introduced. Theoretical foundations of literature data analysis are listed in terms of digital humanities, e.g., author and style analysis and sentiment analysis. In order to enhance creativity, a multi-dimension poetry data analysis method is developed based on the combination of state-of-the-art analysis methods and chemical principles.

3.3.1 Common Data Analysis Methods

Data analysis is usually the core of data manipulation. There are a wide variety of data analysis methods, e.g., machine learning, statistics, regression analysis and exploratory analysis etc. Machine learning is a branch of Artificial Intelligence. The basic hypothesis is to develop algorithms that make computers ‘learn’ automatically. Machine learning implies the sort of algorithms that are able to extract patterns from data and to predict unknown information. There are some well-known algorithms in machine learning, such as Naive Bayes [88], K-Nearest Neighbours [1, 61], Decision Tree [77], Support Vector Machines [17], etc. From a perspective of data combination and transformation, this research creatively brings about together the KNN and Decision Tree algorithms of machine learning. Specifically, qualitative data analysis in machine learning and the proposed creative rules are being combined to perform creative analysis on data.

Statistics also play an important role in data analysis. Such analysis consists of automated or semi-automated techniques for discovering previously unknown patterns in data, including relationships that can be used for the prediction of user-relevant quantities [62].

There are many algorithms in statistics, such as Linear Regression, Exploratory Analysis and Random Forest, etc. From a data similarity and diversity point of view, interdisciplinary creative analysis is carried out by incorporating regression and exploratory algorithms in the field of statistics, in which quantitative data analysis methods are joined with the proposed creative rules.

Regression Analysis aims to build a regression function based on the analysis of the relationship between variables and dependant variables. The regression function will be regarded as a prediction model which estimates the related trend of dependant variables based on observation of the variables. Regression analysis is a practical, effective and valuable prediction method [33]. Regression is primarily used for prediction and causal inference. In its simplest (bivariate) form, regression shows the relationship between one independent variable x and a dependent variable y , as in the formula below:

$$Y = \beta_0 + \beta_1 x + u \quad (3.1)$$

Before applying regression analysis it is necessary to make sure the variables are correlated. If not, error results are expected.

The so called Exploratory Data Analysis (EDA) refers to exploring obtained data (especially raw data gathered by survey or observations) through plot, table, regression and calculating characteristics in order to draw patterns of the data with the least predetermined assumptions. EDA has three significant features including the least manipulation of data, flexible analysing methods and straightforward analysing tools [97].

The above methods are more commonly applied to numerical data. For literature data, especially poetry data, it is more significant to interpret the underlying emotions and the main theme of expression.

3.3.2 Literature Analysis Methods

In recent studies of the Digital Humanities, the emphasis has been on literary text analysis, including the personalities of the author, style, emotion and the relationships between characters, etc.

Author and Style Analysis

Literature analysis, especially classical literature and historiography, often involves the discriminating compositions of unknown origin. The aim of using text mining is to identify the author by analysing the style of a composition. Style analysis relies on a quantitative evaluation of the pattern of words and grammar using a computer. The basic hypothesis is that each author has his or her immutable writing characteristics. The aim of style analysis is to address the characteristics of different authors. Researchers have also investigated the gender of authors by text mining a wide range of data. Emily [83] researched the gender of authors among ancient Greek compositions. Shlomo et al. [31] studied that of French compositions and drew quantitative evidence on different writing styles between male and female.

Sentiment Analysis

Analysing the emotion of a composition is one of the most significant purposes of literature analysis. Conventional methods of analysing emotion can only perform simple statistical analysis on short texts. However, text mining is able to analyse emotions of a large amount of literature; their degree of intensity as well as representative words. Analysing emotion is always related to the context, i.e., the content and theme of a composition, whereas, the style which presents personal preferences is irrelevant to the context. There is much successful research that has been accomplished by analysing emotions in literature. Plaisant et al. [105] performed text mining on the letters of Emily Dickinson, a famous female poet of the 19th century, through categorisation of sexual features in her poems. This work was completed with the aid of experts on Dickinson's poetry. Based on a combination of automatic categorisation and artificial discrimination, indicators of sexual features were found which help to enhance the accuracy of categorisation.

Mining of the Relationships between Characters Analysis

Mining of the relationships between characters grammar analysis refers to the method of constructing relationships networks within a composition. This involves the techniques of name-object identification and relationship extraction. Using information extraction of text mining, one can delineate the social network in a novel which enables researchers to better understand the relationships between characters as well as social backgrounds. This approach overcomes the drawback of conventional methods in which only one piece of work can be analysed at a time. This approach can also calibrate the conventional analysis results.

Many researchers have attempted to work out the relationship network of compositions using computer modelling. Elson et al. [28] performed text mining on 60 novels and periodicals of the 19th century. By addressing conversations between characters, this method successfully built relationship networks. Then, the characteristics of the social networks, as well as the correlation between the social network and the background of the composition, were investigated. The result shows that novels of that time do not conform to the two features revealed by empirical analysis. Therefore, new theories of the social characteristics of that time

were proposed which describes the social characteristics as a function of the number of characters in a novel, its format, and the background of the time. This analysing approach significantly challenges traditional conclusions and provides new perspectives. In another work, Elson et al. [28] expanded on how to draw social networks from a composition.

Mining the relationships between characters is a frontier research method which mainly applies to novels. These approaches also provide inspiration for the research, in which mining poetry data may lead to creative analysis results.

3.3.3 Poetry Analysis Methods

People like poems not only because reading poems brings joy but because they meet the needs, offer a sense of comfort, and inspire us. However, to comprehend poems is not as simple as reading a newspaper or novel, since poems employ multiple aspects of language. Usually, information or knowledge are communicated with only one degree of language, i.e., the degree of comprehension, which refers to the thoroughness of understanding. Poems, on the other hand, also possess a degree of sense, a degree of emotion and a degree of imagination. Therefore, in order to reveal the underlying emotion and meaning it is essential that poems should be approached from multiple dimensions.

According to the Poetry Analysis Example [18] proposed by the Oregon Arts Commission, poetry analysis can be divided into ten dimensions:

1. Poem Published Time and Age
2. Poet Information
3. Structure Analysis, including word count, lines and stanzas.
4. Topic
5. Theme
6. Mood
7. Personal Reflection
8. Vocabulary
9. Poetic Devices, including rhyme, meter, alliteration, repetition and imagery.
10. Effect of the poem

In this research, the emotional aspects of the multiple dimensions are of the interest. Hence, a multi-dimensional poetry data analysis method is proposed based on chemical principles in which three different rules are designed to evaluate mood, theme and personal reflection in a poem.

3.4 Poetry Data Presentation Methods

This section reviews common data presentation methods. In pursuit of enhancing creativity, the idea of a multi-media poetry data presentation method, which utilises the visual effects of geometrical shapes and curves, is explained. Three rules are designed to present the mood, theme and personal reflection of a poem, respectively.

3.4.1 Common Data Presentation Methods

A table is a critical object in terms of a database. It is the basis of other objects. Depending on different categories of information, a database may contain several tables, which are correlated with others to form the database [46].

Graphs are a useful method of displaying quantitative data. The standard graph uses two rectangular co-ordinates (called the x and y axes) [96]. There are several different types of graphs, such as histograms [89], arithmetic scale line graphs [4], frequency polygons [62] and scatter diagrams [26].

The stem-and-leaf display, also called the stem plot, was designed by Arthur Bowley [98]. It compares the digits of data in the data set, and regards a relatively constant digit as a stem while digits with larger variations are regarded as branches. There are two merits of using the Stem-and-Leaf display: firstly, no data is lost on the plot; secondly, data are easily accessible. Yet a stem-and-leaf display is merely good at presenting data with a closed number of digits; and it can be tricky to record more than two sets of data.

It is often convenient to present data as a picture. Information can be delivered much more quickly by a diagram than by a table of numbers. Charts and diagrams can help a reader or audience quickly grasp the salient point of an analysis or report [3]. Unfortunately, because diagrams can be very misleading, they should be treated as a

complement to numbers, not a replacement. There are several types of charts which can be used to illustrate data: pie charts [20], bar charts [80] and line charts [2].

In brief, various types of data presentation methods have been presented. The most significant concern is to choose an effective data presentation tool according to user needs. For example, pie charts show the parts compared to the whole, while a continuous line graph shows a trend, explicitly. With any method, it is necessary to create easily understandable labels and position them properly.

3.4.2 Poetry Data Presentation Methods

A paper in the Scientific American in February, 2017, synthesises previous research results on analysing literature using mathematic approaches [32]. Researchers from the University of Vermont carried out a study on the Gutenberg Project, the very first digital library, in which they visualised the emotions of 1327 works of fiction in English. The trends of emotions throughout the works, i.e., the measures of grief and joy, were plotted against chapters. Researchers from the University of Vermont named the envelopes as the ‘emotional curves’ of these works [19].

The Nuclear Physics Association of Poland [27] has published an interesting piece of research in which the lengths of sentence from 113 literary compositions of different languages were ranked and plotted as a bar chart. When they looked at these bars, they found that the variation of the bars formed a fractal image. A fractal image indicates a complex geometric shape consisting of multiple fractions, which assemble the entire shape. Therefore, there is surely a sort of mathematic regularity within these compositions.

It can be seen from the above examples that the underlying information of literature can be revealed using different mathematics approaches. In the research, geometric shapes and curves have been exploited to present mood, theme and personal reflections of poetry data to generate creativity when reading.

3.5 Relevant Laws of Physics, Principles in Chemistry and Mathematics Formulae

The poetry data management methods introduced in previous sections also illuminate the research into creative poetry data manipulation. The adoption of Creative Computing, which combines scientific theories with human art, aims to solving problems in either area or generating new ideas. In this thesis, poetry data are regarded as objects and analysed quantitatively and qualitatively to bring new interpretation and comprehension.

Based on the theories of Creative Computing, this research utilises the laws of physics, chemical principles and geometric shapes and curves to collect, analyse and present poetry data, in pursuit of inspiring creative minds. This section briefly goes through the related concepts in science and engineering.

3.5.1 Laws of Physics

Newton's law of universal gravitation [82] states that a particle attracts every other particle in the universe using a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

Newton's law of universal gravitation can be written as below:

$$F = G \frac{m_1 m_2}{r^2} \quad (3.2)$$

in which G is the gravitational constant. Terms m_1 and m_2 are masses of two objects. Term r denotes the distance. One of the most significant characteristics of this formula is that a quantified measure of relationship can be derived based on the nature of each object.

Archimedes' buoyant force principle [96] considers any object of arbitrary shape and volume V surrounded by a liquid. The force the liquid exerts on an object within the liquid is equal to the weight of the liquid with a volume equal to that of the object. If the buoyancy force is greater or less than the gravity of an object, it will float up onto the surface of the liquid or sink. In this way, the volume of an object is

the key to determine the floating state which can be used to filter objects of different densities.

Besides the above filtering processes, the properties of objects can also be quantified using laws of Physics. Ohm's law [5] states that the current, I , through a conductor between two points is always equal to the voltage, V , across them divided by the resistance, R , which can be represented as:

$$I = \frac{V}{R} \quad (3.3)$$

This can be adopted to define the conductivity of electronic circuits in terms of the resistance of elements being used.

3.5.2 Principles in Chemistry

In chemistry, pH is a numeric scale used to specify the acidity or basicity of an aqueous solution which is approximately the negative of base 10 logarithm of the molar concentration, measured in units of moles per litre, of hydrogen ions [98]. Such bipolarity definition of a solution may be used to measure the degree of positive or negative of literature data.

As organic substances are constructed from carbon based atoms, an organic molecule is simply an accumulation of the number of elements it is comprised of [101]. The statistical fashion of naming chemical organics may be used to count the intensity of key data.

The reduction-oxidation reaction [86] is a chemical reaction in which the oxidation states of atoms are changed. Any such reaction involves both a reduction process and a complementary oxidation process in which new chemical compound is generated. Hence, such chemical reactions may inspire the generation of brand new literature data analysis results.

3.5.3 Mathematics Formulae

According to chromatics and geometry, colourful curves and shapes of mathematical elegance take effect on the performance of presentation. An emoticon is a pictorial representation of a facial expression using punctuation marks, numbers and letters

[69]. It usually is written to express a person's feelings or mood. The most famous example is that in which Ambrose Bierce [6] proposed an improvement in punctuation- the snigger point, or note of cachinnation; it is written thus “◡”and represents a smiling mouth. Based on this theory, shapes and curves with special emotional indications can be explored and adopted to present literature data in pursuit of unique reflections for audience.

3.6 Summary

In this chapter, common poetry data collection methods are introduced and reviewed firstly; then, the related studies of poetry data analysis methods and poetry data presentation methods have been discussed:

- Common poetry data collection methods are studied for this research. This section presents, classifies and explores the data collection methods. Common data collection methods, such as surveys, interviews and documentation, have been investigated. However, traditional data collection methods are too simple to meet diverse needs. Prevailing web sources about poetry are reviewed along with a discussion on their strengths and weaknesses. These web sources generally lack creativity and systematisms. A web crawler is introduced and discussed as the kernel collection technique of this research.
- Related poetry data analysis methods are summarised, common data analysis principles and techniques are introduced. Theoretical foundations of literature data analysis are listed in terms of digital humanities, e.g., author and style analysis, sentiment analysis and the mining of the relationships between characters' analysis. Although these researches provide a good foundation, creativity and deep excavation of data are still missing. In order to enhance creativity, a multi-dimensional poetry data analysis method is developed, based on the combination of state-of-the-art analysis methods and chemical principles.

- This chapter provides an overview of data presentation methods, which includes tables, graphs, stem-and-leaf plotting and charts. It also presents two examples showing that underlying information in literature can be revealed using different mathematics approaches. In the research, geometric shapes and curves have been exploited to present mood, theme and personal reflections in poetry data, to generate creativity when reading.
- Finally, this section goes through related concepts in the laws of Physics, Chemistry principles and geometric shapes which will be applied in the proposed creative poetry data manipulation approach.

Chapter 4

A Creative Computing Approach to Poetry Data Collection, Analysis and Presentation

Objectives

- To present an overview of the proposed creative poetry data manipulation approach
 - To introduce the framework of the approach
 - To describe the poetry data collection, poetry data analysis and poetry data presentation phases.
-

4.1 Overview

Through a critical review of the literature, it can be seen that the complexity of current data has become so complex that traditional data manipulation schemes are inefficient. Furthermore, there are diverse formats whereby useful data could be stored, e.g., text, video, and audio. In order to access and exploit the information in these data, a comprehensive and inspiring data manipulation approach is required. This data manipulation approach is also expected to be able to reveal valuable information underlying the plain text. Moreover, the accuracy of a data manipulation scheme is no longer the only measure of its success. An ideal data manipulation scheme should consider user preference, representativeness and the contingency to inspire creative minds.

Many researchers have worked on poetry data manipulation and proposed different methods. However, owing to the new challenges and creative requirements, traditional methods need adjustments. For example, most poetry data collection

methods, such as surveys, are based on single target searching; that is, only relying on keywords and themes. Thus, the results can be restricted. Moreover, the accuracy of algorithms for poetry data analysis is no longer the only benchmark. The underlying meaning of data has attracted people's attention. Thus, traditional poetry data presentation methods need to be enhanced in terms of diversity and media richness.

In order to meet the need of creative manipulation of poetry data, this research launches from the theoretical basis of Creative Computing. The aim of Creative Computing is to combine or transform diverse knowledge from different areas to solve a problem or bring about valuable new product. In terms of poetry literature of humanities subject, it is an innovative and meaningful attempt to introduce scientific laws and principles. In order to reveal the humanity comprehension of poetry data, this research regards poems as objects which are being analysed quantitatively and qualitatively to present new discoveries.

In this chapter, a comprehensive creative computing approach is presented. The proposed data collection method aims to improve poetry data collection based on Section 3.2.2. It is explained that the proposed method covers a broader scale and types of data sources, and provides pre-processing on the data compared to the web crawler methods [15, 30]. The proposed data analysis method is able to provide a more systematic analysis which outperforms traditional data analysis methods [83, 31]. Based on references [105, 18], a multi-dimension poetry data analysis method is proposed to explore the emotions as well as meaningfulness of poetry data. In [32, 27, 19] mathematics methods are adopted to study literature which inspires us to present poetry data in a creative way.

In the following sections, the framework of the approach is presented in details. Furthermore, various phases, rules and algorithms are explained.

4.2 Framework of Creative Computing Approach to Poetry Data Manipulation

A unified flowchart is generated to illustrate the proposed approach as displayed in Figure 4.1, which details the poetry data manipulation procedures. The entire process can be deconstructed into separate phases, activities and tasks, and structured into different abstract levels with different system models. Generally, Figure 4.1 demonstrates the layout of the proposed approach upon which diverse models and hierarchy analysis methods are adopted.

As is sketched in Figure 4.1, the entire framework is comprised of three primary phrases including:

- **Multi-Purpose Poetry Data Collection.** The task of this phase is to acquire related poetry data from different sources with the help of web crawler. Data relevancy is achieved by applying the designed word gravitation rule. The data are then stored in a unified system, according to a series of standards. After that, the data are pre-processed through the designed flotation filter and conductivity calculator rules. The processed data can then be stored and imported into the data analysis models. This process is intended to give users more options to obtain the related poetry data, and inspire creative minds.
- **Multi-Dimension Poetry Data Analysis.** This is the phase in which the processed poetry data are analysed, and more attention is paid to digging into the meaning of the data. Some creative models and algorithms are proposed and used to analyse the mood, theme and personal reflection. The relationships between the data and the chosen poetry data are established, and then the results of the analysis are generated for data presentation. This phase aims to provide users with experiences that accurately reflect their wants and needs in an authentic and creative manner.
- **Multi-Media Poetry Data Presentation.** In this stage, the poetry analysis results are presented, with a focus on showing the diversity and media richness of the data. It builds the bridge between the colour and the emotional data, and

presents them through mathematics curves and shapes. In order to provide users with experiences that accurately reflect their wants and needs in an aesthetic and creative manner, the proposed approach not only demonstrates how to present data, but also comes up with creative rules to enhance the poetry data visualisation. The mood, theme and personal reflection analysis results are presented following different creative rule

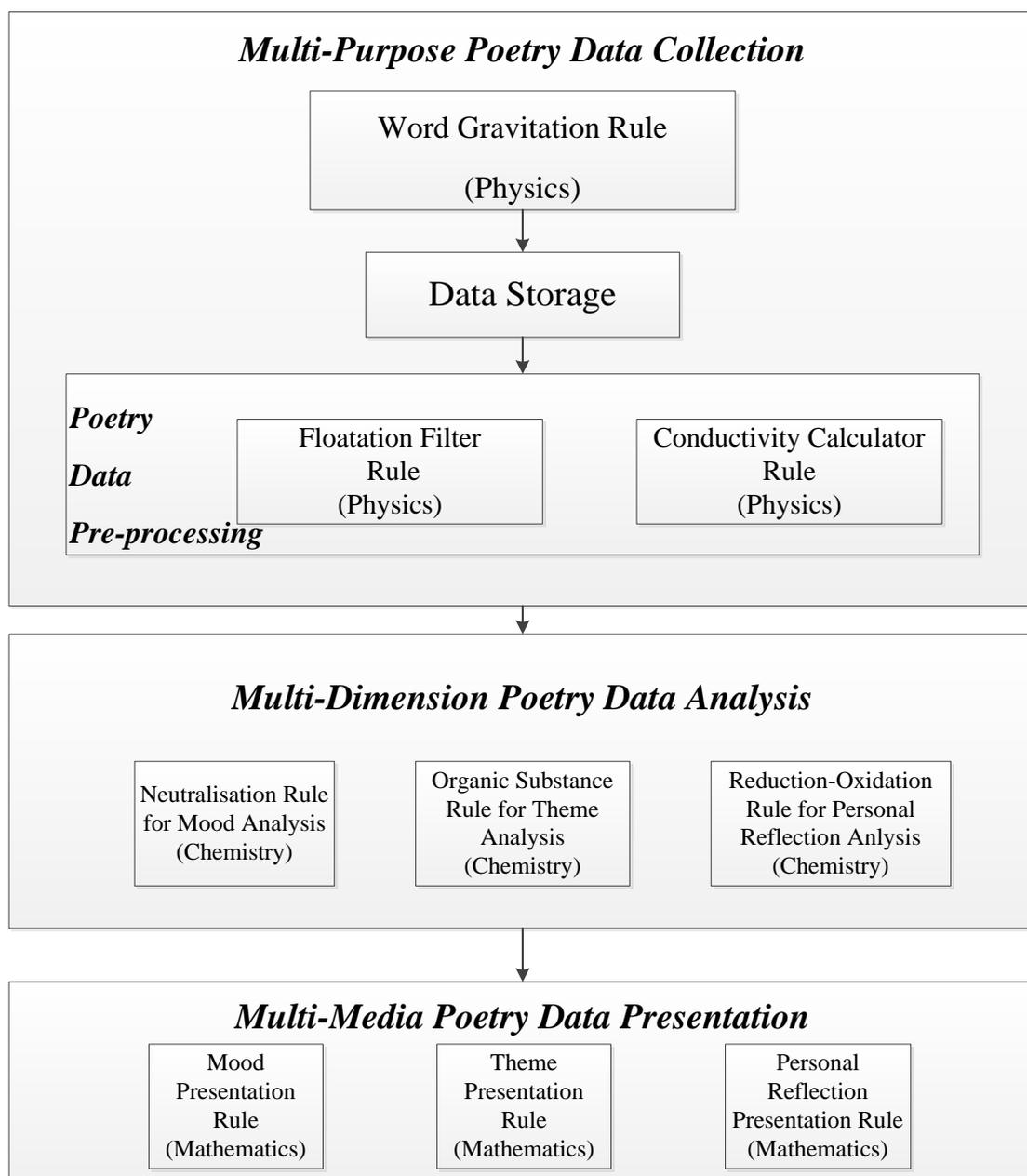


Figure 4.1 Framework of Creative Computing Approach to Poetry Data Manipulation

4.3 Creative Poetry Data Manipulation Phases

4.3.1 Multi-Purpose Poetry Data Collection Phase

The data collection phases include two stages: data collection and data pre-processing.

The first step of the proposed approach is to collect various types of poetry data for the later poetry data analysis and presentation phases. This work focuses on the meaning of poetry data and their relationships. It introduces a multi-purpose approach which considers the poetry data in terms of three physical properties: mass, volume and conductivity. The proposed approach not only considers the meaning of the data but also introduces physical theories into literature in order to develop creative rules. By improving the web crawler algorithm, the work aims to offer the user more options for obtaining the related data, and inspire creative minds.

There are some common poetry data quality issues to confront, such as noise. In order to solve these problems, this work presents a series of steps of poetry data pre-processing based on a set of designed creative Flotation Filter and Conductivity Calculator rules. After that the processed data will be stored in the database for analysis and presentation.

- **The Word Gravitation Rule (WGR)** has been designed based on Newton's law of universal gravitation [84], and which is intended to calculate the attractive force, i.e., the gravity, between the keyword of user input and any word among the list of its synonyms. It provides more creative possibilities and diversity for data collection. In order to realise the proposed word gravitation data collection method, an improved focused crawling process, which takes the proposed creative rule as input (i.e., considers the keyword and candidate words as search conditions), is presented.

- **The Flotation Filter Rule (FFR)** is designed to reduce the scope and size of data for a more precise poetry data analysis, based on Archimedes' buoyant force principle [85]. The scope and size of data can be quantified as the shape and volume of objects.
- **The Conductivity Calculator Rule (CCR)** is designed based on Ohm's law [86]. It aims to act as a final filter for the collected data through their internal properties.

4.3.2 Multi-Dimension Poetry Data Analysis Phase

The second step of the proposed approach is to analyse the processed poetry data. This phase introduces a multi-dimension approach which considers the poetry data meanings from three aspects: mood, theme, and personal reflection. This phase uses chemical principles for reference, creatively regarding data as chemical elements, treating data combinations as chemical reactions in order to infer the sentiment contained in therein. This phase presents three creative multi-dimension poetry data analysis rules. They help to excavate the sentiment of poetry from three aspects - mood, theme and personal reflection in order to satisfy the user requirements of enlightenment and creativity.

- **The Neutralisation Rule (NR)** is designed to analyse the mood of the poetry data based on chemical neutralisation reactions. By analysing the *pH* of a poem, novel or paper, an innovative acidity measure is applied to convey the mood for the users.
- **The Organic Substance Rule for Theme Analysis (OSR)** is inspired by the concept of organic compounds in chemistry. It equates six themes of poems with six chemical elements. The overall theme can be estimated by accumulating the credit of each word in a poem, and the most significant count of an element will be recognised as the main theme.
- **The Reduction-Oxidation Rule for Personal Reflection Analysis (ROR)** is based on the idea that a substance transforms differently under specific conditions, triggering the development of the personal reflection analyser. Based on the six categories of themes of a piece of literature, personal reflection

varies as the measure of mood differs. In other words, the *pH* of a text plays the role of an oxidising agent or reducing agent, which leads to personal reflection.

4.3.3 Multi-Media Poetry Data Presentation Phase

The final step of the proposed approach is to present the poetry data analysis results. To provide users with experiences that accurately reflect their wants and needs in an aesthetic and creative manner, this thesis introduces a multi-media poetry data presentation approach which builds connections between geometric shapes in mathematics and the emotional data. The proposed approach not only considers how to present data, but also uses creative rules to enhance the poetry data visualisation. By introducing creative presentation rules, an emotion-based multi-media poetry data presentation method is provided and examples are used to demonstrate how the proposed method works.

- **The Mood Presentation Rule (MPR)** is designed to present the mood of the poetry data based on a parabolic curve in mathematics. Based on this method, this chapter combines emoticons and the parabolic curve in mathematics to propose a creative mood presentation rule. It aims to present the mood of a poem immediately and creatively.
- **The Theme Presentation Rule (TPR)** uses a spectrum of colours to show the theme derived from an organic histogram analyser. It offers a vivid and polychromatic theme presentation to users.
- **The Personal Reflection Presentation Rule (PRPR)** presents 12 kinds of personal reflections, presenting literature along 12 types of curves or shapes to display these reflections. It provides a diversity of visual personal reflections to the users.

4.4 Summary

The aim of this thesis is to propose and implement a creative poetry data manipulation approach. Since the traditional way of comprehending literary data is subjective, laws and principles are integrated in science and engineering to the

poetry data manipulation approach. Such a multidisciplinary approach synthesises both qualitative and quantitative analysis patterns to generate brand-new results in terms of creativity. A Poetrypedia system, which incorporates the poetry data manipulation approach, is developed and opened to users in pursuit of inspiring creative minds, as well as fulfilling the usefulness in terms of creativity.

- An overview of the proposed approach of creative poetry data manipulation is presented, which is based on creative computing and relevant techniques. Reasoning rules, which are the main part of collecting, analysing and presenting poetry data are detailed into its theoretical foundations and design mechanism.
- A Multi-Purpose Poetry Data Collection phase is depicted with specific processing procedures and supporting techniques. The collection rules and algorithms, which are illustrated in details in Chapter 5, are also introduced.
- The phase of Multi-Dimension Poetry Data Analysis is explained in more detail, with its supporting techniques. It presents how the reasoning rules work to support the poetry data analysis, including mood analysis, theme analysis and personal reflection analysis. The reasoning rules are demonstrated in Chapter 6.
- The phase of Multi-Media Poetry Data Presentation is introduced by designing processes and relevant techniques. This section displays how the design rules support creative data presentation. The rules are further discussed in Chapter 7.

Chapter 5

A Multi-Purpose Poetry Data Collection Method

Objectives

- To present a creative multi-purpose poetry data collection method
 - To propose a word gravitation rule designed for poetry data collection
 - To introduce the architecture and algorithms for web crawler implementation
 - To depict the flotation filter and conductivity calculator rules designed for poetry data pre-processing
-

5.1 Overview

Poetry data collection is the source and technical foundation of poetry data manipulation. There are extremely rich sources of poetry data and diverse data types. Many approaches have been proposed to implement data collection; for example, surveys, interviews and document studies. However, most of the methods in the literature are based on single target searching, which only relies on keywords and a theme. Consequently, the results can be monotonous. Such information hardly paints a holistic picture of the data, and leaves out the most important information: its features, forms and relationships with others. Though some of the results possess a certain degree of relativity, creativity is missing such that they would offer neither effective nor inspiring information.

To provide users with experiences that accurately reflect their wants and needs in an authentic and creative manner, poetry data collection needs to pay more attention to various forms of related poetry data. This chapter introduces a multi-purpose

approach which considers the poetry data in terms of three physical properties: mass, volume and conductivity. The proposed approach not only considers the meaning of the data, but also introduces physical theories into literature to develop creative rules. By improving the web crawler algorithm, the work contributes to allowing users more options to obtain the related poetry data and inspire creative minds. In order to improve the quality of poetry data analysis, poetry data pre-processing is crucial. This chapter presents flotation filter and conductivity calculator rules designed for filtering poetry data.

5.2 Using Web Crawler for Multi-Purpose Poetry Data Collection

This section introduces a creative approach by applying web crawler to implement multi-purpose poetry data collection. Firstly, the definition and categorisation of poetry data relevancy are introduced. Then, a creative word gravitation rule is proposed for poetry data collection. Finally, this chapter presents the architecture and algorithms designed for web crawler implementation.

5.2.1 Poetry Data Relevancy

Data relevance is defined as a measure of data quality, a level of consistency between data content and the field of interests of users [52], which cares about whether the available data sheds light on the issues crucial for the users. Evaluating relevance can be quite subjective, and relies largely upon the various needs of users. The trick lies in balancing the trade-off between the needs of current and future users to generate data which agrees with the most significant needs with the given resource limits.

This work focuses on the meaning of poetry data and their relationships. In this research, Newton's law of universal gravitation [101] is borrowed and amended to quantify the data relevancy. Newton's law of universal gravitation states that a particle attracts every other particle in the universe using a force that is directly

proportional to the product of their masses and inversely proportional to the square of the distance between them, as shown in Figure 5.1 [101].

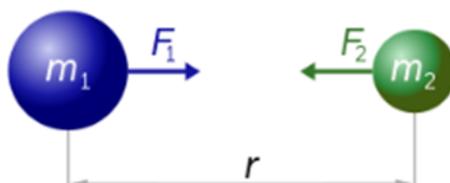


Figure 5.1 Newton's Law of Universal Gravitation [101]

This work treats a word as one object, whose mass is determined by its characteristics. The attractive force between the words is obtained where the larger the attraction, the higher relevancy there is. This following section presents the creative word gravitation rule. It helps to excavate the meaning of poetry data from different aspects, and seeks the relationships between data fulfil the user requirements of diversity and creativity.

5.2.2 Word Gravitation Rule (WGR)

The Word Gravitation Rule (WGR) is designed to calculate the attractive force, i.e. the gravity, between the keyword of user input and any word among the list of its synonyms. Newton's law of universal gravitation is borrowed to quantify the relevancy between words as below:

$$F = G \frac{m_1 m_2}{r^2} \quad (5.1)$$

Since the core of this law lies in the negative correlation between object masses and their separation distance, it needs to specify that the mass of a word is equal to a nine-element (the characteristic or property of a word has been divided into nine general classes [70]) vector comprising the number of meanings in each characteristic of the word. The vector dimension is defined in the order of Table 5.1 below:

Table 5.1 Word Classes

Noun	Verb	Adj.	Adv.	Pron.	Prep.	Conj.	Det.	Excl.
------	------	------	------	-------	-------	-------	------	-------

For example, the mass of 'blue' yields

$$m('blue') = [7, 3, 4, 0, 0, 0, 0, 0, 0]$$

The numbers denote the fact that 'blue' has seven meanings as a noun, three meanings as a verb and four meanings as an adjective, according to its definitions in the Oxford Dictionary. The distance is equivalent to a constant in different contexts, such as thesis, poems, novels, etc. In this way, the vector product is constructed by product of meanings for each characteristic of the word, and those pair of words with multiple meanings in common characteristics will get more credit, in other words, have a stronger attractive force.

Hence, m_1 and m_2 are the nine-dimensional masses of meanings for the keyword and the retrieved word, respectively. In terms of a poem, which is of interest, r is assigned 1 for the ease of computation. Constant G designates the scale at which the attractive force is measured.

Taking keyword '*wind*' for example, there are six given interpretations as a noun and nine meanings as a verb in the Oxford Dictionary. The mass of '*wind*' can be represented as:

$$m_1('wind') = [6, 9, 0, 0, 0, 0, 0, 0, 0]$$

With reference to the Thesaurus Synonym Dictionary [93], there are 22 synonyms of '*wind*'. Here the top six words '*air*', '*blast*', '*breeze*', '*cyclone*', '*gale*', and '*gust*', which are selected as candidate words m_2 . The attractive forces between '*wind*' and its candidate synonyms can be calculated as below:

$$F('wind', 'air') = 1 * \frac{[6, 9, 0, 0, 0, 0, 0, 0, 0] * [4, 2, 0, 0, 0, 0, 0, 0, 0]^T}{1^2} = 42$$

$$F('wind', 'blast') = 1 * \frac{[6, 9, 0, 0, 0, 0, 0, 0, 0] * [5, 5, 0, 0, 0, 0, 0, 0, 0]^T}{1^2} = 75$$

$$F('wind', 'breeze') = 1 * \frac{[6, 9, 0, 0, 0, 0, 0, 0, 0] * [3, 1, 0, 0, 0, 0, 0, 0, 0]^T}{1^2} = 27$$

$$F('wind', 'cyclone') = 1 * \frac{[6, 9, 0, 0, 0, 0, 0, 0, 0] * [1, 0, 0, 0, 0, 0, 0, 0, 0]^T}{1^2} = 6$$

$$F('wind', 'gale') = 1 * \frac{[6, 9, 0, 0, 0, 0, 0, 0, 0] * [2, 0, 0, 0, 0, 0, 0, 0, 0]^T}{1^2} = 21$$

$$F('wind', 'gust') = 1 * \frac{[6, 9, 0, 0, 0, 0, 0, 0, 0] * [1, 1, 0, 0, 0, 0, 0, 0, 0]^T}{1^2} = 15$$

Therefore, the relevancy between '*wind*' and its six candidate synonyms can be sorted as below:

1. F('wind', 'blast')
2. F('wind', 'air')
3. F('wind', 'breeze')
4. F('wind', 'gale')
5. F('wind', 'gust')
6. F('wind', 'cyclone')

The top three candidate words: '*blast*', '*air*' and '*breeze*' can be treated as preferential candidate words. Now, the search keyword, '*wind*' and the three candidate words will be put into web crawler together to collect related poems and poetry information.

5.2.3 Web Crawler Implementation

In order to realise the proposed word gravitation data collection method, an improved focused crawling, which takes the proposed creative rule as input (i.e., considers the keyword and candidate words as search conditions), is implemented. Focused crawler scan pre-downloaded poems based on the input keywords. It searches deeper when relevant poems are found, based on the selected candidate keywords and main keywords, and stops searching at poems irrelevant to the keywords.

5.2.3.1 Architecture of the Focused Crawling

Figure 5.2 shows the architecture of an adapted focused crawling system. A set of unvisited poems maintained by the crawler is stored in the poem database. The poem database obtains pre-downloaded poems from the internet. A relevance calculator estimates the relevance of a poem with the keywords, based on the word gravitation rule. If the poem is relevant, it will be added to the relevant database, otherwise it will be discarded.

5.2.3.2 Word Gravitation Implementation

As a core step in the proposed approach, the word gravitation rule plays an important role for candidate keyword selection. After calculating the attractive

forces between a keyword and its candidate words, the keyword and the top three candidate words will be used as input for the relevancy calculator.

Currently the crawler attempts to enrich this original list of keywords, by adding related candidate keywords which it has intelligently located during the crawling process [100]. In the poems downloaded by the focused crawler, the ones with a relevancy score are likely to be relevant to the search topic, in the case 'wind', since the range of relevancy score lies between 0 and 1, and the relevancy increments as this value rises. The table consists of related keywords with the highest frequency from each of these poems, with weight equivalent to the relevance score of the corresponding page.

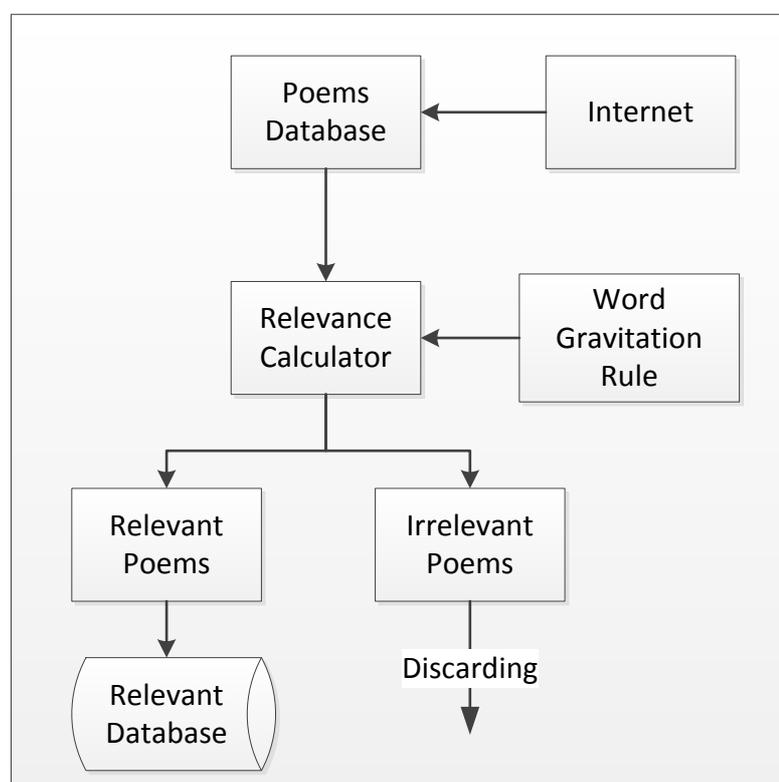


Figure 5.2 Architecture of Focused Crawling

5.2.3.3 Poems Relevance Calculator

The Poems Relevance Calculator computes the weight of words in a poem referring to the keyword in the table. Notice that the same words appearing at different locations of a poem contain information of varying significance. For example, the title text usually delivers more information about the topic of a poem than the body

text. As a result, the approach not only considers the number of candidate keywords appearing k_i , but also the location where it appears. The location weights $f(l, p)$ are adjusted as follow:

$$f(l, p) = \begin{cases} 2 & \text{poem title text} \\ 1 & \text{poem body text} \end{cases} \quad (5.2)$$

$f(l, p)$ is the weight of keyword k_i in different locations of a poem p . So we can get the overall weight $\sum_i f(l, p)$, namely wlp , of keyword k_i in poem p by adding the weights of k_i in different locations. Figure 5.3 shows an example.

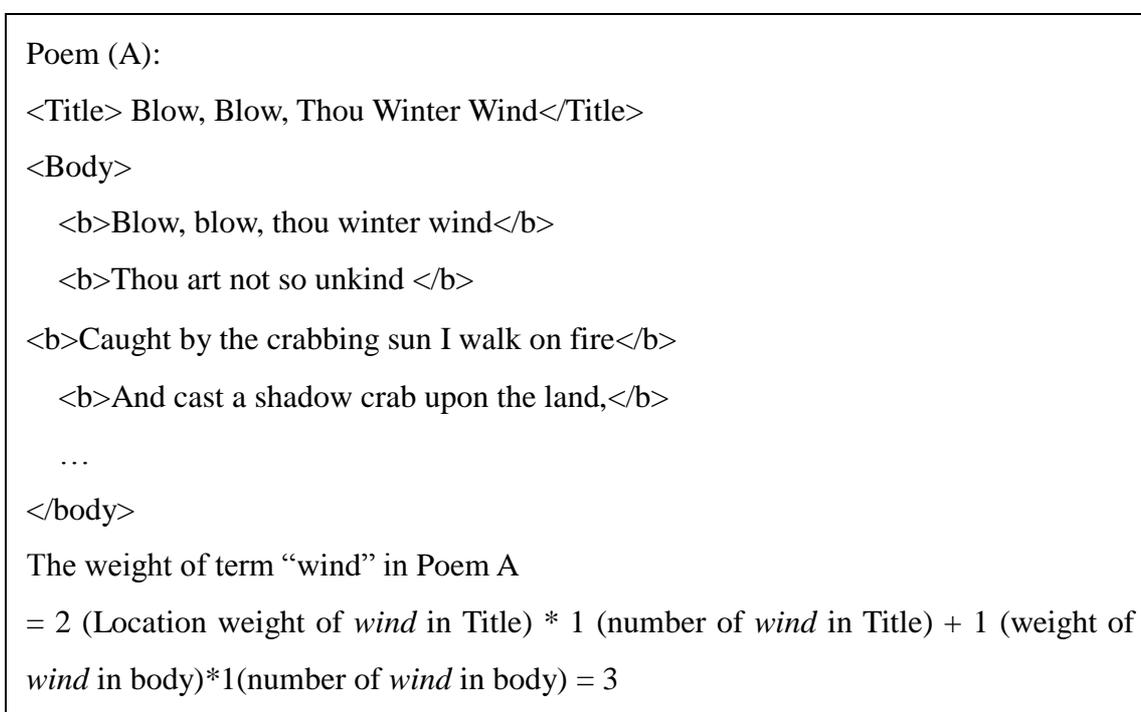


Figure 5.3 Poem Weighing Method

The pseudo code of the focused crawler algorithm is shown in Figure 5.4.

Firstly, all related poems are pre-downloaded and stored. Secondly, a list of poems is created, and the corresponding poems are parsed and extracted by using the keywords. Then the poem’s relevancy is calculated using equation (5.2). Finally, if the poem is relevant, it will be saved in the database; otherwise, it will be discarded.

```
1. if (poems is not empty)
2. {
4.   poem_list = extract_poems(Poems);
      // extract all the poems from the poem database
5.   calculate relevance of the poem using equation (15);
6.   if ( poem is relevant) {
7.     // put poem into relevant page database
8.     RelevantPoemDB = poem;
9.   }
10. }
11.}
```

Figure 5.4 Pseudo Code of the Focused Crawler Algorithm

5.3 Using Flotation Filter and Conductivity Calculator Rules for Poetry Data Pre-processing

This section introduces some common poetry data qualitative issues, such as noise. In order to solve these problems, this work presents a series of steps for poetry data pre-processing, based on designed creative Flotation Filter and Conductivity Calculator rules. The Flotation Filter aims to solve the poetry data of superficial dissimilarity problems, and the Conductivity Calculator highlights the poetry data with higher fluency and conductivity. The processed data will be stored in the database for poetry analysis and presentation.

5.3.1 Common Poetry Data Quality Issues

Currently practical databases have become astonishingly huge in size (several gigabytes or even more) which makes them highly susceptible to noisy data. Moreover, their possible origins from multiple, heterogeneous sources give rise to worse susceptibility. Low-quality data generally produces poor analysis results. Data pre-processing, thus, becomes an indispensable procedure to help enhance the quality of poetry data and effectiveness of poetry data analysis.

A great number of possible reasons may account for noisy data, for example, possessing incorrect attribute values. Maybe errors stem from the use of faulty data collection instruments. In addition, human or computer mistakes inevitably introduce errors when storing data. Data transmission can be susceptible to generate errors. There may be technology limitations, such as a limited buffer size, for coordinating synchronised data transfer and consumption. It is likely that inconsistencies in naming conventions or the data codes used are to blame. Also, inconsistent formats of input, such as *date*, can bring about incorrect data, and duplicate tuples may be due to lack of data cleaning.

5.3.2 Flotation Filter Rule (FFR)

With the help of the above word gravitation rule and web crawler, related poetry data about ‘*wind*’ has been collected. However, due to the huge amounts, in order to provide accurate data to users, pre-processing is necessary. Based on Archimedes’ buoyant force principle [40], this work proposes a Flotation Filter Rule.

The Flotation Filter Rule (FFR) is designed to reduce the scope and size of poetry data for more precise poetry data analysis. The scope and size of data can be quantified as the shape and volume of objects. Archimedes’ buoyant force principle, considers any object of arbitrary shape and volume V when surrounded by a liquid. The force the liquid exerts on an object within the liquid is equal to the weight of the liquid with a volume equal to that of the object. This force is applied in a direction opposite to gravitational force, which is of magnitude as shown in Figure 5.5.

And the force formula is as below:

$$F = \rho_f \cdot V \cdot g \quad (5.4)$$

Apparently, only if the magnitude of floating force is equivalent to its gravity would an object be able to stay static inside the liquid, where the equilibrium reads:

$$\rho_f V g - mg = 0 \quad (5.5)$$

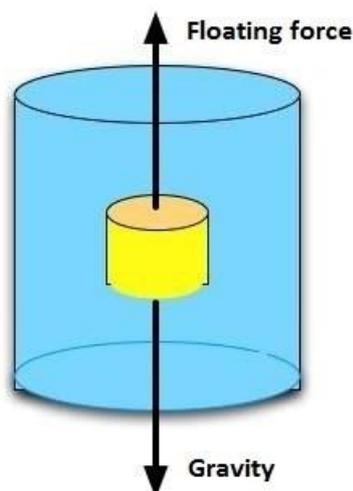


Figure 5.5 Archimedes' Buoyant Force Principle

Consequently, an object with a certain mass would float inside a liquid only with a specified volume. Based on this concept, the flotation filter rule is applied to data selection when a related *data 2* is expected to appear in similar length, i.e. volume, compared to *data 1*. Thus, the difference of volume can be defined as:

$$V_d = |V(\text{data1}) - V(\text{data2})| \quad (5.6)$$

Then all V_d derived from the candidates of retrieval should be rearranged in ascending order, from which users could request to filter out a specified percentage, e.g. 30% overlong poems and 30% too short ones, of results and return only the rest.

The mechanism of the flotation filter can be express as

$$\begin{cases} a \cong 30\% \times n \\ b \cong 70\% \times n \\ \{data_i \in \{filter(data)\}\} \end{cases}$$

$$for\ i = a, a + 1, a + 2 \dots \dots + b$$

where n denotes the total number of search results using the web crawler. Alternatively, the search results of volume difference greater than a threshold can be filtered out.

Let us take poems for instance. With the help of the above gravitation rule and web crawler implementation, 100 poems about 'wind' have been collected. In order to return more appropriate search results according to the user requirements, 30%, or any specified percentage, of overlong or too brief, candidate poems will be filtered

out. Alternatively, the filter may work relying on the threshold. If the threshold is specified as 10, the candidate poems which have over 10 or fewer than 10 words, compared to the average length of all results, will be filtered out. The rest will be stored for further processing.

5.3.3 Conductivity Calculator Rule (CCR)

After filtering by the flotation filter rule, the collected poetry data has been given an initial pre-processing. However, the flotation filter rule only considers the external properties of data, i.e., the shape and volume; the internal properties of data also need to be considered.

The Conductivity Calculator Rule is designed based on Ohm's law. It aims to do a final filter of the collected poetry data through their internal properties. Ohm's law states that the current through a conductor between two points is always equal to the voltage across them divided by the resistance, which can be represented as:

$$I = \frac{V}{R} \quad (5.7)$$

Inspired by the elegance of this equation, the idea was borrowed by defining: voltage as the type of context, such as a thesis, poems or novels; resistance as the compatibility of a word; and, thus, current as the fluency using the word. Given a certain type of literature, e.g., a poem, the voltage is constant 1. The resistance of a word is specified by the number of meanings available according to the dictionary, in which meaning according to different characteristics is allocated various values as adjectives are generally more compatible than nouns. The allocation table is as below:

Table 5.2 Word Classes Values

Noun	Verb	Adj.	Adv.	Pron.	Prep.	Conj.	Det.	Excl.
6	5	4	4	3	2	1	0.5	0.5

The resistance of a word is derived from a synthesis of parallel connected resistors, of which each resistance is equivalent to its parameter corresponding to the characteristic. Since the more characteristics and meanings a word has the more flexible it can be used in different contexts, the conductivity of a poem comprising

different words can be measured in which flexible words contribute to the overall fluency. Such fashion of quantification is comparable to the resistance calculation of parallel connection. In electronics, several parallel connected resistors can be regarded as one alternative resistor, whose value of resistance is equivalent to the reciprocal of the sum of reciprocals of those resistors in parallel, as shown in Figure 5.6 [43].

The equation of the Resistors Calculator is given below:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots + \frac{1}{R_n} \quad (5.8)$$

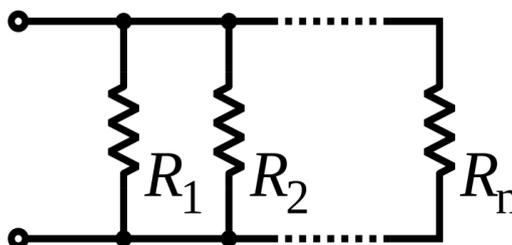


Figure 5.6 Resistors Connected in Parallel [43]

where n is the total number of interpretations for a word. For example, the word ‘content’ has three meanings in terms of being a noun, ‘everything that is included in a collection’, ‘what a communication that is about’, and ‘the proportion of a substance that is contained in a mixture or alloy, etc.’, and two as verb: ‘satisfy in a limited way’, and ‘make content’, and one as an adjective, ‘satisfied or showing satisfaction with things as they are’. However, one of its synonyms ‘satisfy’ possesses only three interpretations as a verb ‘fulfil the requirements or expectations of’; ‘make happy or satisfied’ and ‘fill or meet a want or need’. Therefore, their resistances can be obtained:

$$R('content') = 1 \div \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{5} + \frac{1}{5} + \frac{1}{4} \right) = 0.8695$$

$$R('satisfy') = 1 \div \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6} \right) = 2$$

Their fluencies in the context of a poem yield:

$$I('content') = \frac{1}{R('content')} = 1.15$$

$$I('satisfy') = \frac{1}{R('satisfy')} = 0.5$$

From the fluency point of view, '*content*' appears a more compatible word to use in poems than '*satisfy*'. The same rule can be applied to the synonym list of keyword retrieval such that words with a higher compatibility become preferable outcomes.

If the conductivity of collected poems needs to be calculated, the Conductivity Calculator Rule can also be applied. For instance, a keyword might occur multiple times in a poem. Thus, frequency of occurrence should definitely be highlighted. In order to evaluate the overall fluency of a poem, more frequently presented words are regarded as having a higher level of parallel connection, whose overall resistance is further reduced by the rate of occurrence.

If '*content*' occurs three times and '*satisfy*' occurs twice in a poem, the anticipated current for the poem yields

$$R = \frac{1}{\frac{3}{R('content')} + \frac{2}{R('satisfy')}} \\ I = \frac{1}{R} = \frac{3}{R('content')} + \frac{2}{R('satisfy')} = 4.45$$

The overall fluency of a poem is described as the conductivity of a circuit in electronics. By comparing and rearranging the conductivities of candidate poems, users are able to select the best matches in which keywords and their highly relevant synonyms are not only weighted by their compatibility of use but also considered by how frequently they appear.

5.4 Summary

In this chapter, a multi-purpose poetry data collection method has been illustrated, which can be reviewed as follows:

- It proposes the definition and aspects of poetry data relevance. The method quantifies the relevancy of poetry data depending on the law of universal gravitation, and derives the candidature database for the keyword. Moreover, it explains the creative word gravitation rule. The Word Gravitation Rule (WGR) is defined to calculate the attractive force, i.e. the gravity, between the keyword of user input and any word among the list of its synonyms. It provides more creative possibilities and diversity for poetry data collection.
- In order to implement the proposed poetry data collection method, an improved focused crawling that takes the proposed word gravitation rule as input (i.e., considers the candidate keywords as a search condition), is presented. The architecture and algorithm are also demonstrated in detail.
- This chapter introduces some common poetry data qualitative issues, such as noise. In order to solve these problems, this work presents a series of steps for poetry data pre-processing, including the Flotation Filter Rule and the Conductivity Calculator Rule. The Flotation Filter Rule is inspired by Archimedes' principle of buoyant force. This is used to solve the problem of data superficial dissimilarity. The Conductivity Calculator Rule highlights higher fluency and conductivity in the data, and is designed based on Ohm's law. The proposed pre-processing rules lead to providing more accurate and creative poetry data for users.

Chapter 6

A Multi-Dimension Poetry Data Analysis

Method

Objectives

- To present a creative multi-dimension poetry data analysis method
 - To propose multi-dimension rules designed for data sentiment analysis
 - To introduce the design principles and examples for the multi-dimension rules
-

6.1 Overview

Poetry data analysis is the core and key technology of poetry data manipulation. The value of data depends on the data analysis process. After poetry data collection and poetry data pre-processing, processed data are stored for poetry data analysis. Based on different demands, the processed data can be analysed selectively, or as a whole. Common data analysis technologies, such as data mining, machine learning, statistical analysis, etc., play an important role in data analysis. However, in the current poetry data era, the accuracy of algorithms is no longer the only benchmark. The meaning of the data has drawn people's attention. As a result, due to the new challenges and creative requirements, traditional poetry data analysis technologies need adjusting.

To provide users with experiences that accurately reflect their wants and needs in an authentic and creative manner, poetry data analysis needs to pay more attention to digging the meaning and sentiment out of data. This chapter introduces a multi-dimension approach, which considers the meanings of the data from three aspects: mood, theme and personal reflection. In the last chapter, the laws in physics

have been transformed and leveraged to implement creative poetry data collection, thanks to the processes whereby physical changes do not alter the properties of original words or texts. However, during data analysis, novel discoveries and explanations are expected to arise from the original data to generate creativity. Thus, characteristics of chemical changes are borrowed in pursuit of innovative and surprising analysis results.

This section describes the creative rules which aim to provide potential options and introduce various directions during the poetry data analysis. In order to explain the rules clearly, the design principle and examples are also presented.

6.2 Data Sentiment Analysis

Sentiment analysis refers to the processing, analysing and reasoning of subjective text material [58]. As the data related technologies thrive, people are not only concerned about data computing, but fascinated by uncovering the underlying sentiment that the data is trying to convey. Take poetry analysis for instance: normal poetry analysis always focuses on the following aspects [18]:

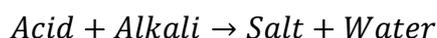
- Poem Published Date
- Poet Information
- Physical Analysis: Word Count, Lines Number and Stanzas
- Poem Topic
- Poem Theme
- Poem Mood
- Poem Vocabulary
- Poetic Devices: Rhyme, Meter and Imagery
- Poem Personal Reflection

Many researchers had worked on analysing the structure and poetic devices of poems [36, 37, 44]. However, how to analysis the sentiment part of poems, such as mood, or theme, are still challenging issues.

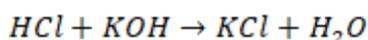
This chapter uses chemical principles for reference, creatively regarding poetry data as chemical elements, treating data combinations as chemical reactions in order to infer the sentiment contained in poetry data. This following section presents the creative multi-dimension poetry data analysis rules. They help to excavate the sentiment of poetry data based on three aspects - mood, theme and personal reflection - for satisfying the user requirements of enlightening and creativity.

6.3 Neutralisation Rule for Mood Analysis

The Neutralisation Rule (NR) is designed to analysis the mood of data based on chemical neutralisation reactions [86]. In the context of a chemical reaction, the term neutralisation is used to describe a reaction between an acid and a base or alkali. An empirical expression for this reaction is



For example:



In chemistry, *pH* is a numeric scale used to specify the acidity or basicity of an aqueous solution which is approximately the negative of base 10 logarithm of the molar concentration, measured in units of moles per litre, of hydrogen ions. Generally, the entire vocabulary can be categorised into positive, negative and neutral words in terms of mood [22]. The research aims to make use of this chemical concept to determine the mood of a context via pre-allocating *pH* for 2000 words of frequently-used vocabulary.

Assuming each word in a text is equivalent to one litre of solution with a specified pH value, the overall acidity or basicity of the mixture solution could be obtained. For instance, to calculate the moods of ‘gorgeous cherry blossom’ three portions of solutions are being mixed where neutralisation reaction takes place. The pH of these three components read

Table 6.1 *pH* Values of Gorgeous Cherry Blossom

Vocabulary	Gorgeous	Cherry	Blossom
<i>pH</i>	4	7	6

Thus, their acidity concentration, i.e. essentially the concentration of emotion, in one portion of solution yields:

$$H^+('gorgeous') = 10^{-4}$$

$$H^+('cherry') = 10^{-7}$$

$$H^+('blossom') = 10^{-6}$$

As for alkali solutions, the dominant ion is OH^- whose *pH* value is alternatively the complement to make up 14. In mixture calculations, the concentration of OH^- neutralises H^+ , so that OH^- should be marked negative. For example, the word 'tear' is assigned the *pH* value of 9:

$$OH^-('tear') = -10^{-(14-9)} = -10^{-5}$$

At the same time, the concentration of OH^- for neutral words, e.g. 'cherry', is 10^{-7} as well, the same as H^+ . Therefore, neutral words are assigned zero credit in the mixture calculation. In brief, the concentration assignments are listed below

$$Concentration(pH) = \begin{cases} 10^{-pH}, & pH < 7 \\ 0, & pH = 7 \\ -10^{-(14-pH)}, & pH > 7 \end{cases} \quad (6.1)$$

Literally, 'gorgeous' and 'blossom' are positive words of different degree, whereas, 'cherry' is a neutral word that acts as diluent to neutralise the overall *pH* towards 7, i.e. neutralisation. The *pH* of overall solution can be derived from

$$pH('gorgeous\ cherry\ blossom') = -\log_{10} \left(\frac{10^{-4} + 0 + 10^{-6}}{3} \right) = 4.4728$$

Hence, the overall *pH* of the phrase equals 4.4728, which implies a positive mood. Notice that if the majority of words in a text possess extreme *pH* less than 7 and occur frequently, the mood will be more likely positive, and vice versa. On the other hand, a great number of neutral words would reduce the concentration of the solution, thus, neutralising the mood. By analysing the *pH* of a poem, novel or paper, an innovative acidity measurement is expected to convey the mood for the users.

6.4 Organic Substance Rule for Theme Analysis

Besides the mood of a poem, what it is mainly talking about, i.e. the theme, is certainly of users' interest. The synthesis of words are believed that comprise the text is able to reveal the main theme of a piece of literature based on an accumulation of contribution from each word. The **Organic Substance Rule for Theme Analysis (OSR)** is inspired by the concept of organic compounds in chemistry [65].

As organic substances are constructed from carbon based atoms, an organic molecule is simply an accumulation of the number of elements it is comprised of. A typical glucose molecule is shown in Figure 6.1 [94].

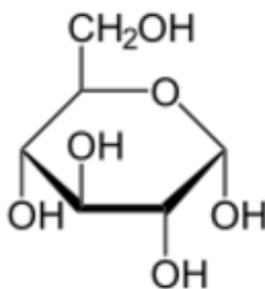


Figure 6.1 Glucose Molecules [94]

It consists of six carbon atoms, twelve hydrogen atoms and six oxygen atoms such that could be written as $C_6H_{12}O_6$. Common organic elements include *carbon (C)*, *oxygen (O)*, *nitrogen (N)*, *sulphur (S)*, *hydrogen (H)*, *phosphorus (P)*, etc. Here the research sets six genres of themes in poetry: telling ambition, expressing emotion, criticising, commemorating, praising and celebrating. They can be assigned to the six common elements correspondingly, as shown in the Table 6.2.

Table 6.2 Six Themes for Poems and Their Organic Elements

Telling ambition	C
Expressing emotion	O
Criticising	N
Commemorating	S
Praising	H

Celebrating	P
-------------	---

Given pre-allocated theme credits for 2000 words of frequently used vocabulary, the overall theme can be estimated by accumulating the credit of each word in the piece of literature. The most significant count of an element will be recognised as the main theme. Take one sentence, for example: ‘*Rough winds do shake the darling buds of a date*’, in which none zero credit words include

$$\text{Organic}('rough') = (0, 2, 3, 0, 0, 0)$$

$$\text{Organic}('winds') = (1, 3, 0, 1, 0, 0)$$

$$\text{Organic}('shake') = (1, 2, 2, 0, 0, 0)$$

$$\text{Organic}('darling') = (2, 5, 0, 3, 1, 2)$$

$$\text{Organic}('date') = (0, 1, 0, 2, 0, 1)$$

And the elements count is shown in Table 6.3.

Table 6.3 Elements Count of ‘Rough Winds do Shake the Darling Buds of a Date’

Elements	C	O	N	S	H	P
Count	4	13	5	6	1	3

Therefore, the molecule expression for this organic compound is, as can be shown in the histogram below

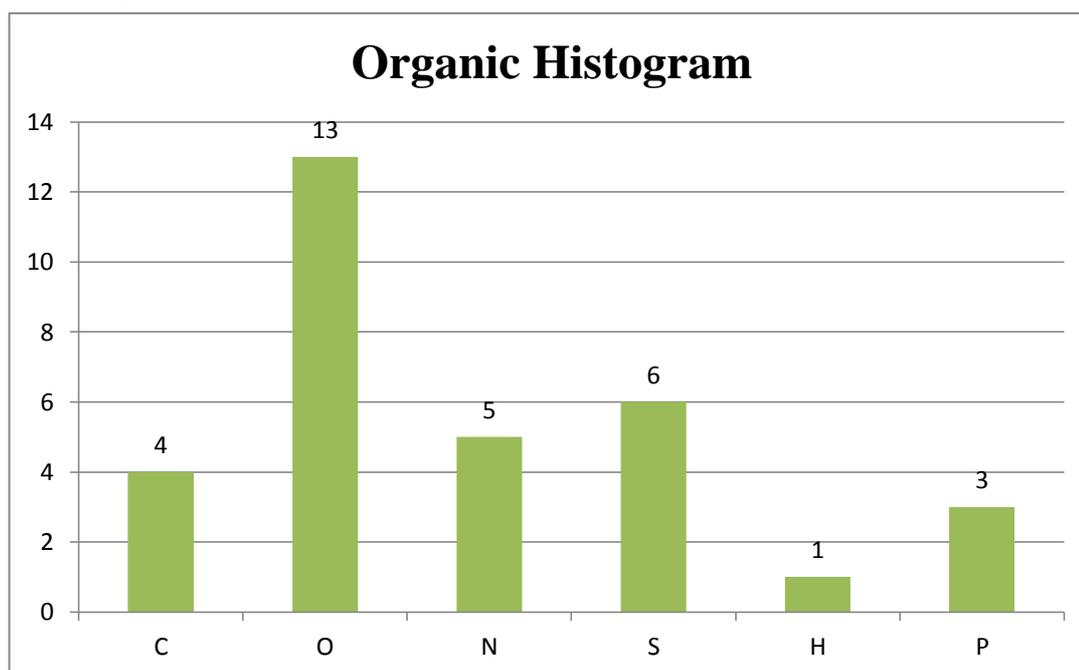


Figure 6.2 Organic Histogram for ‘Rough Winds Do Shake the Darling buds of a Date’

Obviously, the *oxygen* element is the dominating component in this compound, i.e. expressing emotion mostly accounts for the theme of this sentence. Thus, it is concluded that the sentence probably aims at expressing the author’s emotions.

The above shows the analysis result applying OSR to a sentence. If a poem is to be analysed, all the words in all the sentences should be disassembled and their individual credit accumulated to determine the theme of the poem. Certainly, the words comprising the title may be assigned greater weight to highlight their overall contribution.

6.5 Reduction-Oxidation Rule for Personal Reflection Analysis

Apart from estimating the theme, personal reflection on browsing the text also draws the interest. Personal reflection refers to the user’s intuition, feelings and thoughts after reading the poem. Although it is subjective, a personal reflection should always be related to the theme that the author wanted to deliver. Given the mood of a poem, the message originating with the author arrives at the reader’s heart differently. In this proposal, this reflection with reference would like to be analysed and quantified to a classical chemical reaction: the reduction-oxidation reaction [55].

The reduction-oxidation reaction is a chemical reaction in which the oxidation states of atoms are changed. Any such reaction involves both a reduction process and a complementary oxidation process, two key concepts involved in the electron transfer processes [55]. The processes of reduction and oxidation are illustrated in Figure 6.3 below

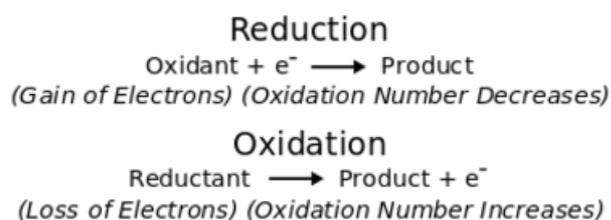
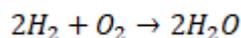
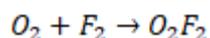


Figure 6.3 The Processes of Reduction and Oxidation [55]

Generally speaking, a substance might transform into other substances depending on whether it is gaining or losing electrons. Dioxygen is a common oxidising agent in a chemical reaction, e.g. the burning of hydrogen gas:



Wherein dioxygen difluoride can be obtained by subjecting a 1:1 mixture of gaseous fluorine and oxygen at low pressure (7–17 mmHg (0.9–2.3 kPa) is optimal) to an electric discharge of 25–30 mA at 2.1–2.4 kV [70], in which dioxygen is acting as the reducing agent:



The Reduction-Oxidation Rule for Personal Reflection Analysis (ROR) is the idea that as a substance transforms differently under specific conditions, it triggers the development of the personal reflection analyser. Based on the six categories of themes of a poem, personal reflection varies as the measure of mood differs. In other words, the *pH* of a text plays the role of either an oxidising agent or reducing agent, which leads to personal reflection. Here, oxidation is regarded as a positive reflection, whereas, reduction is treated as negative feedback. After a piece of literature, of which the formula reads:

$$Theme \leftrightarrow \begin{cases} \text{Personal reflection1, } pH \leq 7 \\ \text{Personal reflection2, } pH > 7 \\ \dots \dots \end{cases} \quad (6.2)$$

the lookup table is defined as below:

Table 6.4 Themes of Poems and their Oxidation and Reduction

Theme	Oxidation	Reduction
<i>Telling ambition</i>	Confident	Sense of lost
<i>Expressing emotion</i>	Cheerful	Miserable
<i>Criticising</i>	Irony	Hate
<i>Commemorating</i>	Yearning	Regret
<i>Praising</i>	Passionate	Encouraging

<i>Celebrating</i>	Gratitude	Delight
--------------------	-----------	---------

With respect to the above six themes, the oxidising agent and the reduction agent denote the two emotional orientations for each theme. Some orientations refer to opposite expressions of feeling while others are merely expressing a feeling to different degrees.

- *Telling ambition*: If it is positive, confidence is being expressed; otherwise, there is a sense of lost.
- *Expressing emotion*: Positive emotion suggests cheerfulness; otherwise, it conveys sadness.
- *Criticising*: If positive criticism is treated as irony; negative criticism becomes hate.
- *Commemorating*: Positive commemorating is recognised as missing, whereas, it could be considered regret in a negative sense.
- *Praising*: Positive praising implies intense devotion. On the other hand, it means encouraging in a mild way.
- *Celebrating*: An intense celebration is regarded as expressing gratitude, while it could be just delight if enjoyed gently.

Regarding the different reflections, the analysis returns a sentence which stems from a similar emotion to inspire creative minds.

Table 6.5 Themes and Their Analysis Results

Theme	Reaction	Analysis Results
Telling Ambition	Oxidation	Confident people believe in themselves, and because they believe, they achieve.
	Reduction	It'll be impossible to make up the lost time.
Expressing Emotion	Oxidation	A red dress brings warmth and cheer to the saddest winter day
	Reduction	There was nothing at all in this miserable place to distract me.
Criticising	Oxidation	The author's tone was tinged with irony.

	Reduction	It is never hard to imagine that when the poet wrote down these words, his eyes were full of hate.
Commemorating	Oxidation	The trees come up to my window like the yearning voice of the dumb earth.
	Reduction	'Don't do anything you might regret.' said the author.
Praising	Oxidation	The enthusiasm expressed from the text overflows so much that it almost grows into an obsession.
	Reduction	The encouraging poem was like a shot in the arm.
Celebrating	Oxidation	'Thank you!' She put her palms together devoutly, with pricking tears.
	Reduction	Her screams of delighted laughter filled the air.

Certainly, detailed categorisation is applicable as the value of pH varies. For ease of demonstration, only the acidity or basicity is being discussed in terms of personal reflection. Apply this rule to the previous example, '*Rough Winds Do Shake the Darling buds of a Date*'. The acidity of this sentence gives us:

$$pH = -\log_{10} \left(\frac{-10^{-(14-8)} + 0 + 0 - 10^{-(14-8)} + 0 + 10^{-4} + 0 + 0 + 0 + 10^{-6}}{10} \right)$$

$$= 5.0044$$

The pH of this sentence is less than 7, which indicates an oxidation reaction is taking place. The theme of the sentence was designated expressing emotion in the previous example. Hence, the estimated personal reflection is the oxidation state of expressing emotion: 'Cheerful'. In this way, a creative analysis results: '*A red dress brings warmth and cheer to the saddest winter day*' is presented to the user based on the synthesis of the mood and theme of the line.

6.6 Summary

In this chapter, a creative multi-dimension poetry data analysis method is explained, which can be summarised as follows.

- It proposes the Neutralisation Rule (NR) that is designed to analyse the mood of poetry data based on chemical neutralisation reactions. Based on NR each word in a text is equivalent to one litre of solution with a specified pH value, so that to obtain the overall acidity or basicity of the mixture solution. According to the final pH value, the mood of poetry data can be classified into positive, negative and neutral. By analysing the pH of a poem, novel or paper, an innovative acidity measure is expected to convey the mood for the users.
- In order to analyse the theme of poetry data, the Organic Substance Rule for Theme Analysis (OSR) is inspired by the concept of organic compounds in chemistry. It considers six themes of poetry as six chemical elements. The overall theme can be estimated by accumulating the credit of each word in a poem, with most significant count of an element being recognised as the main theme.
- The Reduction-Oxidation Rule for Personal Reflection Analysis (ROR) is the idea that a substance transforms differently under specific conditions, triggering the development of the personal reflection analyser. Based on the six categories of themes of a poem, the personal reflection varies as the measure of mood differs. In other words, the pH of a text plays the role of an oxidising agent or reducing agent, which leads to the personal reflection. With regards to the different reflections, the analysis returns a sentence which stems from a similar emotion to inspire user creative minds.

Chapter 7

A Multi-Media Poetry Data Presentation

Method

Objectives

- To present a creative multi-media poetry data presentation method
 - To introduce designing creative rules and principles for the mood, theme and personal reflection data presentation
 - To show the presentation examples and curves
-

7.1 Overview

Poetry data presentation is the display and techniques used in the demonstration of poetry data manipulation. It is usually what users care about. If the poetry analysis result is not appropriately interpreted, the result would be incomprehensible, and even misleading. There are multiple ways of presenting data, such as: tables, graphs, stem-and-leaf plots and charts, etc. These methods may be good choices for dealing with small amounts of data. However, the type of poetry data varies, and the correlations among the results are complex. Traditional poetry presentation methods are not working any longer, which leads to the need for more approaches to enhance poetry data presentation.

In the current data era, multi-media data such as audio, video, image and animation have gradually become the mainstream. People have started to pay attention to the means of delivering the most valuable information to users via data visualisation methods.

To provide users with experiences that accurately reflect their wants and needs in an aesthetic and creative manner, data presentation needs to pay more attention to diversity, especially regarding visualisation. In previous sections, creative poetry data collection and poetry data analysis techniques have been proposed and used to obtain the mood, and theme of, as well as personal reflections to a poem. This chapter aims to present these analysis results in innovative ways. This chapter introduces a multi-media poetry data presentation approach which firstly builds the connection between geometric shapes in mathematics and the emotional data. The proposed approach not only considers how to present poetry data, but also uses creative rules to enhance the data visualisation. By introducing creative presentation rules, an emotion based multi-media poetry data presentation method is provided, and examples are used to demonstrate how the proposed method works.

7.2 Mood Presentation Rule

7.2.1 Mood and Emoticons

An emoticon is a pictorial representation of a facial expression using punctuation marks, numbers and letters [39]. It usually is written to express a person's feelings or mood. The most famous example is that which Ambrose Bierce [6] proposed as an improvement in punctuation- the snigger point, or note of cachinnation. It is written thus: “☺”, and represents a smiling mouth.

Based on this method, this chapter combines emoticons and parabola curves in mathematics to propose a creative mood presentation rule. It aims to present the mood of a poem immediately and creatively.

7.2.2 Rule Definition

The Mood Presentation Rule (MPR) is designed to present the mood of data based on parabola curves in mathematics. Owing to the briefness and elegance of geometric shapes in mathematics, presenting the poetry data along a certain curved trajectory will help convey the emotional content of poetry data intuitively. The two-dimensional expression can be written:

$$y = kx^2 \quad (7.1)$$

The sign and magnitude of k determines the trend of the parabola. Because the mood measure has been defined, positive and larger values to k to represent a positive mood of the poem can be allocated, and use negative k , i.e. a sad face, to express a negative mood.

Since $pH=7$ is specified to stand for an indifferent mood, it makes sense to assign

$$k = 7 - pH \quad (7.2)$$

Such that positive, known as acid, text will generate a smiling face and vice versa.

7.2.3 An Example for MPR

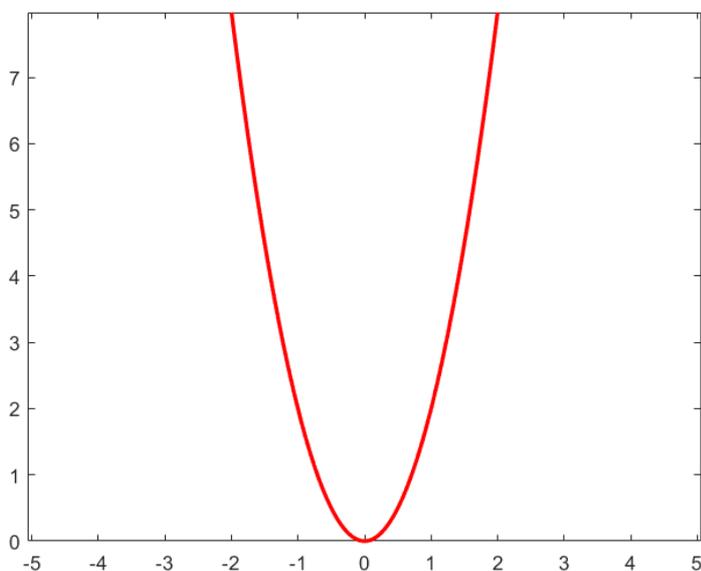


Figure 7.1 Mood Presentation of ‘Rough Winds do Shake the Darling Buds of a Date’

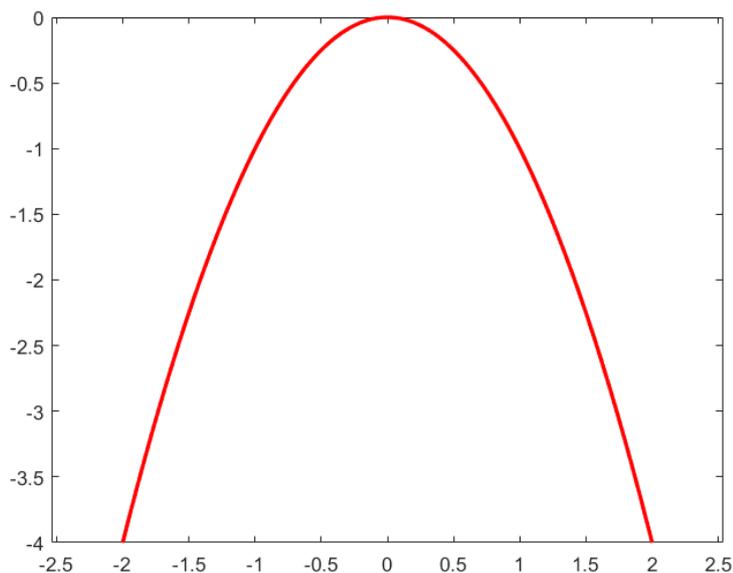


Figure 7.2 Mood Curve of the Line with pH=8

In this case, the pH value of the above line ‘*Rough winds do shake the darling buds of a date*’ has been calculated to 5.004. According to rule (7.2), the k of this sentence can be set as

$$k_{rough\ winds} = 7 - 5.004 = 1.996$$

So this poem belongs to a positive mood, and a plot would be shown as in Figure 7.1.

On the contrary, if there is another line with pH=8, its k value can be calculated as:

$$k_{example2} = 7 - 8 = -1$$

Because -1 is a negative value, so the curve of this poem will look like Figure 7.2.

7.3 Theme Presentation Rule

7.3.1 Colour and Meaning of Data

Colour is a characteristic of human visual perception described through colour categories, with names such as red, yellow, purple or bronze [50]. It plays an important role in daily life, and has the ability to affect human emotions and moods.

Table 7.1 Colours and Related Meanings [34]

Colour	Meanings			
White	Mourning(In some cultures)	Purity	Innocence	Neutrality
Black	Evil	Death/Mourning	Authority	Power
Red	Passion	Joy	Sensitivity	Anger
Orange	Happy	Warmth	Energetic	Excitement
Yellow	Hunger	Laughter	Intensity	Cheery
Green	Natural	Health	Envy	Tranquillity
Blue	Cold	Melancholy	Wisdom	Calmness
Violet	Royalty	Mystery	Exotic	Spiritual
Pink	Romance	Love	Agitation	Gentle
Brown	Security	Reliability	Stability	Friendship
Indigo	Devotion	Justice	Impartiality	Faithful

For example, warm colours – such as red, yellow and orange – can spark a variety of emotions ranging from comfort and warmth to hostility and anger. Cool colours – such as green, blue and purple – often spark feelings of calmness as well as sadness [10]. Colour psychology is the study of the hues as a determinant of human behaviour [7]. The psychology of colour is based on the mental and emotional effects colours have on sighted people in all facets of life. Different colours can represent various emotions: red usually means passion, blue presents melancholy, green stands for tranquillity, etc. The approach introduces this idea to help to present the theme of data.

In order to present data creatively, this thesis proposes a multi-media data presentation method, which utilises different colours to present the meaning of data. In this thesis, different colours are used to present different meanings, as shown in Table 7.1 [34].

7.3.2 Theme Presentation Rule

After using a shape to present the mood of a poem, a spectrum of colours is used to derive from an organic histogram analyser, to show the theme. One of the most recognisable spectrums is that of a rainbow, whose seven prime colours are red, orange, yellow, green, blue, indigo and violet [81]. Since there are six categories of themes, and orange is not as distinguishable as the other colours, only six of the prime rainbow colours are adopted to express the theme.

The rainbow colours are aligned in a specific sequence such that it is necessary to reallocate the themes according to whether they are of cool or warm tone. One possibility of allocation is listed in Table 7.2 below (certainly, a user is welcomed to assign colours as they wish).

The third column lists the corresponding R, G, B values of these colours in Matlab, which will be used for plotting. According to the histogram Figure 6.2 in the last chapter, the proportion that each theme accounts for is specified.

Table 7.2 Six Colours for Themes

Theme	Colour	[R G B]	Proportion
Celebrating	Red	[1 0 0]	3/32
Praising	Yellow	[1 1 0]	4/32
Commemorating	Green	[0 1 0]	6/32
Criticising	Indigo	[0 1 1]	5/32
Telling ambition	Blue	[0 0 1]	13/32
Expressing emotion	Violet	[1 0 1]	1/32

7.3.3 An Example for TPR

The rainbow plot is shown in Figure 7.3. As can be seen, violet, which refers to Expressing emotion, takes up the largest part, whereas, warm tones, like red and yellow, i.e. Celebrating and Praising, are less significant in the example: *'Rough winds do shake the darling buds of a date'*.

This rainbow proposal is merely showing one of our ideas in presenting data. The rainbow can be applied as a background, or accessory, or upgraded to a 3D effect picture to aid the presentation of theme. Other patterns, spectrum components or shapes of rainbows are also applicable, as long as they help vividly express the plain text.

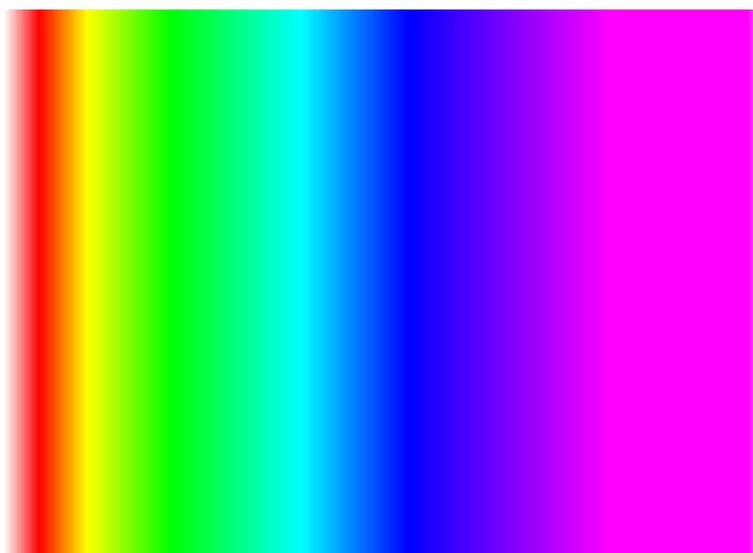


Figure 7.3 Rainbow Plot of ‘Rough Winds do Shake the Darling Buds of a Date’

7.4 Personal Reflection Presentation Rule

7.4.1 Rule Definition

Now that 12 kinds of personal reflections have come up, 12 types of curves or shapes are expected to display these reflections. In geometry, curves can be generally divided into finite or infinite in terms of range. Therefore, finite curves or shapes which are usually more conserved, or even depressed, can be used to show relatively negative reflections with pre-defined pH values greater than 7. In contrast, infinite curves often end up with a remarkable uptrend, which could convey positive feedback. A possibility of curves or shapes allocations is shown in the table below

Table 7.3 Twelve Kinds of Personal Reflections and Their Curves

Oxidation	Curve or Shape	Reduction	Curve or Shape
Confident	<i>‘Victory’</i>	Sense of Lost	<i>‘Helix’</i>

Cheerful	<i>'Rising'</i>	Miserable	<i>'Descending'</i>
Ironic	<i>'Sinusoidal'</i>	Hate	<i>'Downwards Triangle'</i>
Yearning	<i>'Hyperbola'</i>	Regretful	<i>'Rhombus'</i>
Passionate	<i>'Exponential'</i>	Encouraging	<i>'Square'</i>
Grateful	<i>'Logarithmic'</i>	Delighted	<i>'Circle'</i>

7.4.2 Twelve Kinds of Curves for Personal Reflection

The mathematics expressions for the 12 curves or shapes above are listed as follow:

$$1. \quad f(\text{'Victory'}) = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (7.4)$$

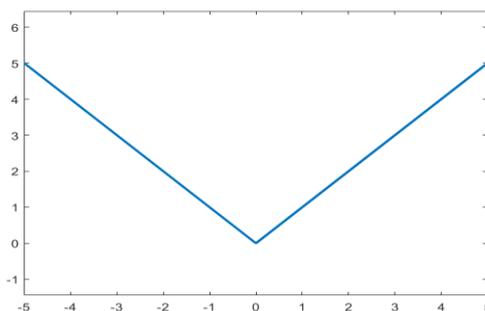


Figure 7.4 'Victory' for Confident

$$2. \quad f(\text{'Rising'}) = x \quad (7.5)$$

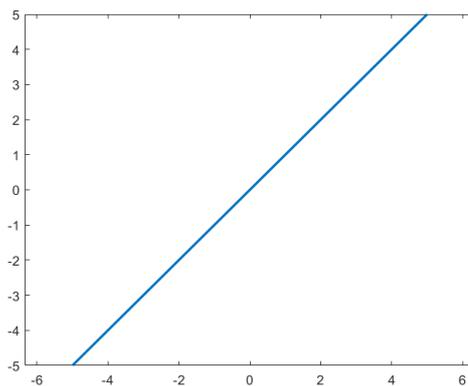


Figure 7.5 'Rising' for Cheerful

$$3. \quad f(\text{'Sinusoidal'}) = \sin(x) \quad (7.6)$$

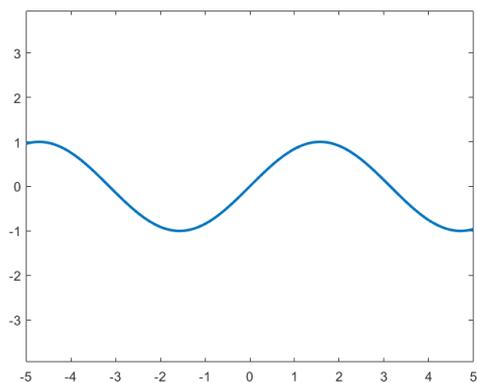


Figure 7.6 ‘Sinusoidal’ for Ironic

4. $f(\text{'Hyperbola'}) = \frac{1}{x} \quad x \neq 0$ (7.7)

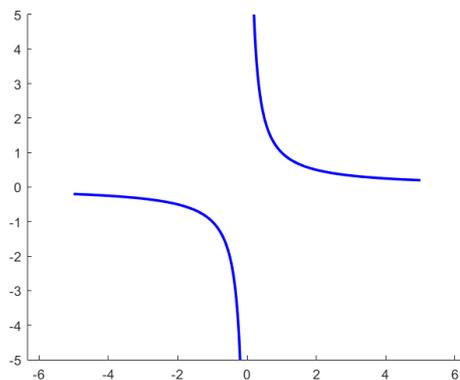


Figure 7.7 ‘Hyperbola’ for Yearning

5. $f(\text{'Exponential'}) = 2^x$ (7.8)

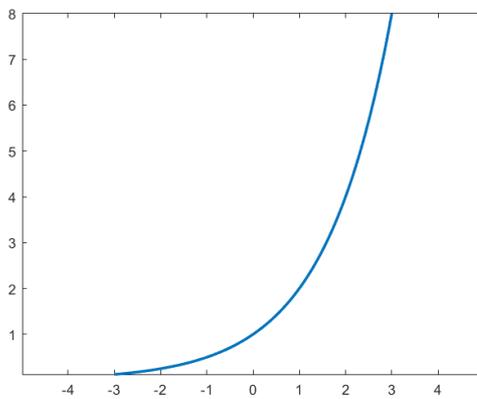


Figure 7.8 ‘Exponential’ for Passionate

$$6. \quad f(\text{'Logarithmic'}) = \ln(x) \quad x > 0 \quad (7.9)$$

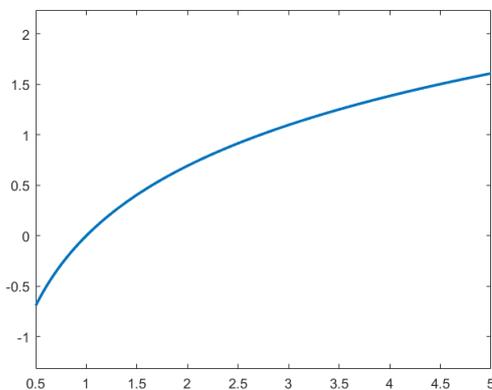


Figure 7.9 'Logarithmic' for Grateful

$$7. \quad \text{'Helix'} \rightarrow \begin{cases} x(t) = \cos(t) \\ y(t) = \sin(t) \\ z(t) = kt \end{cases} \quad (7.10)$$

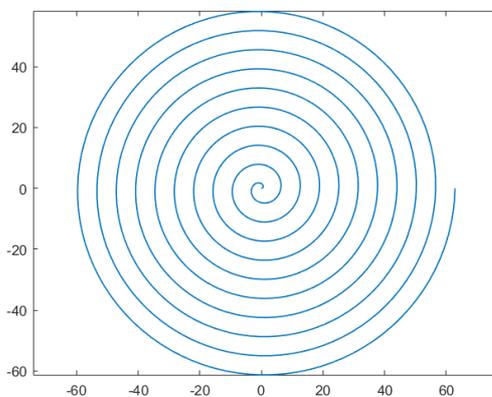


Figure 7.10 'Helix' for Sense of Lost

$$8. \quad f(\text{'descending'}) = -x \quad (7.11)$$

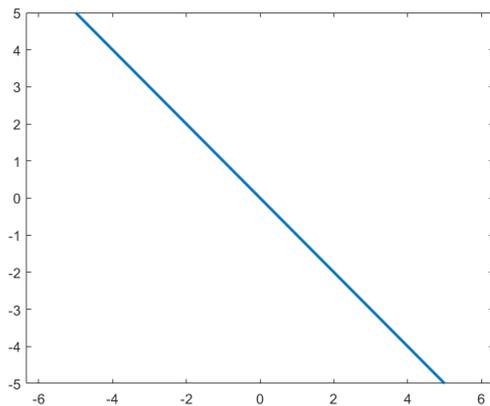


Figure 7.11 ‘Descending’ for Miserable

9. 'Downwards triangle' $\rightarrow \{(0,0), (1,2), (-1,2)\}$ (7.12)

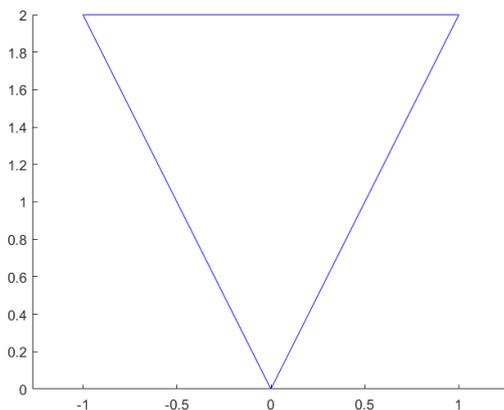


Figure 7.12 ‘Downwards Triangle’ for Hate

10. 'Rhombus' $\rightarrow \{(0,0), (1,2), (0,4), (-1,2)\}$ (7.13)

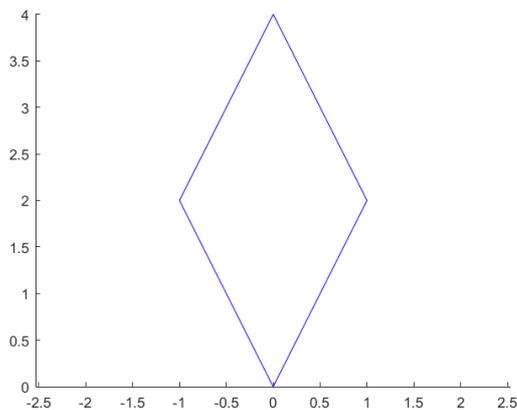


Figure 7.13 'Rhombus' for Regretful

$$11. \text{'Square'} \rightarrow \{(0,0), (2,0), (2,2), (0,2)\} \quad (7.14)$$

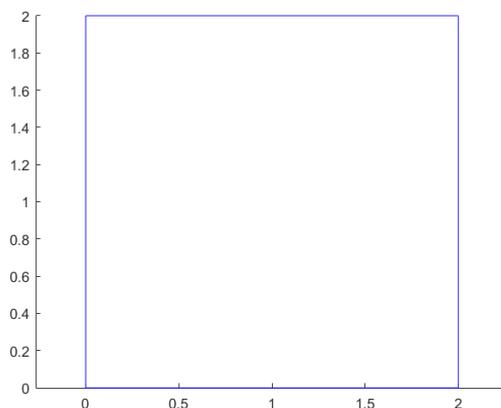


Figure 7.14 'Square' for Encouraging

$$12. \text{'Circle'} \rightarrow x^2 + y^2 = 1 \quad (7.15)$$

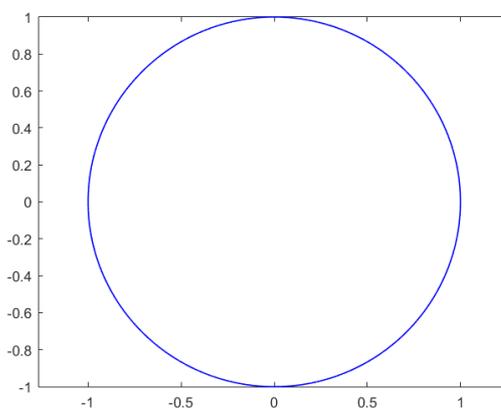


Figure 7.15 'Circle' for Delighted

7.4.3 An Example for PRPR

Since the example, '*Rough winds do shake the darling buds of a date*', is categorised as the personal reflection of Cheerful, the presentation of this line should be along a rising line. An effect picture presenting mood, theme and personal reflection may adopt a rising rainbow as background, and display the rolling scripts

along a 'smiling face' parabola. Users are always welcome to access and customise patterns of presentation.

7.5 Summary

In this chapter, a creative multi-media poetry data presentation method is explained, which can be summarised in the following aspects.

- It proposes a creative mood presentation rule which combines emoticons and parabola curves in mathematics. Through the parabola curve, it presents a positive mood as a smiley face and a negative mood as a sad face. It aims to present the mood of poems immediately and creatively.
- Based on the principles and underlying emotions of colours, the theme presentation takes advantage of the spectrum of the rainbow, which delivers the theme to users vividly.
- 12 personal reflections are presented in various curves and shapes to offer users both innovative and interesting expression.

Chapter 8

Case Studies

Objectives

- To illustrate the proposed creative poetry data manipulation approach through a Poetrypedia system
 - To present the design details and user interfaces of Poetrypedia
 - To illustrate the functions of the Exploration module in Poetrypedia
 - To show the functions of the Appreciation module in Poetrypedia
 - To specific the functions of the Help-to-Write module in Poetrypedia
 - To present the functions of the Dissemination module in Poetrypedia
-

8.1 Overview

Nowadays, thanks to the all-pervading accessibility and rich resources from the Internet, information, websites and applications relating to poetry are easy to find. However, these resources are usually dispersed, many are only concerned with content or publications, and they lack imagination.

In order to provide a comprehensive and creative system for poetry analysis and appreciation, this chapter presents how the proposed creative approach is implemented in the Poetrypedia system. Poetrypedia is a comprehensive multi-function poetry system, which not only helps users explore a wide range of information about poems, but assist in writing, comprehension and composition. Four main modules of Poetrypedia, - Exploration, Appreciation, Help-to-Write and Dissemination- are specifically proposed to examine the approach as a set of case studies. The design details, user interfaces and workflows are presented respectively.

8.2 Design Details and User Interfaces of Poetrypedia

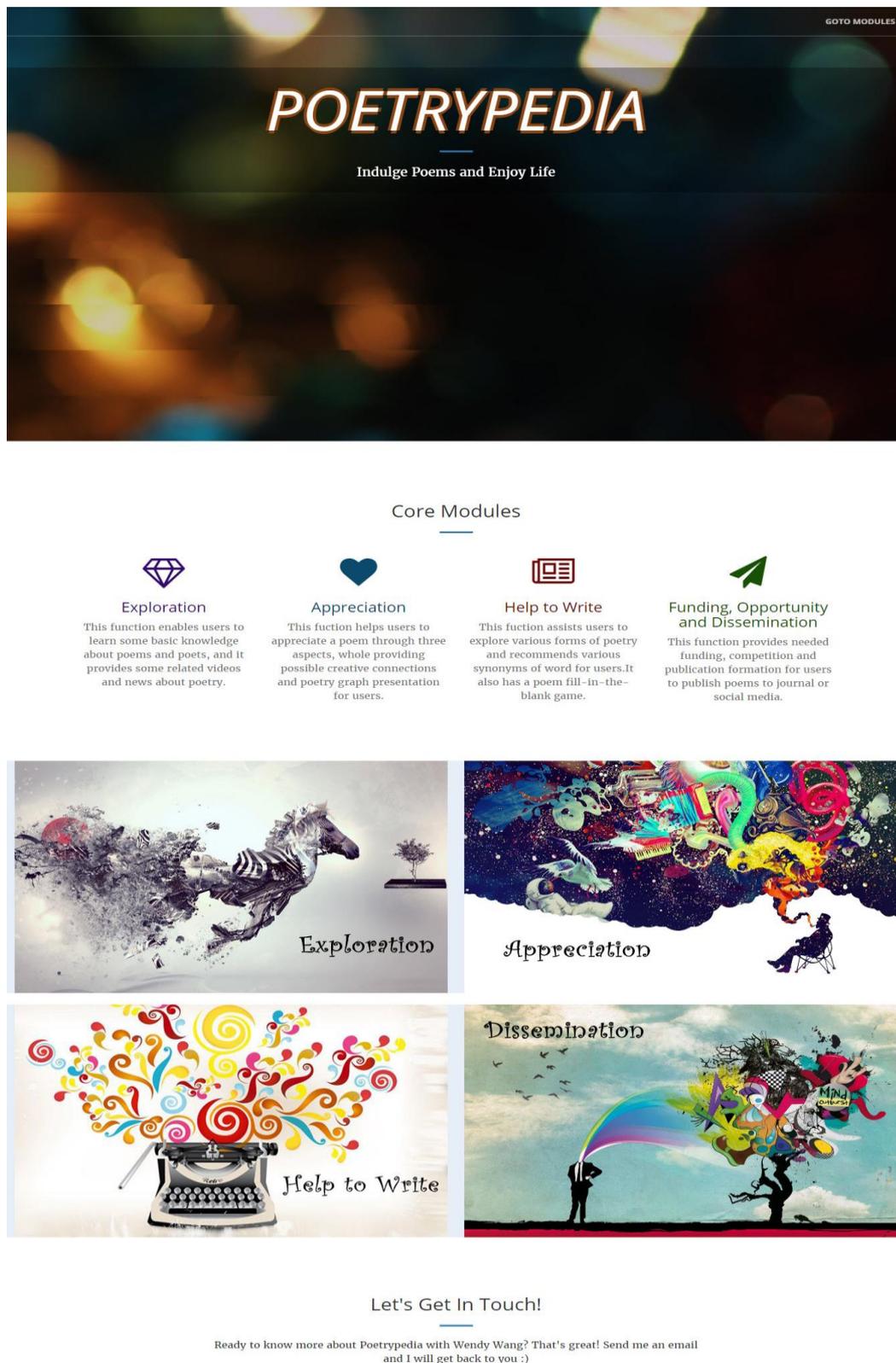
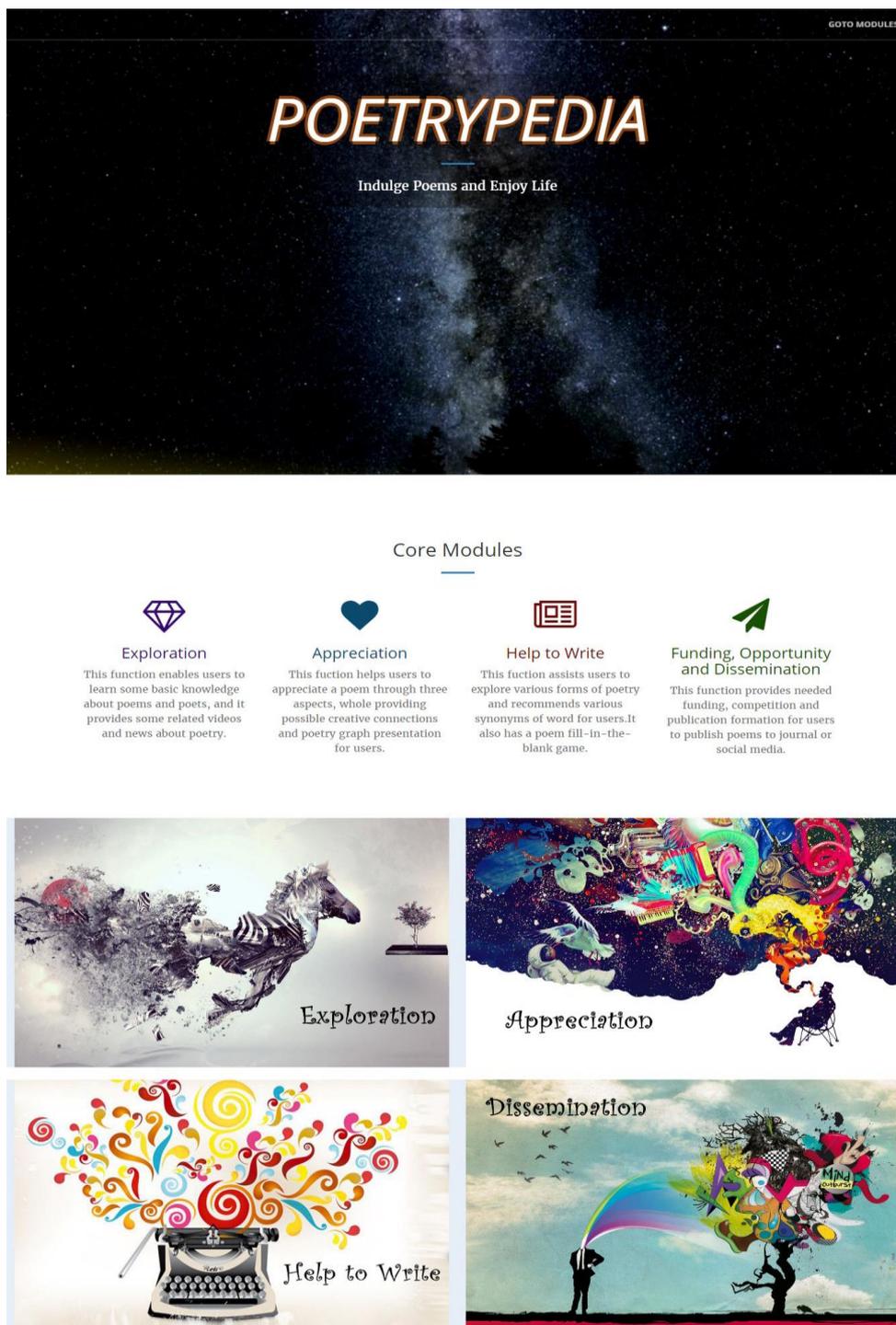


Figure 8.1 User Interface of Poetrypedia



Let's Get In Touch!

Ready to know more about Poetrypedia with Wendy Wang? That's great! Send me an email and I will get back to you :)



wendywang8822@hotmail.co.uk

Figure 8.2 Second User Interface of Poetrypedia

Poetrypedia is a human-computer interaction system for poetry analysis and appreciation. There are four main modules in Poetrypedia: Exploration, Appreciation Help-to-Write, and Dissemination. The user interface is shown in Figure 8.1.1. There are six different dynamic videos displaying randomly when the user scrolls the cursor. The starry night is one of them, as shown in Figure 8.1.2.

The Exploration module is designed for users to explore multiple aspects of poetry, including poems and poets, related news and videos etc. The design details are shown in Figure 8.3. The Basic Knowledge for Poems module presents some basic knowledge about poetry, such as the structure and content of poems. The user is able to explore different kinds of poems according to his requirements. On the other hand, the Basic Information for Poets module presents information on poets, such as life stories, background, representative works etc. If the user finds anything interesting, he can directly click on a link to enter the Audio & Video module to look into it. The News & Events module provides the latest news and events on poetry. The user interface of the Exploration module is shown in Figure 8.4.

The Appreciation module is designed to help users to appreciate a poem from three aspects; the user can input a poem and get the results of an analysis based on comments by readers and other selected sources; moreover, it provides some possible creative connections and poetry graph presentation for users. The design details of the Appreciation module are shown in Figure 8.5. The Appreciation module not only assists users to understand the structure and meaning of poems, but also aims to inspire users through creative analysis results. There are four functions in this module: Comments by Readers, Analysis by Aspects, Possible Creative Connections and Poetry Presentation. Comments by Readers presents related comments on the poem from user input, which helps the reader comprehend the poem. The other three functions are designed to inspire creative views of appreciation for the user based on the creative rules proposed by the thesis Section 8.3 will expand on these functions. The user interface of the Appreciation module is shown in Figure 8.6

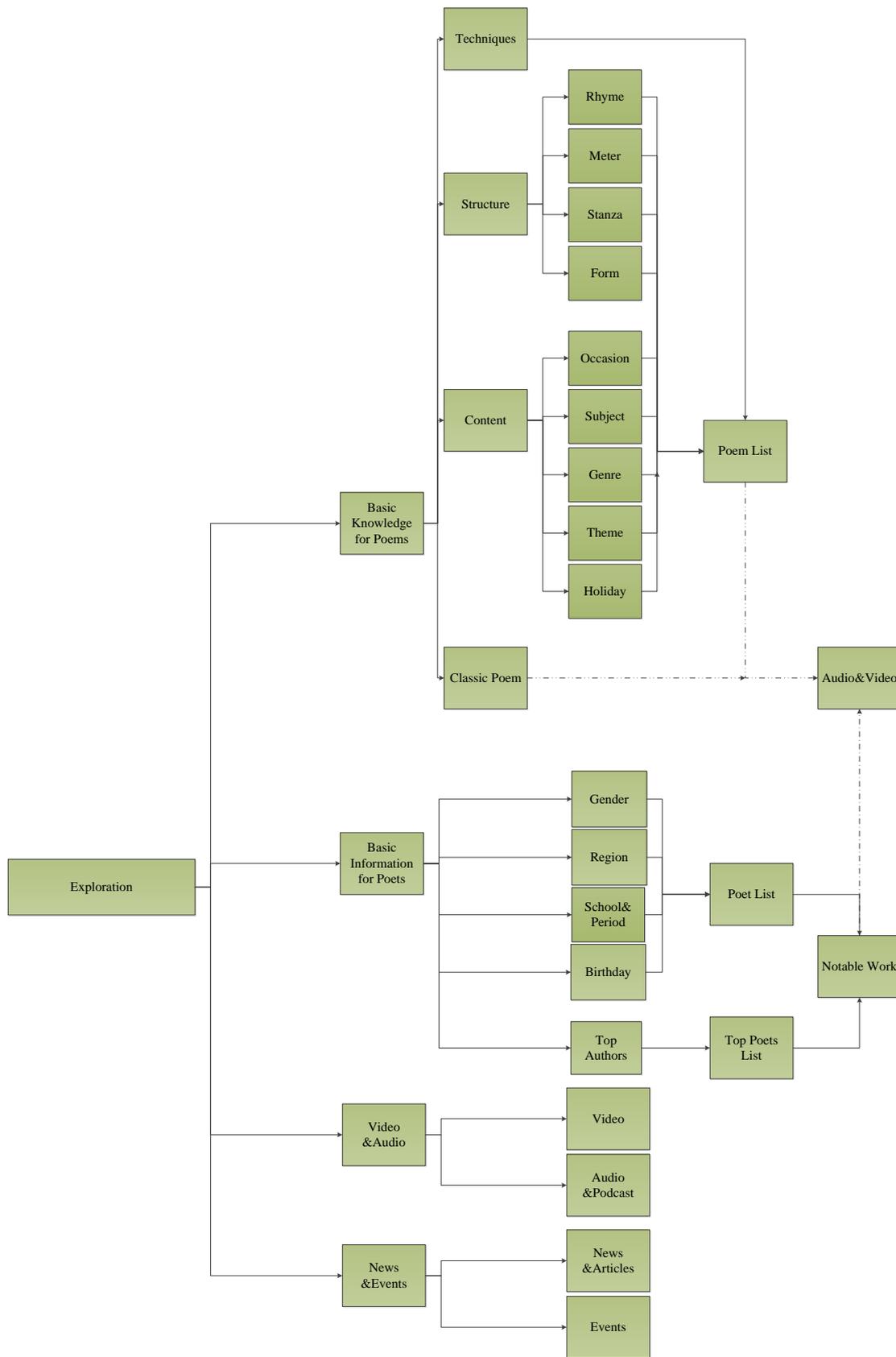


Figure 8.3 Design Details of Exploration Module

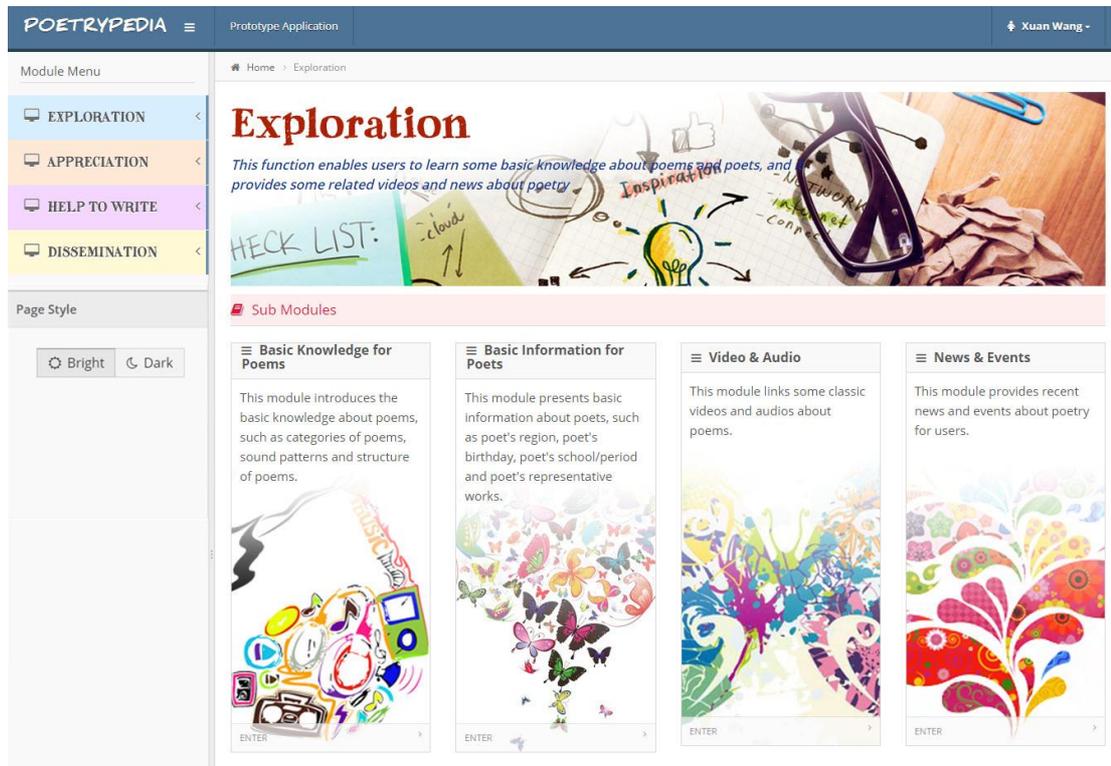


Figure 8.4 User Interface of Exploration Module

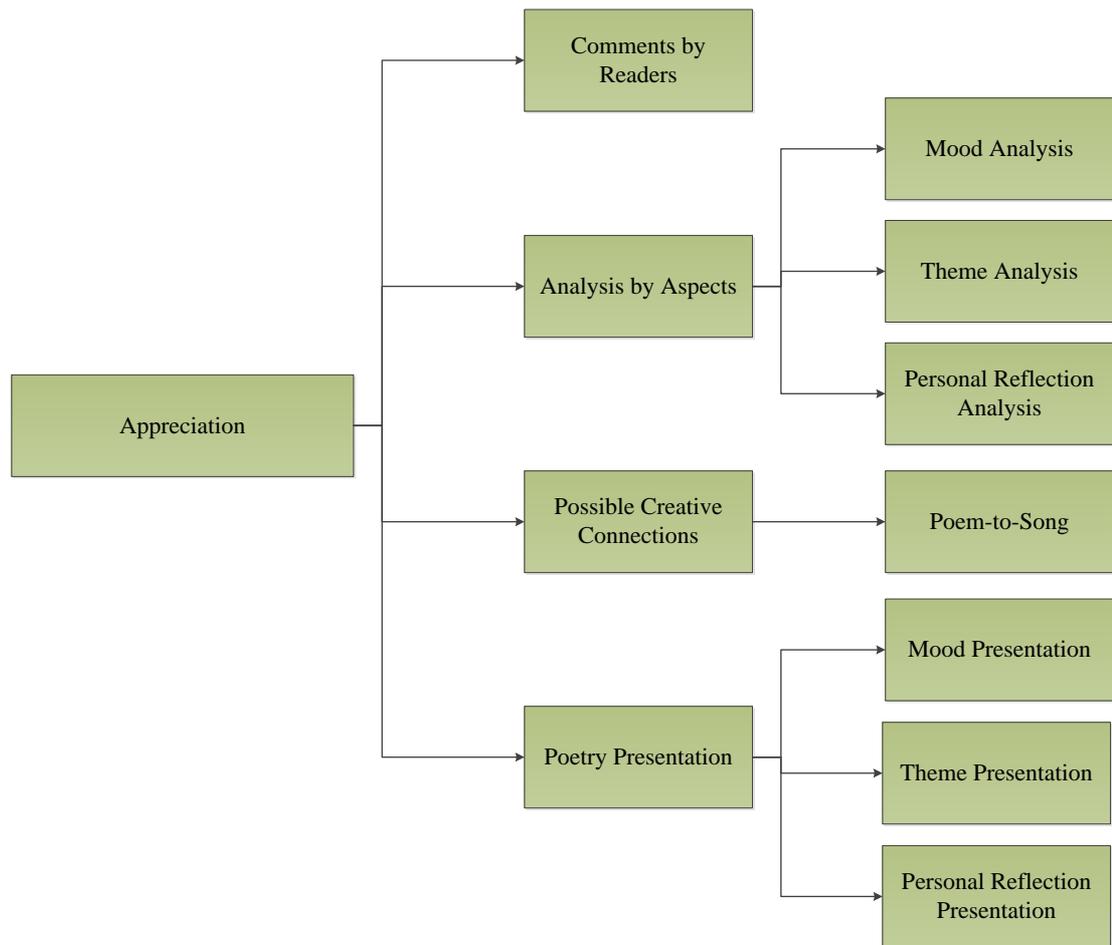


Figure 8.5 Design Details of Appreciation Module

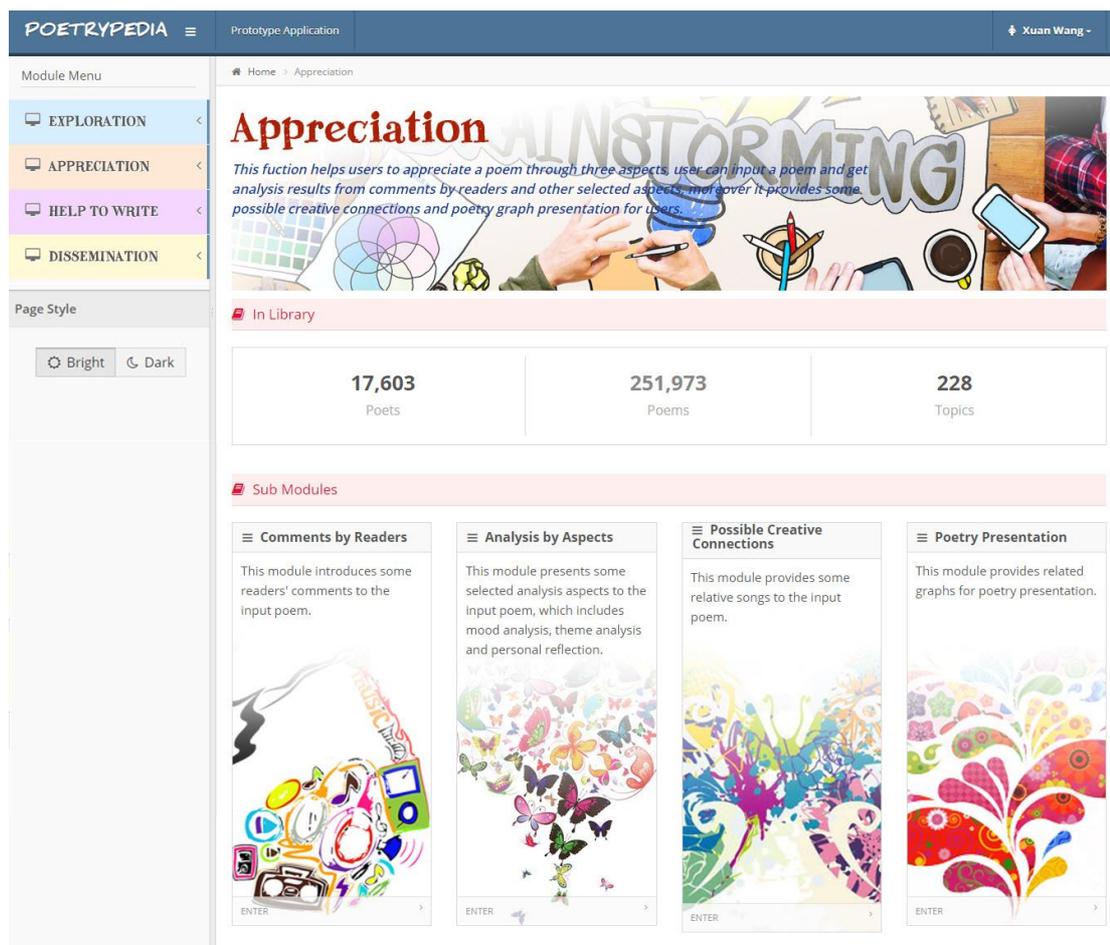


Figure 8.6 User Interface of the Appreciation Module

The Help-to-Write module assists users to explore various forms of poetry, and recommends various synonyms of words for users. It also has a poem fill-in-the-blank game. The design details of the Help-to-Write module are shown in Figure 8.7. The Help-to-Write module not only helps user compose, but provides opportunities for the user to innovate. There are three functions in this module: Poem Generation, Word Recommendation and Poem Fill-in-the-Blank. Section 8.3 will explain these functions in detail. The user interface of the Help-to-Write module is shown in Figure 8.8.

The Funding, Opportunity and Dissemination module provides services about funding about poetry and competition information. Meanwhile, if users want to publicise poems on related journal or magazines, the related information about publication formats, time, and copyright will also be provided. The design details and user interface are shown in Figure 8.9 and Figure 8.10

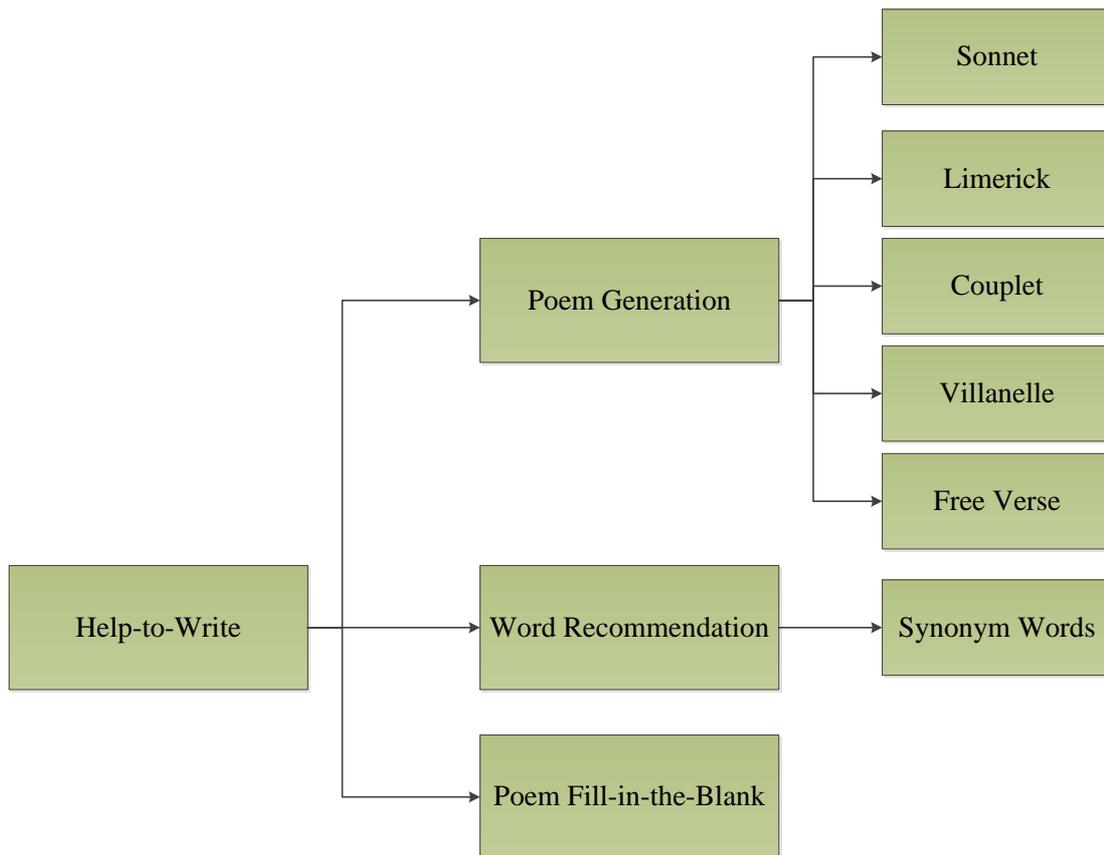


Figure 8.7 Design Details of Help-to-Write Module

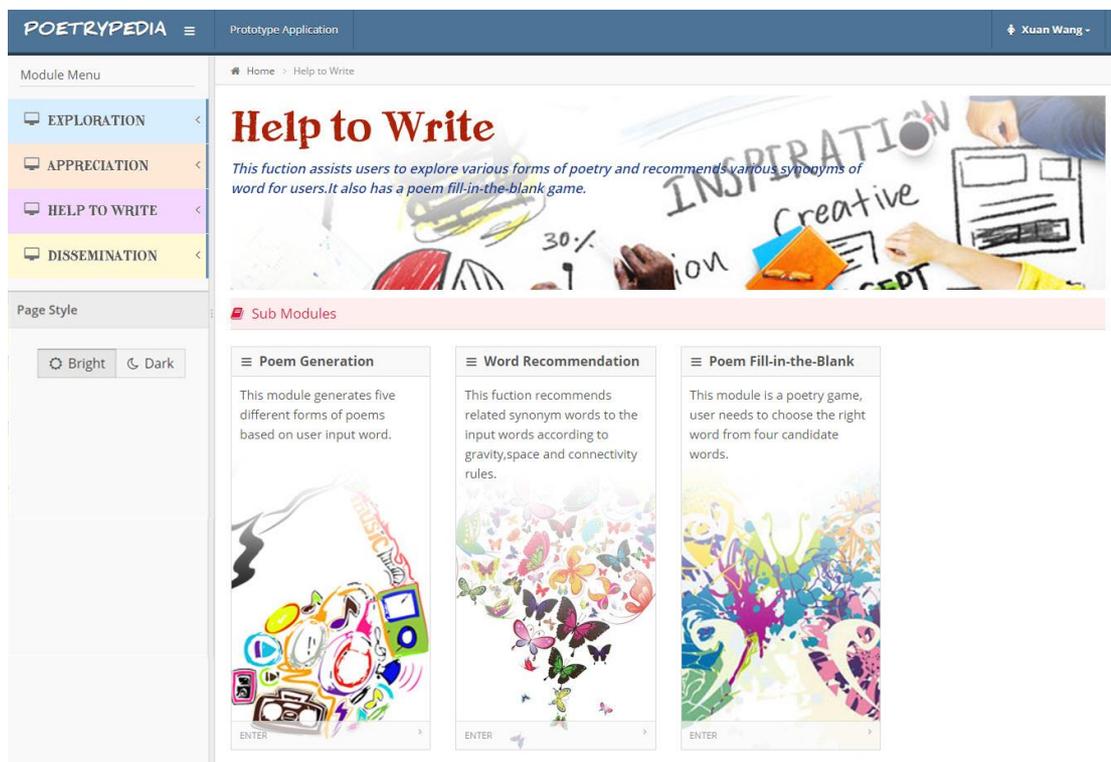


Figure 8.8 User Interface of Help-to-Write Module

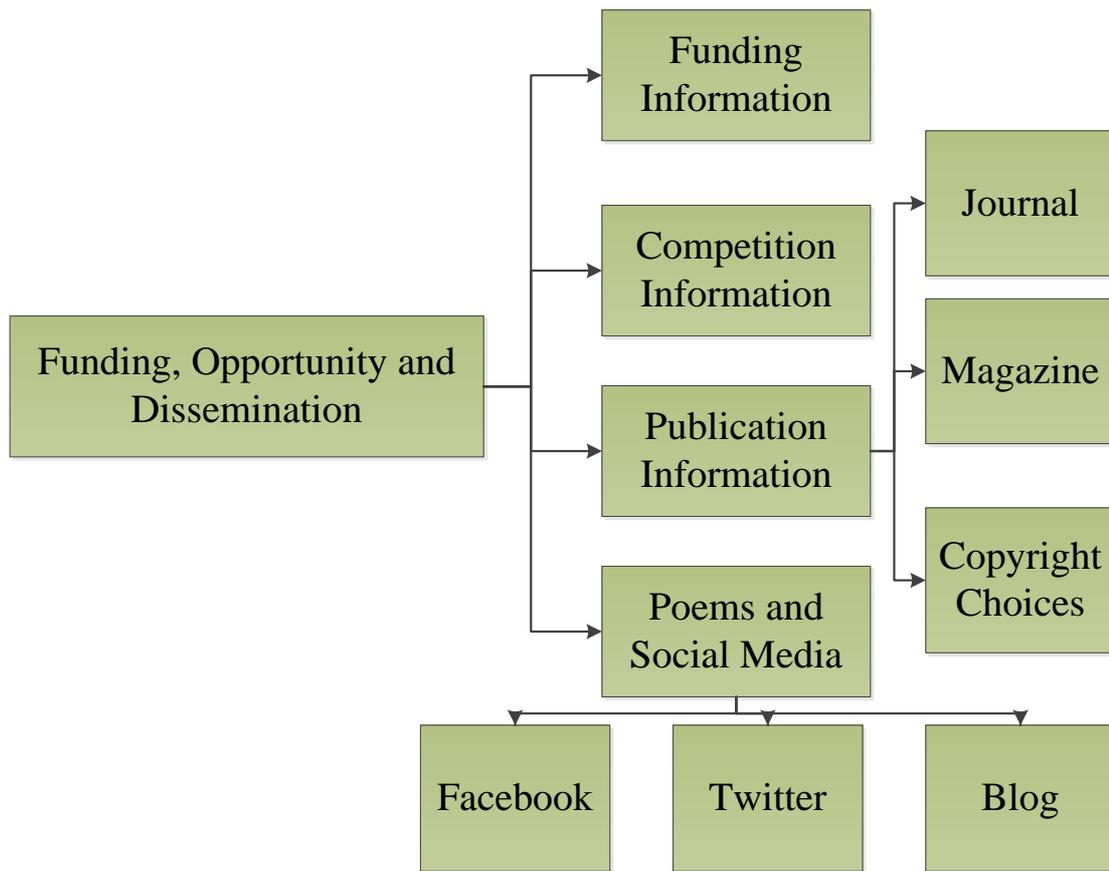


Figure 8.9 Design Details of Dissemination Module

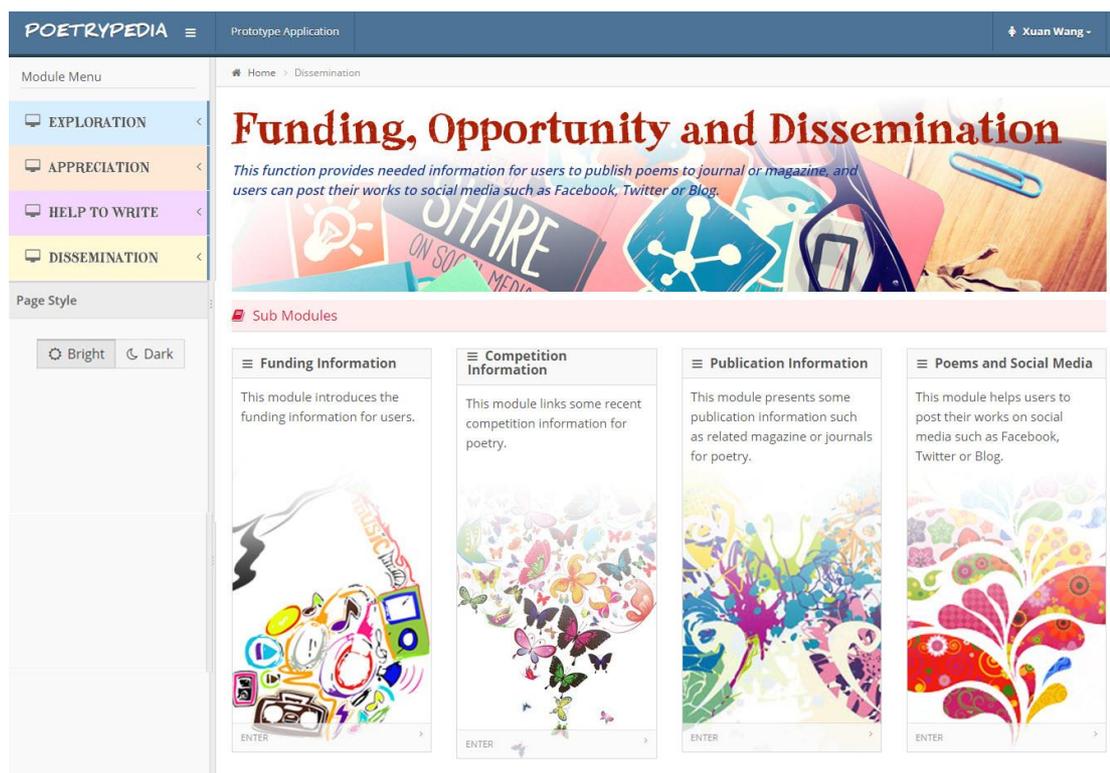


Figure 8.10 User Interface of Dissemination Module

8.3 Case Studies Summary

In order to illustrate the proposed creative data manipulation approach, this section presents six functions of PoetryPedia based on the proposed creative rules which apply to the approach: Analysis by Aspects, Possible Creative Connections, Poetry Presentation, Poem Generation, Word Recommendation, and Poem Fill-in-the-Blank. The user interfaces and workflows of the six functions are explained respectively.

8.3.1 Function of Analysis by Aspects in PoetryPedia

This section introduces the function of Analysis by Aspects in Appreciation module. It is designed to provide creative analysis results for poetry based on three aspects: mood analysis, theme analysis and personal reflection analysis. To use the Analysis by Aspects function, a user needs to enter a title or choose one sample poem at the input interface. In this case, the poem *Dreams* is input which is shown in Figure 8.11.

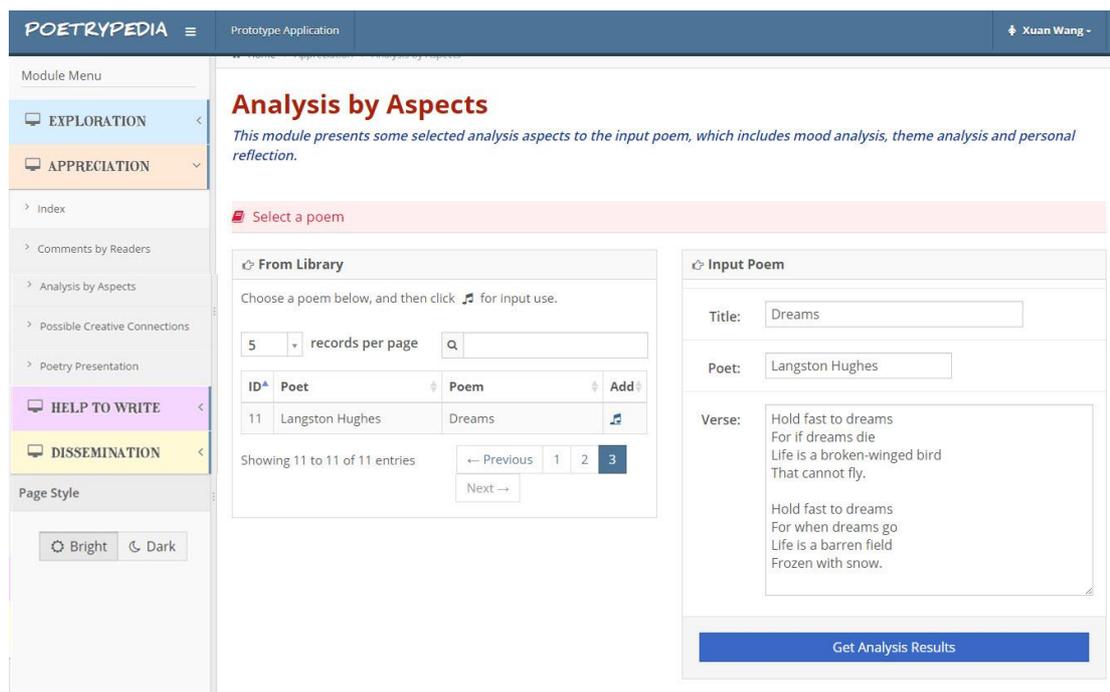


Figure 8.11 User Input of Analysis by Aspects

The design workflow of Analysis by Aspects is as follows:

- Step 1. Applying web crawler to collect words and poem information.
- Step 2. Storing the collected data into the poem database.
- Step 3. Selecting the chosen poem and splitting each word of the poem.
- Step 4. Applying the Neutralisation Rule to calculate the pH value of the chosen poem.
- Step 5. Confirming the theme value of the poem based on the Organic Substance rule.
- Step 6. Selecting the corresponding personal reflection sentence and math shape based on the Reduction-Oxidation rule.
- Step 7. Returning the creative analysis results to users.

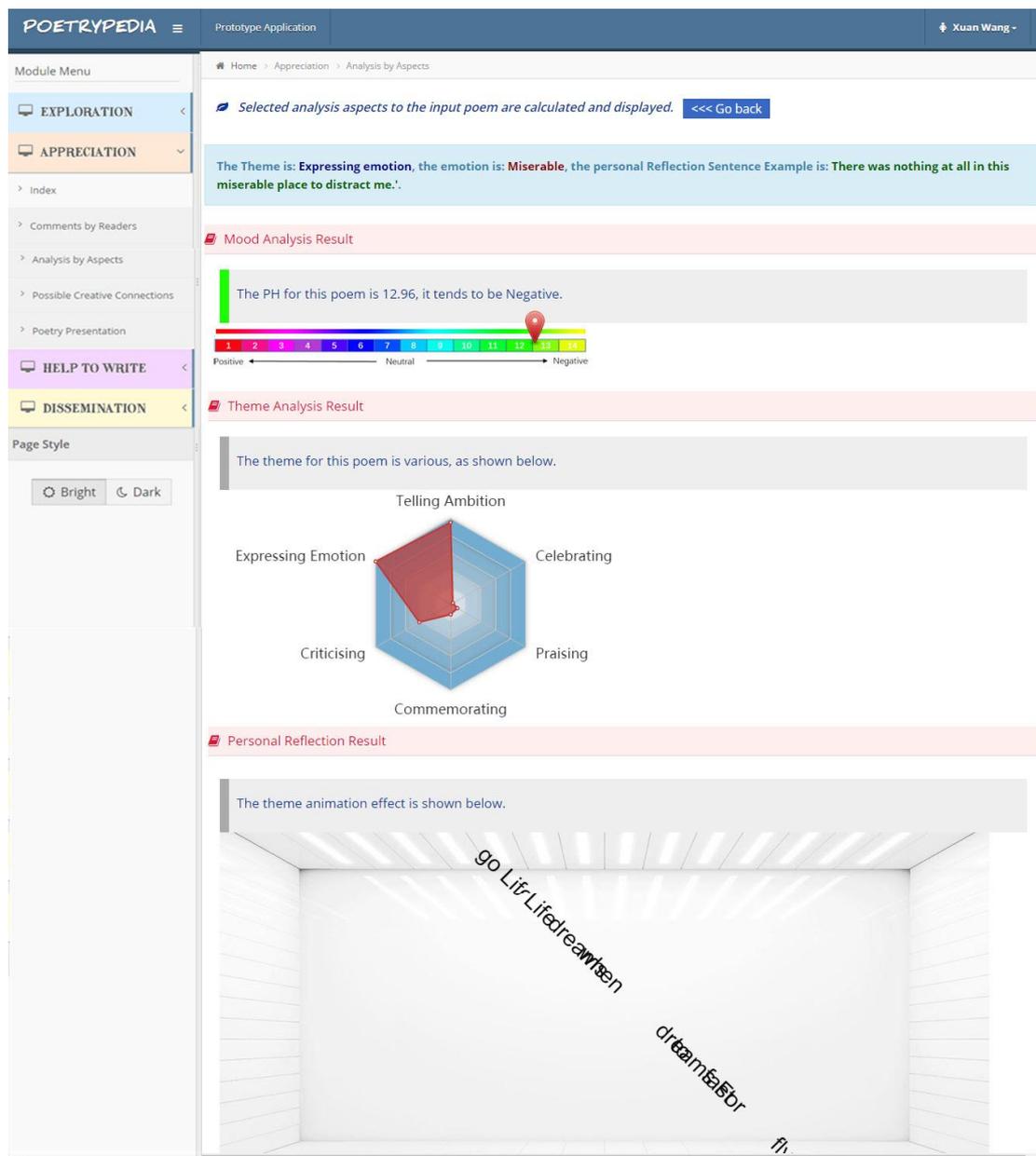


Figure 8.12 User Output of Analysis by Aspects

According to the rules and the workflow, the output results are shown in Figure 8.12.

A *pH*-bar is used to show the *pH* value of the solution. For instance, the *pH* of *Dreams* equals 12.96, which indicates negative emotion. A hexangular diagram is used to present the theme of the poem in which proportions of the six prime themes are disassembled. It can be inferred that *Dreams* is mainly for expressing emotion. Based on these analysis results, the mathematics curves of ‘miserable’ could be used to present this poem.

8.3.2 Function of Possible Creative Connections in Poetrypedia

The current function of the Possible Creative Connections in Poetrypedia is to create related song lists for input poems, and its ultimate goal is to allow users to explore innovative ways to appreciate poetry.

The design of this function is based on the Association Analysis. Association analysis is a simple, widely-used analysis technique that uncovers the associativity and relevancy among large quantities of data to extract patterns occurring simultaneously in things or objects. A typical example is the basket analysis, which reveals shopping habits of a customer through the association analysis of the products in their basket. The frequencies at which certain categories of products are purchased at the same time would help retailers plan strategies for sales. Other applications incorporate price list design, sales promotion, the layout of products, and customer categorisation based on shopping habits.

However, traditional association analysis has constraints in terms of patterns. In the pursuit of changing ways of thinking, inspiring creativity and offering diversity for the users, this work borrows the concept of Allotrope in chemistry to propose the transformational rule.

Allotrope refers to those simple substances consisting of identical chemical elements, but with different natures (disparate appearances of physical and chemical properties due to different location modes of elements). Primary differences among allotropes are reflected in physical properties: nevertheless, their activities behave differently in chemistry. For example, two allotropes of Phosphorus, - Red Phosphorus (P, amorphous) and White Phosphorus (P_4), - possess ignition points of **240°C** and **40°C**, respectively, and both become Phosphorus Pentoxide after sufficient burning. On the other hand, White Phosphorus is extremely poisonous and soluble in Carbon Disulfide, whereas, neither applies to Red Phosphorus. Allotropes are transformable under certain conditions which are in essence chemical changes.

This work combines the chemical characteristics with association analysis to uncover similar features among data. Not only homogeneous data is being investigated, but relationships between different types of data, in order to provide diverse results for users.

The Song Inference Rule based Association Analysis: The Song Inference Rule is to retrieve candidate data segments *poem* which has n_{th} features x_i ranked by users' preference ration w_i (i.e., weight for the segment feature). The formal definition of Song Inference Rule is as follows:

$$poem \xrightarrow{w_0x_0, w_1x_1, \dots, w_nx_n} song \quad (8.1)$$

Where *song* is the corresponding candidate data which is the most features related to the candidate data segments *poem*, and the sum of w_i should be equal to 1.

$$w_0 + w_1 + \dots + w_i = 1 \quad (8.2)$$

The screenshot shows the 'POETRYPEDIA' web application. The main content area is titled 'Possible Creative Connections' and includes a search bar and a table of poems. The 'Input Poem' section on the right has the following fields:

- Title:
- Poet:
- Verse:

A 'Get Related Songs' button is located at the bottom of the 'Input Poem' section.

Figure 8.13 User Input of Possible Creative Connections

To use the Possible Creative Connections function, a user needs to input a title of a poem or choose a sample poem from the menu at the input interface. In this case, the poem *Dreams* is input as shown in Figure 8.13.

The workflow of the Possible Creative Connections is as follows:

- Step 1. Applying word gravitation rule and web crawler to collect words, poems and poets' information, songs.
- Step 2. Filtering the collected data according to the Flotation Filter and Conductivity Calculator rules.
- Step 3. Storing the collected data in three databases: sentiment word database, poem style database and song library database.
- Step 4. Defining the core features of the keywords.
- Step 5. Applying the Neutralisation rule to decide the mood of keywords.
- Step 6. Confirming the style of keywords based on the Organic Substance rule.
- Step 7. Selecting candidate songs from the database, which have a similar style to the keyword.
- Step 8. Assigning different weights for the songs according to users requirements, and ranking the songs by the weights
- Step 9. Returning the candidate song list to the user.

According to the workflow of this function, by clicking the *Get Related Songs* button, the system returns the results to user, which includes the 49 candidate song lists and related lyrics, as shown in Figure 8.14.

There are 49 relevant songs shown in the song list. When the user clicks on a title, the Lyric module on the right will show the corresponding lyrics. The song list is incorporated in a dynamic circle in which a small circle represents a song. If one thinks 49 options are too many, the *Top 7 Songs* button will lead to the best matched seven songs found, as shown in Figure 8.15.

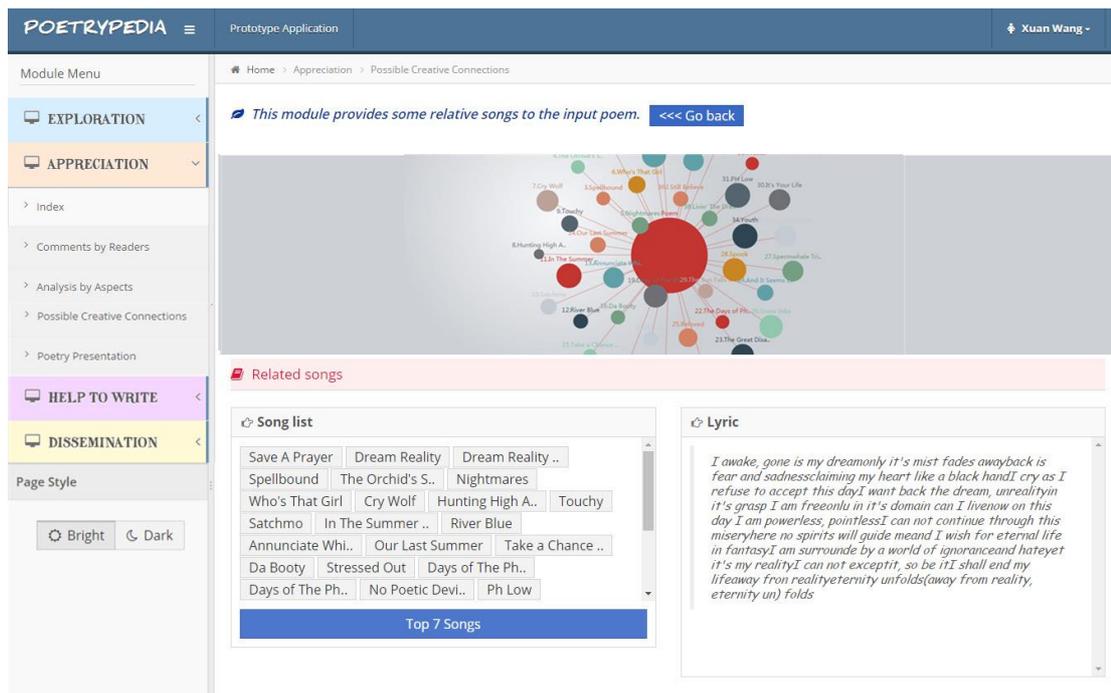


Figure 8.14 Songs List of Possible Creative Connections

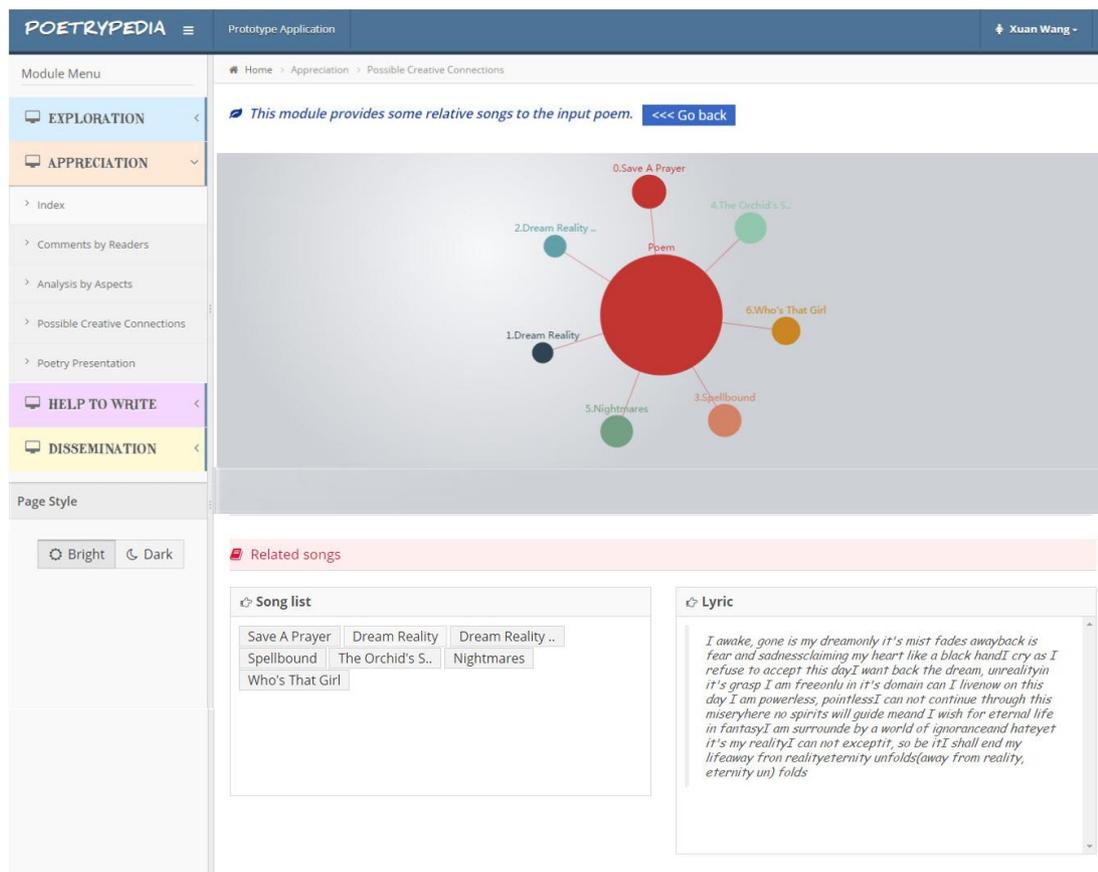


Figure 8.15 Top Seven Songs of Possible Creative Connections

8.3.3 Function of Poetry Presentation in Poetrypedia

If the user only wishes to see how to present a poem creatively, he could use the function of the Poetry Presentation in Appreciation module. The function of the Poetry Presentation attempts to exploit dynamic geometric shape to present the mood, theme and personal reflection of poetry, based on the creative presentation rules. The user is allowed to type in either existing poems, or his composition or a piece of dialogue. Taking *The Tiger* for example, the user interface is as shown in Figure 8.16.

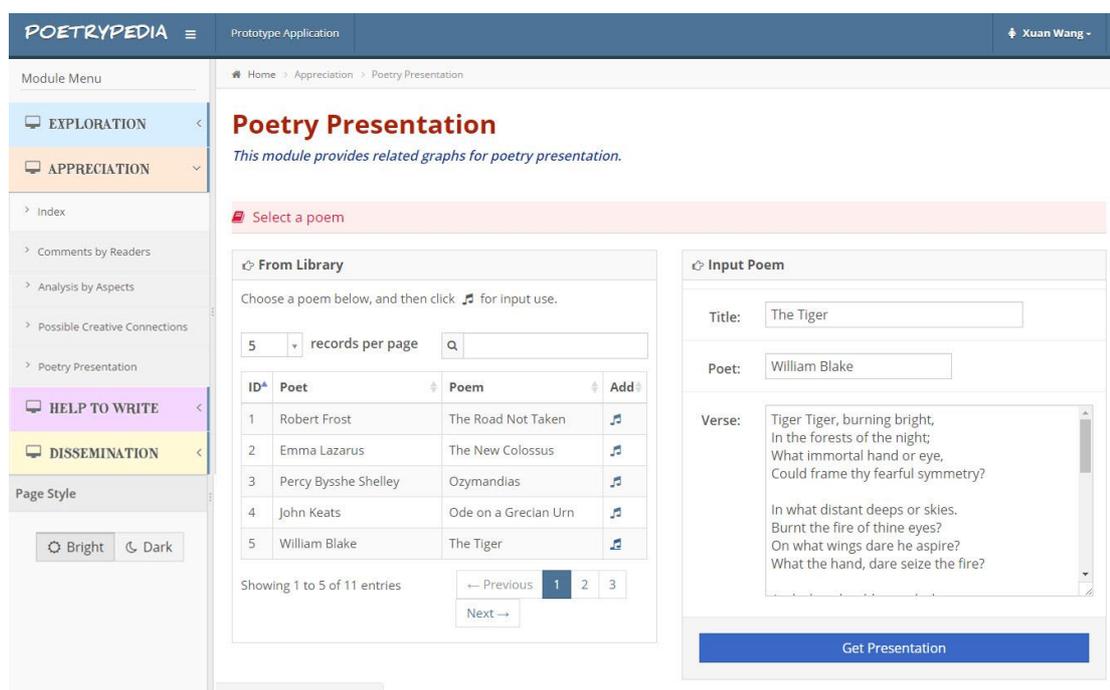


Figure 8.16 User Input of Poetry Presentation

Unlike Analysis by Aspects, Poetry Presentation directly demonstrates the geometric shapes and curves of the input. Via calculations, the emotion of *The Tiger* conforms to Sense of Lost. Its pattern of presentation is shown in Figure 8.17.

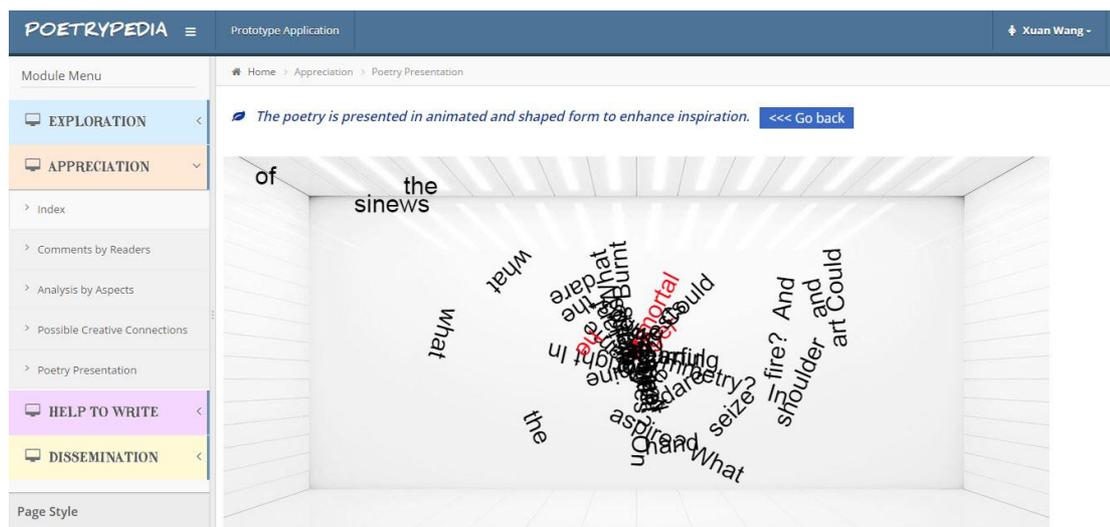


Figure 8.17 User Output of Poetry Presentation

8.3.4 Function of Poem Generation in PoetryPedia

This section introduces the function of the Poem Generation in the Help-to-Write module. This function is designed to help the user to compose different forms of poems, according to the theme. To use the Poem Generation function, a user needs to enter a concept at the input interface. The user is also allowed to choose a topic out of the list given. In this case, the concept of *dream* is input, which is shown in Figure 8.18.

Then, by clicking the *Get Various Poems* button, the system returns the results to the user, which involves poems about *dream* in five different forms. In order to broaden the view and inspire creativity for users, here this module presents the users with five different common formats of poems, including *Sonnet*, *Limerick*, *Couplet*, *Villanelle* and *Free Verse*. In order to provide room for creation, this module firstly addresses five poems of different forms referring to ‘*dream*’. Then the adjectives found according to the Word Gravitation Rule are replaced by new words and marked in red, as shown in Figure 8.19.

The design workflow of Poem Generation is as follows:

- Step 1. Applying web crawler to collect words and poem information.
- Step 2. Storing the collected data in the poem database.
- Step 3. Returning five poems of different forms based on the poem database.

- Step 4. Replacing the keywords and marking them in red, according to the Word Gravitation Rule.
- Step 5. Returning the creative poems to users.

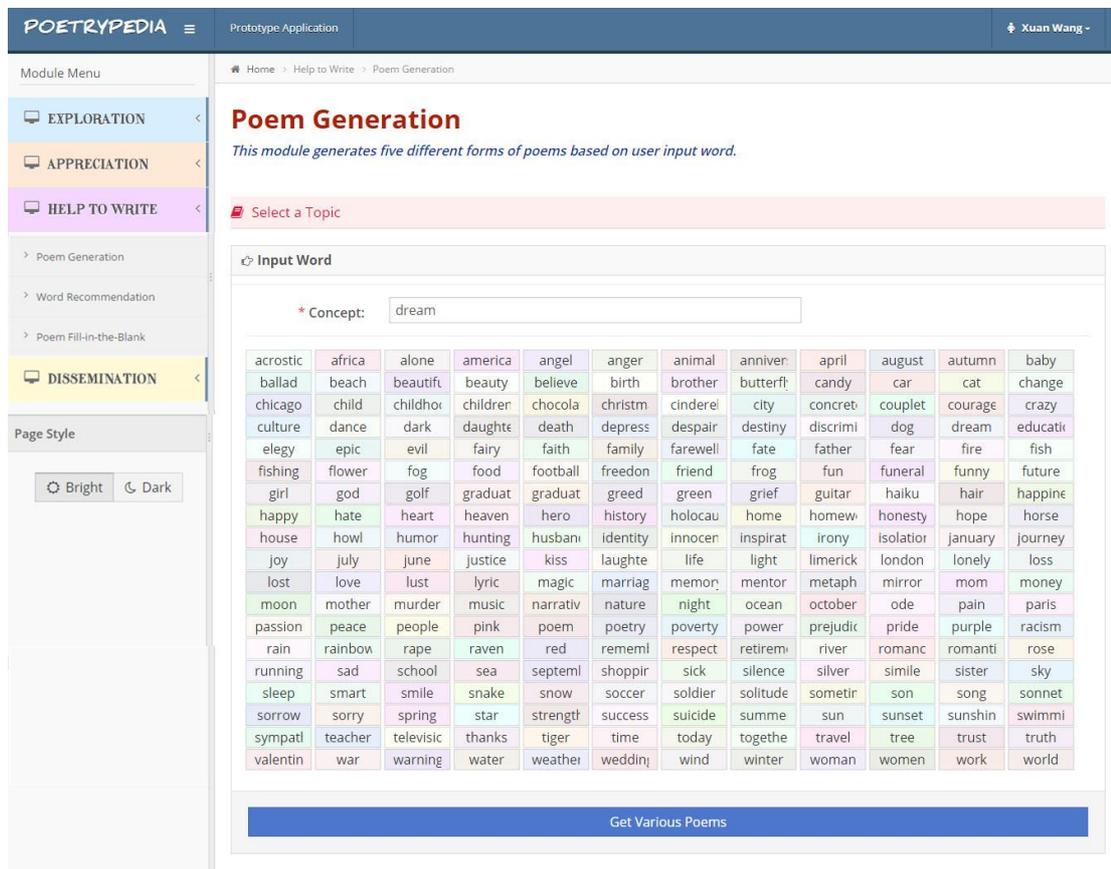


Figure 8.18 User Input of Poem Generation

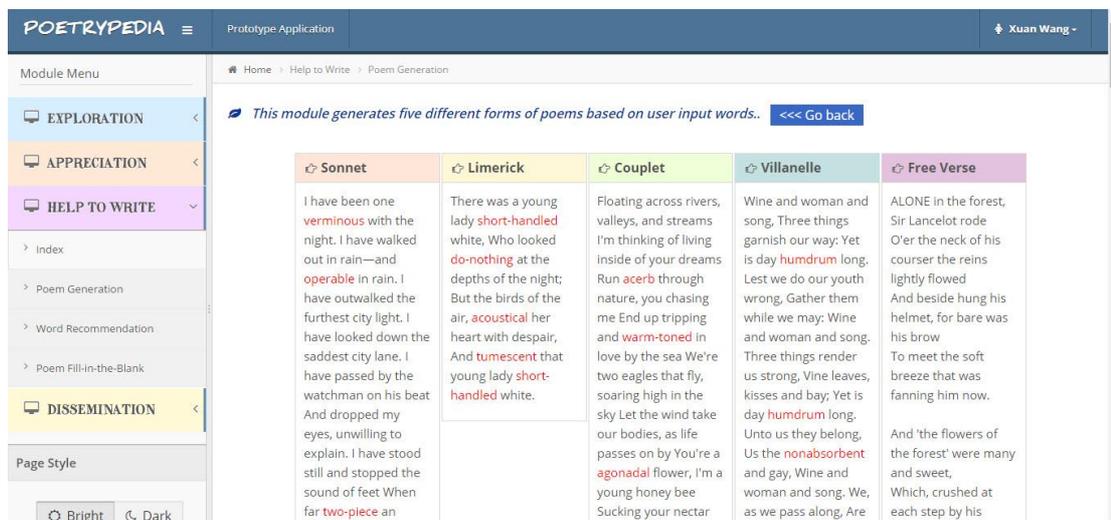


Figure 8.19 User Output of Poem Generation

8.3.5 Function of Word Recommendation in Poetrypedia

The current function of the Word Recommendation in Help-to-Write module recommends creative synonyms depending on the user input. The user is expected to type in a word, e.g., *dream*, as shown in Figure 8.20. According to the proposed creative rules, synonyms of ‘*dream*’ are presents from three different points of view, and ranked based on relevancy, as shown in Figure 8.21.

The design workflow of Word Recommendation is as follows:

- Step 1. Applying web crawler to collect words.
- Step 2. Storing the collected data in the word database.
- Step 3. Applying the Word Gravitation rule to calculate the value of words and ranking the results.
- Step 4. Listing the synonyms according to the Flotation Filter rule.
- Step 5. Ranking the synonyms based on the Conductivity Calculator rules.
- Step 6. Returning the creative synonyms results to users.

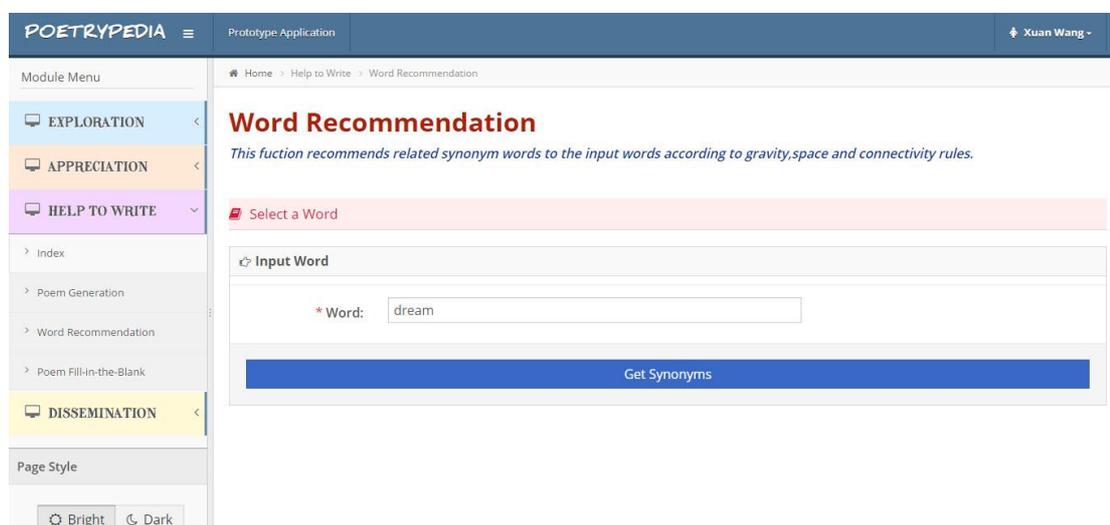


Figure 8.20 User Input of Word Recommendation

The screenshot shows the POETRYPEDIA interface with the word 'dream' selected. The interface is divided into three main sections for word recommendations:

- By Meaning:** A table showing words and their relevancy percentages.

Words	Value
image	Relevancy: 9 %
fantasy	Relevancy: 7 %
idea	Relevancy: 6 %
nightmare	Relevancy: 4 %
thought	Relevancy: 4 %
delusion	Relevancy: 2 %
- By Space:** A table showing words and their relative weights compared to 'dream'.

Words	Value
image	Weight equal to dream
idea	Weight lighter than dream
fantasy	Weight heavier than dream
thought	Weight heavier than dream
delusion	Weight heavier than dream
nightmare	Weight heavier than dream
- By Connectivity:** A table showing words and their conductivity values.

Words	Value
image	conductivity: 0.71
fantasy	conductivity: 0.54
idea	conductivity: 0.51
nightmare	conductivity: 0.34
thought	conductivity: 0.34
delusion	conductivity: 0.17

Figure 8.21 User Output of Word Recommendation

8.3.6 Function of Poem Fill-in-the-Blank in Poetrypedia

This section introduces the function of the Poem Fill-in-the-Blank in Help-to-Write module. It is designed to inspire users such as students, interest on poetry by playing a poem Fill-in-the-Blank game. Here this function takes *Dreams* as an example. Firstly, there is a poem with blanks and users need to complete it as shown in Figure 8.22. Subsequently, the user is allowed to select a word from the given options as displayed in Figure 8.23. Then the user is able to check whether his choice is correct by clicking on the *Check your Answer* button. This will help the user memorise the poem, as illustrated in Figure 8.24.

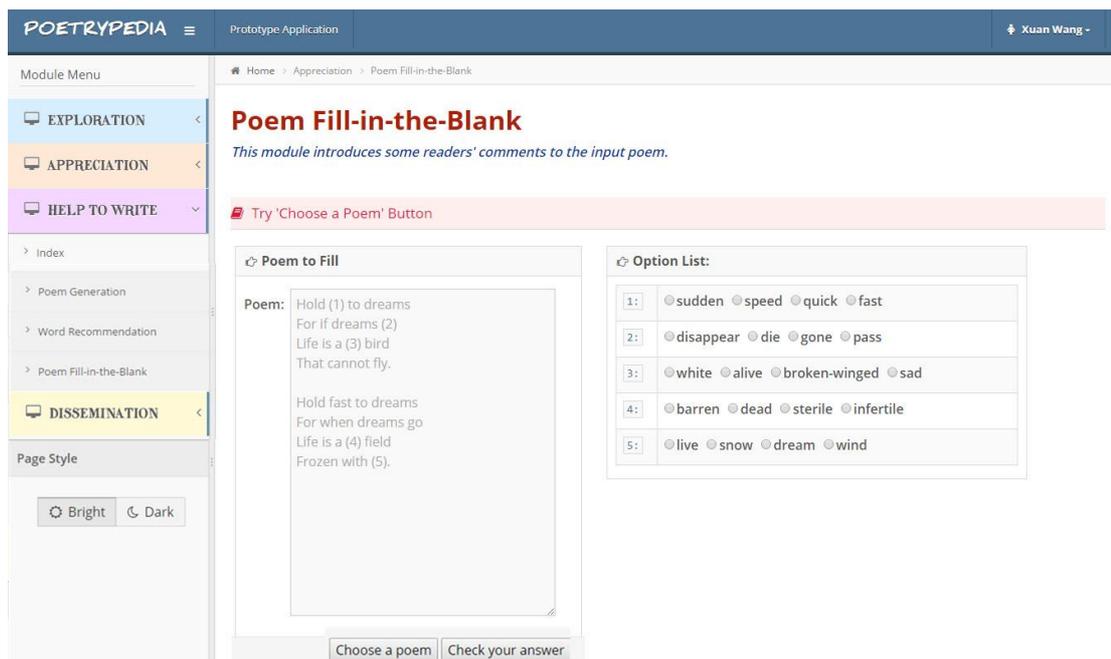


Figure 8.22 User Input of Poem Fill-in-the-Blank

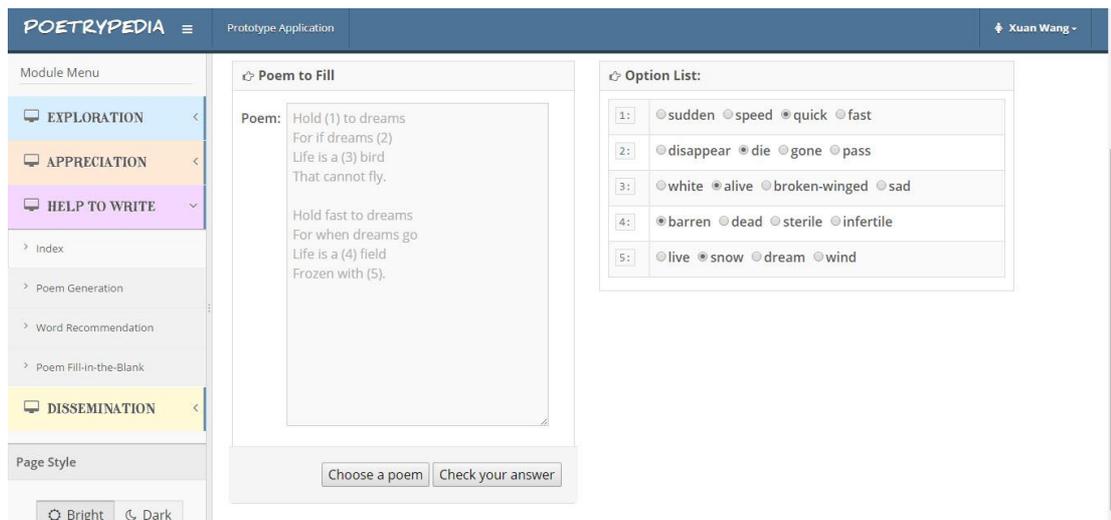


Figure 8.23 User Chosen of Poem Fill-in-the-Blank

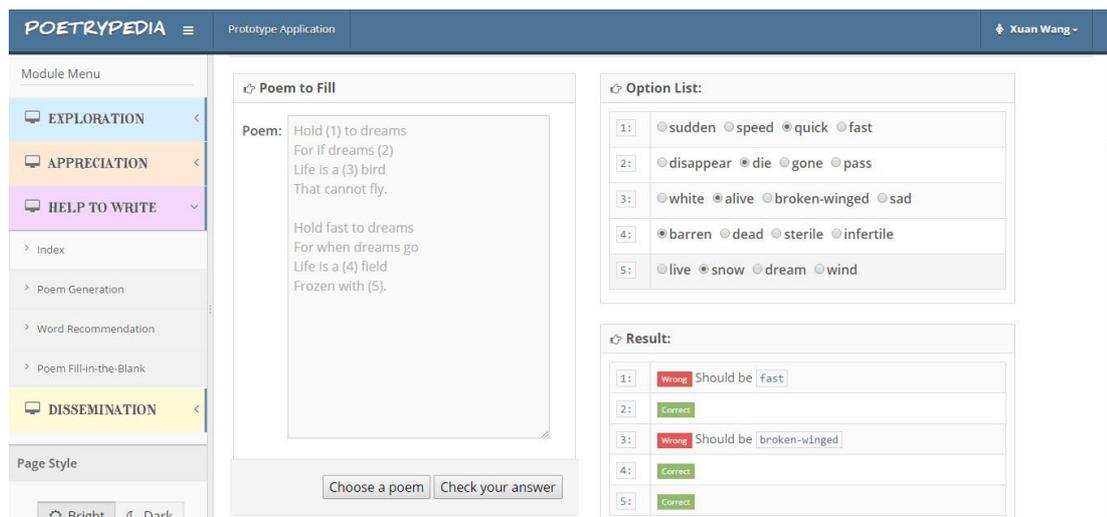


Figure 8.24 User Output of Poem Fill-in-the-Blank

8.4 Summary

This chapter explains how the proposed creative data manipulation approach is applied in a PoetryPedia system and presents its four major modules. Particularly, the prototypes of six functions, i.e., Analysis by Aspects, Possible Creative Connections, Poem Presentation, Poem Generation, Word Recommendation and Poem Fill-in-the-Blank are proposed as case studies. PoetryPedia verifies the effectiveness of our approach and, at the same time, presents a comprehensive, diverse and inspirational poem collection.

- The Analysis by Aspects function stems from creative analysis rules proposed. It not only provides conventional poem interpretation and analysis but introduces scientific principles into poem analysis, which offers opportunities for innovative appreciations of poems.
- It provides a Possible Creative Connections function which returns a list of related songs based on the user input of a poem's title or content. This function explores the relationship between two types of human arts and gives rise to the possibility of creative composition.

- The Poem Generation function offers five different formats of poems depending on the input concept. It serves the purposes of poetry exploration and creative composition.
- It details the Poem-Fill-in-the-Blank function which is essentially a blank filling game designed to inspire interest in learning poems. Besides that, it is intended to help users better memorise poems and get the hang of poetry.

Chapter 9

Conclusions and Future Work

Objectives

- To evaluate the research by answering the research questions and revisiting the success criteria
 - To revisit original contributions
 - To illustrate the limitation of the work
 - To propose future work
-

9.1 Summary

This research aims to investigate and implement a creative computing approach to the poetry data manipulation. Since traditional poetry data collection, processing and presentation methods lack versatility of implementation for different users and purposes, the major objective of this research is to explore poetry data manipulation strategies to inspire creative minds. In this thesis, it focusses on the development of a poetry data manipulation approach which adopts a combination of multidisciplinary concepts to boost creativity. In order to validate this poetry data manipulation approach, a Poetrypedia system has been proposed to retrieve, analyse and present real-world poetry databases.

The proposed poetry data manipulation approach can be divided into three phases: Multi-Purpose Poetry Data Collection, Multi-Dimension Poetry Data Analysis and Multi-Media Poetry Data Presentation. The web crawler and laws of physics are deployed to assist poetry data collection and pre-processing. Principles in chemistry are exploited to aid poetry data analysis. Geometrical curves and shapes are used to present poetry data analysis results in an inspiring fashion. The Poetrypedia system,

which takes the form of a website, incorporates all the above methods, and provides an interactive platform for different users to touch and feel poems in a creative way.

9.2 Evaluation

The definition and forms of creativity have been broadly reviewed in Chapter 2. This research has also attempted to define creativity in terms of poetry data manipulation. Since the knowledge in science and engineering is transformed to solve the problems in the literature, the proposed creative data manipulation approach is capable of generating creativity. Specifically, the laws of physics are borrowed to design multi-purpose data collection rules (Chapter 5). Principles of chemistry are used to design the creative multi-dimension analysis rules (Chapter 6) and geometrical curves and shapes are used for multi-media data presentation (Chapter 7).

To provide users with experiences that accurately reflect their wants and needs in an authentic and creative manner, poetry data analysis needs to pay more attention to digging deeply for the meaning and sentiment of data. This work introduces a multi-dimension approach which considers the meaning of the data from three aspects: mood, theme and personal reflection.

This work presents nine creative reasoning rules to support the poetry data manipulation. The Word Gravitation Rule, Flotation Filter Rule and Conductivity Calculator Rule are designed based on laws of physics to realise multi-purpose poetry data collection. The Neutralisation Rule, Organic Substance Rule and Reduction-Oxidation Rule come from principles of chemistry, and they help to analyse the poetry data. The Mood Presentation Rule, Theme Presentation Rule and Personal Reflection Presentation Rule are used to present poetry analysis results based on the implications of curves and shapes in mathematics.

In this thesis, case studies are carried out to validate the proposed creative poetry data manipulation approach through the Poetrypedia system. There are four modules in this system: Exploration, Help to Write, Appreciation and Dissemination. All the

functions are purposefully designed to assist users to comprehend and to compose in a creative way. The developed applications in the case studies illustrate how the creative poetry data manipulation approach works for the four modules. By presenting and evaluating the case studies, the feasibility and generality of the proposed approach have been verified.

In addition, the web crawler is applied for multi-purpose poetry data collection (Section 5.2.3). C# is the main programming language for poetry data analysis. The curves and shapes for data presentation are generated by Matlab (Section 7.2-7.4).

9.3 Contributions

This research contributes to extending the scope of creative computing in terms of poetry data manipulation. The proposed approach particularly advances the understanding of poetry data accessible from the Web. The details of contributions are as below:

- C1: This research presents and validates an innovative and applicable approach to providing creative poetry data manipulation.
- C2: This research develops a series of algorithms to bring creativity into multi-purpose poetry data collection, multi-dimension poetry data analysis and multi-media poetry data presentation processes. These rules are proposed upon inspirational and reasonable combinations of classic laws in physics, principles in chemistry and the symbolism of geometric shapes.
- C3: Via the development of the Poetrypedia, this research has demonstrated the capability, as well as significance, of assisting educational studies of literary data, providing heuristic poetry analysis and stimulating human artistic creation.

9.4 Future Work

The research presented in this thesis is not the end of this study. Based on the present progress, it is suggested that future work may be undertaken in the following parts to enrich the data manipulation theories.

Creativity can be a subjective concept which varies according to the different people working with it, the purposes to which it is being put, or the situations in which it occurs. Creativity is also temporally dynamic. Thus, a broader, in-depth and constant study of the definition of creativity would help us determine how data manipulation schemes should be designed.

Although the analysis and presentation work well on the poetry database in the case studies, it is worth investigating the effectiveness of this poetry data manipulation approach on different types of poems. Surveys and feedback from the Poetrypedia website are significant in the evaluation of the performance of this approach on different types of poems in terms of generating creativity.

In this research, specific laws and principles of physics, chemistry and mathematics are adapted to enhance the poetry data manipulation approach. Generally, these selected laws and principles possess a similar format to the target data problem. For instance, the Word Gravitation Rule calculates the relevancy between two words by specifying masses and the separation distance of two words, for which it uses Newton's law of universal gravitation. Future work will exhaustively search similarities between laws in science and engineering and the needs of data manipulation. This would be able to show what kind of laws and principles are particularly helpful in inspiring poetry data manipulation.

The horizons of poetry data analysis need to be broadened. Apart from mood, theme and personal reflection, a great number of other sentiment analysis parts of poems require further study. Other topics to be investigated include metaphor, style, as well as underlying effects. Many proposed concepts need further enrichment. For example, rather than defining a word of positive or negative tone, i.e., acidity or basicity, a vector of pH can be built to measure happiness, excitement or

patriotism to different degrees. An extended collection of presentation forms is also indispensable, to express the corresponding poetry analysis results.

It is also worth expanding this approach towards a wider range of literary data such as reports, novels, drama, etc. Inspiring creativity in these contexts would make a more profound contribution to learning, comprehension and the generation of data.

References

- [1] N. S. Altman, “An Introduction to Kernel and Nearest-Neighbour Nonparametric Regression”, *the American Statistician*, Taylor & Francis Group, United Kingdom, vol. 46, 1992, pp. 175–185.
- [2] B. G. Andreas, *Experimental Psychology*, John Wiley & Sons Inc., United States, 1972.
- [3] C. Jensen and L. Anderson, *Harvard Graphics 3: The Complete Reference*, Osborne McGraw-Hill Publisher, United States, 1992.
- [4] L. W. Beineke, “Characterizations of Derived Graphs”, *Journal of Combinatorial Theory*, Elsevier, United Kingdom, vol. 9, 1970, pp. 129-135.
- [5] A. M. Robert and E. S. Bishop, “Elements of Electricity”, *Journal of American Technical Society*, American Technical Publishers, United States, 1917.
- [6] A. Bierce, "For Brevity and Clarity", *the Collected Works of Ambrose Bierce*, the Neale Publishing Company, United States, 1912, pp. 386–387.
- [7] F. Birren, *the Power of Colour: How it Can Reduce Fatigue, Relieve Monotony, Enhance Sexuality, and More*, Carol Publishing Group, United States, 1992.
- [8] D. Kaplan and D. Blei, “A Computational Approach to Style in American Poetry”, *7th IEEE International Conference on Data Mining*, IEEE, Omaha, USA, 2007, pp. 553-558.
- [9] A. Chapman, M. D. Allen and B. Blaustein, “It’s About the Data: Provenance as a Tool for Assessing Data Fitness”, *4th USENIX Workshop on the Theory and Practice of Provenance*, USENIX, Boston, 2012.
- [10] S. Bleicher, *Contemporary Colour Theory and Use*, Cengage Learning Ltd, New York, United States, 2005.
- [11] M. A. Boden, Computer Models of Creativity, *AI Magazine*, American Association for Artificial Intelligence Press, United States, vol. 30, no. 3, 2009, pp. 23–34.

- [12] M. A. Boden, *the Creative Mind: Myths and Mechanisms*, 2nd ed. Routledge, London, 2004.
- [13] N. Bonnardel, "Creativity in Design Activities: The Role of Analogies in a Constrained Cognitive Environment", *3rd Conference on Creativity and Cognition*, ACM, New York, 1999, pp. 158–165.
- [14] Y. Braoudakis, *Why You Should Care About the Creativity Era!*, Online Blog, <http://innovation360group.com/whyyou-should-care-about-the-creativity-era/7>, Accessed Date, 09/12/2017.
- [15] M. Burner, "Crawling Towards Eternity Building an Archive of the World Wide Web", *Web Techniques Magazine*, CMP Technology, United States, vol. 2, 1997.
- [16] J. F. Burrows, *Computation into Criticism: a Study of Jane Austen's Novels and an Experiment in Method*, Clarendon Press, United Kingdom, 1987.
- [17] I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Press, New York, 2008.
- [18] R. R. Christopher, *How to Analyse Poem*, Monarch Press, Canada, 1966.
- [19] A. J. Reagan and J. Christiansen, "The Emotional Arcs of Stories are Dominated by Six Basic Shapes", *EPJ Data Science*, Springer, United States, vol. 5, no. 1, 2016.
- [20] W. S. Cleveland, *The Elements of Graphing Data*, Pacific Grove, Wadsworth & Advanced Book Program, United States, 1985.
- [21] P. Cohen, "Digital Keys for Unlocking the Humanities' Riches", *the New York Times*, A.G.Sulzberger, United States, 2010.
- [22] O. Andrew, G. Clore and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, United Kingdom, 1988.
- [23] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd ed. Sage Publications, Inc., California, United States, 2003.
- [24] S. Easterbrook, J. Singer, M. A. Storey, and D. Damian, "Selecting Empirical Methods for Software Engineering Research", *Guide to Advanced Empirical Software Engineering*, Springer-Verlag, London, United Kingdom, 2008, pp. 285–311.

- [25] R. Damton, *Poetry and the Police: Communication Networks in Eighteenth-Century Paris*, Harvard University Press, United States, 2012.
- [26] F. Michael and D. Dan, “the Early Origins and Development of the Scatterplot”, *Journal of the History of the Behavioral Sciences*, John Wiley & Sons, United States, vol. 41, 2005, pp. 103-130.
- [27] S. Drozd et al, “Quantifying Origin and Character of Long-Range Correlations in Narrative Texts”, *Information Sciences*, SAGE Publication Ltd, United Kingdom, vol. 331, 2016.
- [28] D. Elson, “*Detecting Story Analogies from Annotations of Time, Action and Agency*”, Proceedings of the LREC 2012 Workshop on Computational Models of Narrative, LREC, Istanbul, Turkey, 2012.
- [29] A. C. Fang, et al, “Adapting NLP and Corpus Analysis Techniques to Structured Imagery Analysis in Classical Chinese Poetry”, *AdaptLRTtoND Workshop on Adaptation of Language Resources and Technology to New Domains*, FLReNet Project, Stroudsburg, PA, USA, 2009, pp. 27-34.
- [30] P. Gautam, S. Padmini and M. Filippo, “Crawling the Web”, *Web Dynamics: Adapting to Change in Content, Size, Topology and Use*, Springer, United States, 2004, pp. 153–178.
- [31] S. Argamon, M. Koppel, and J. Fine, “Gender, Genre, and Writing Style in Formal Written Texts”, *Text&Talk*, Mouton de Gruyter Ltd, Germany, vol. 23, issue 3, 2003.
- [32] M. Fischetti, “Great Literature is Surprisingly Arithmetic”, *Scientific American*, Springer Nature, United States, 2017.
- [33] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge University Press, United Kingdom, 2005.
- [34] J. Gage, *Colour and Meaning: Art, Science and Symbolism*, Thames and Hudson Ltd, London, United Kingdom, 2000.
- [35] D. Genzel, et al., “Poetic Statistical Machine Translation: Rhyme and Meter”, EMNLP Conference on Empirical Methods in Natural Language Processing, SIGDAT, Massachusetts, USA, SIGDAT, 2010, pp. 158-166.

- [36] M. Wilk and R. Gnanadesikan, "Probability Plotting Methods for the Analysis of Data", *Biometrika*, Oxford University Press, United Kingdom, vol. 55, 1968, pp. 1-17.
- [37] E. Greene, et al., "Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation", *EMNLP Conference on Empirical Methods in Natural Language Processing*, SIGDAT, Massachusetts, USA, 2010, pp. 524-533.
- [38] L. Polak, J. Green, "Using Joint Interviews to Add Analytic Value", *Qualitative Health Research*, SAGE Publications Inc., United States, vol. 26, 2015.
- [39] J. Gajadhar and J. Green, "the Importance of Nonverbal Elements in Online Chat", *Educause Quarterly*, EDUCAUSE Review, United States, vol. 28, 2005.
- [40] D. Guerrier, *Archimedes: Archimedes' Principle and the Law of Flotation*, Blond & Briggs Publisher, United Kingdom, 1970.
- [41] S. W. Hawking, *the Universe in a Nutshell*. Bantam Books, New York, 2001.
- [42] L. Haas, "Integrating Extremely Large Data is Extremely Challenging", *The Extremely Large Databases Conference at Asia*, ACM, Beijing, China, 2012.
- [43] R. Resnick and D. Halliday, *Physics*, Vol I and II, Wiley International Edition, United States, 1966.
- [44] Z. He, et al, "SVM-based Classification Method for Poetry Style", *IEEE International Conference on Machine Learning and Cybernetics*, IEEE, Hong Kong, China, 2007, pp. 2936-2940.
- [45] J. Hendler and A. Hugill, "The Syzygy Surfer: Creative Technology for the World Wide Web", *ACM Web Science Conference*, ACM, New York, 2011, pp. 1-3.
- [46] "Tables, Rows, and Columns", IBM Article, United States, Retrieved 11 December 2013.
- [47] M. A. Runco and G. J. Jaeger, "The Standard Definition of Creativity", *Creativity Research Journal*, Routledge, United Kingdom, vol. 24, 2012, pp. 92-96.
- [48] C. Chao, J. Yang, and W. Jen, "Determining Technology Trends and Forecasts of RFID by A Historical Review and Bibliometric Analysis from 1991 to 2005", *Technovation*, vol. 27, no. 5, Elsevier Science, Amsterdam, 2007, pp. 268-279.

- [49] M. L. Jockers, *Macroanalysis: Digital Methods and Literary History*, University of Illinois Press, United States, 2013.
- [50] B. Berlin and P. Kay, *Basic Colour Terms: Their Universality and Evolution*, Berkeley: University of California Press, United States, 1969.
- [51] S. Kura, "Qualitative and Quantitative Approaches to the Study of Poverty: Taming the Tensions and Appreciating the Complementarities", *the Qualitative Report*, Sage Publications, United States, vol. 17, 2012, pp. 1–19.
- [52] J. L. Kulikowski, *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends*, IGI Global, United States, 2009.
- [53] Nodus Labs, "US Presidents Inaugural Speeches 1969-2013 Text Network Analysis",
<http://noduslabs.com/cases/presidents-inaugural-speeches-text-network-analysis/>
Accessed Date, 09/12/2017.
- [54] D. M. Levin, *the Opening of Vision: Nihilism and the Postmodern Situation*, Routledge, United Kingdom, 1988.
- [55] K. L. Lerner and B. L. Lerner, *the Gale Encyclopedia of Science*, Gale, Cengage Learning, United States, 2014.
- [56] E. G. Guba and Y. S. Lincoln, *Effective Evaluation*, San Francisco, Jossey-Bass, United States, 1981.
- [57] E. G. Guba and Y. S. Lincoln, *Naturalistic Inquiry*, Beverly Hills, Sage, United States, 1985.
- [58] B. Liu, *Opinions, Sentiment, and Emotion in Text*, Cambridge University Press, United Kingdom, 2015.
- [59] R. Manurung, "Using Genetic Algorithms to Create Meaningful Poetic Text", *Journal of Experimental & Theoretical Artificial Intelligence*, Taylor & Francis, United Kingdom, vol. 24, issue. 1, 2012, pp. 43-64.
- [60] M. Canducci, M. Missikoff, and N. Maiden, *Enterprise Innovation: From Creativity to Engineering*, Wiley, United States, 2015.

- [61] J. Macqueen, "Some Methods for Classification and Analysis of Multivariate Observations", *Mathematical Statistics and Probability*, University of California Press, United States, vol. 1, 1967, pp. 281-297.
- [62] S. Manikandan, "Frequency Distribution", *Journal of Pharmacology & Pharmacotherapeutics*, American Society for Pharmacology and Experimental Therapeutics, United States, vol. 2, 2011, pp. 54-55.
- [63] J. Miles, "Major Adjectives in English Poetry: From Wyatt to Auden", University of California Publications in English, United States, vol. 12, No. 3, 1946, pp. 305-426.
- [64] J. Miles, *Style and Proportion: The Language of Prose and Poetry*, Brown and Co, Boston, United States, 1967.
- [65] L. S. Spencer and R. S. Michael, *Chemistry for Today: General, Organic, and Biochemistry*, Thomson Brooks/Cole, United States, 2004.
- [66] L. Cohen, L. Manion, and K. Morrison, *Research Methods in Education*, 6th ed. Routledge, New York, USA, 2007.
- [67] F. Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, Verso Books, United States, 2007.
- [68] F. Moretti, *Distance Reading: the Formation of An Unorthodox Literary Critic*, Verso Books, United States, 2013.
- [69] Nature, "Big Data: Science in the Petabyte Era", *Nature Magazine*, Nature Publishing Group, Philadelphia, United States, 2008.
- [70] *Word Classes (or Parts of Speech)*, Oxford Dictionary, <https://en.oxforddictionaries.com/grammar/word-classes-or-parts-of-speech> Accessed Date, 09/12/2017.
- [71] V. Weiley and Y. Pisan, "the Distributed Studio: Towards A Theory of Virtual Place for Creative Collaboration", *20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, ACM Press, New York, 2008, pp. 343-346.
- [72] Poetry Foundation Website, <https://www.poetryfoundation.org/> Accessed Date, 09/12/2017.
- [73] Poem Hunter Website, <http://www.poemhunter.com/> Accessed Date, 09/12/2017.

- [74]Poets Website, <https://www.poets.org/> Accessed Date, 09/12/2017.
- [75]Poetry Daily Website, <http://poems.com/> Accessed Date, 09/12/2017.
- [76]Poetry Archive Website, <http://www.poetryarchive.org/> Accessed Date, 09/12/2017.
- [77]J. R. Quinlan, “Induction of Decision Trees”, *Machine Learning*, Kluwer Academic Publishers, USA, vol. 1, 1986, pp. 81-106.
- [78]R. Rosenzweig, “Scarcity or Abundance? Preserving the Past in a Digital Era”, *American Historical Review*, United States, vol. 3, 2003, pp. 735-762.
- [79]B. Franke, J. F. Plante and R. Roscher etc, *Statistical Inference, Learning and Models in Big Data*, Fields Institute, 2015.
- [80]J. R. Beniger and D. L. Robyn, “Quantitative Graphics in Statistics: A Brief History”, *the American Statistician*, Taylor & Francis Ltd, United Kingdom, vol. 32, 1978, pp. 1-11.
- [81]G. Robert, *Rainbows, Halos, and Glories*, Cambridge University Press, United Kingdom, 1980.
- [82]M. A. Runco, *Creativity: Theories and Themes: Research, Development, and Practice*, Philadelphia, Elsevier Science, United States, 2014.
- [83]E. Rutherford, “Arthur Sidgwick’s Greek Prose Composition: Gender, Affect, and Sociability in the Late-Victorian University”, *Journal of British Studies*, Cambridge University Press, United Kingdom, vol. 56, issue 1, 2017, pp. 91-116.
- [84]R. K. Sawyer, *Explaining Creativity: the Science of Human Innovation*, Oxford University Press, United Kingdom, 2012.
- [85]L. Nguyen and G. Shanks, “A Framework for Understanding Creativity in Requirements Engineering”, *Information and Software Technology*, Butterworth-Heinemann, United States, vol. 51, no. 3, 2009, pp. 655–662.
- [86]S. Z. Steven, *Chemical Principles*, Houghton Mifflin Company, New York, United States, 2009.
- [87]S. Steven, *the Algorithm Design Manual*, Springer, United States, 2011, p. 480.
- [88]S. Singh, *Using Naive Bayes for Data Clustering*, University of New Brunswick Press, Canada, 2005.

- [89] D. W. Scott, "On Optimal and Data-based Histograms", *Biometrika*, Oxford University Press, United Kingdom, vol. 66, 1979, pp. 605-610.
- [90] S. Schreibman, *A Companion to Digital Humanities (Blackwell Companions to Literature and Culture)*, Wiley-Blackwell, United States, 2008.
- [91] M. Kobayashi and K. Takeda, "Information Retrieval on The Web", *ACM Computing Surveys*, ACM Press, United States, vol. 32, 2000, pp. 144–173.
- [92] E. A. Carroll, C. Latulipe, R. Fung, and M. Terry, "Creativity Factor Evaluation: Towards A Standardised Survey Metric for Creativity Support", *7th ACM Conference on Creativity and Cognition*, ACM, New York, 2009, pp. 127–136.
- [93] Thesaurus Synonym Dictionary, <http://www.thesaurus.com/browse/synonym>
Accessed Date, 09/12/2017.
- [94] "Glucose", The Columbia Encyclopedia , <http://www.encyclopedia.com>
Accessed Date, 09/12/2017.
- [95] H. Yang and Y. Tian, "Software Engineering Creative Computing", *Conference on Computer Science and Software Engineering (CSSE)*, SCEE, Taiwan, 2010.
- [96] R. J. Trudeau, *Introduction to Graph Theory*, Dover Publisher, United States, 2012.
- [97] J. W. Turkey, "Some Graphic and Semigraphic Displays", *Statistical Papers in Honor of George W. Snedecor*, Iowa State University Press, United States, 1972.
- [98] J. W. Turkey, *Exploratory Data Analysis*, Addison Wesley Publisher, Boston, United States, 1970.
- [99] J. Leskovec, A. Rajaraman and J. D. Ullman, *Mining of Massive Datasets*, Cambridge University Press, Cambridge, United Kingdom, 2011.
- [100] UN Global Pulse, "Big Data for Development: Challenges & Opportunities", UN Global Pulse, United States, 2012.
- [101] D. J. Watts, "A Twenty-First Century Science", *Nature*, Nature Publishing Group, United Kingdom, vol. 445, 2007, pp. 489.
- [102] R. N. Wilson, *Reflecting Telescope Optics: Basic Design Theory and its Historical Development*, Springer Press, United States, 2004.

- [103] A. Hugill and H. Yang, “The Creative Turn : New Challenges for Computing”, *International Journal of Creative Computing*, Inderscience Enterprises Ltd, United Kingdom, vol. 1, no. 1, 2013, pp. 4–19.
- [104] R. Liu and H. Yang, “Chaos and Fractal for Creative Computing”, *IEEE 8th International Symposium on Service Oriented System Engineering*, IEEE, New York, 2014, pp. 470–473.
- [105] C. Plaisant, J. Rose, B. Yu *et al*, “*Exploring Erotics in Emily Dickinson’s Correspondence with Text Mining and Visual Interfaces*”, Proceedings of the 6th ACM Joint Conference on Digital Libraries, ACM, New York, USA, 2006, pp. 141-150.
- [106] J. J. Shaughnessy, E. B. Zechmeister and J. S. Zechmeister, *Research Methods in Psychology*, McGraw Hill, New York, United States, 2011.
- [107] H. Yang and L. Zhang, “Controlling and Being Creative : Software Cybernetics and Creative Computing”, *IEEE 38th International Computer Software and Applications Conference Workshop (COMPSACW)*, IEEE Press, New York, 2014, pp. 19–24.
- [108] Q. Chen, M. Zhang and X. Zhao, “Analysing Customer Behaviour in Mobile App Usage”, *Industrial Management and Data Systems*, Emerald Group Publishing, United Kingdom, vol. 117, issue 2, 2017, pp. 425-438.
- [109] D. Janssen, T. Schlegel, M. Wissen, and J. Ziegler, “MetaCharts - Using Creativity Methods in A CSCW Environment”, *Human-Computer Interaction Theory and Practice (Part II)*, CRC Press, United States, 2003, pp. 939–943.
- [110] D. Janssen, T. Schlegel, M. Wissen, and J. Ziegler, “MetaCharts - Using Creativity Methods in A CSCW Environment”, *Human-Computer Interaction Theory and Practice (Part II)*, CRC Press, United States, 2003, pp. 939–943.

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value (Top 2000 Words in Descending Order)

WORD	pH Val ue	Telling Ambiti on	Expressi ng Emotion	Criticisi ng	Commemora ting	Praisi ng	Celebrat ing
the	7	0	0	0	0	0	0
in	7	0	0	0	0	0	0
that	7	0	0	0	0	0	0
on	7	0	0	0	0	0	0
this	7	0	0	0	0	0	0
at	7	0	0	0	0	0	0
will	6	5	4	1	4	4	1
about	7	0	0	0	0	0	0
all	7	0	0	0	0	0	0
has	7	0	3	0	0	0	0
so	7	0	0	0	0	0	0
like	6	3	5	0	2	2	1
out	7	0	0	0	0	0	0
such	7	2	3	3	1	1	1
different	7	4	0	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

film	7	0	0	0	0	0	0
good	6	0	0	0	0	0	0
very	7	0	0	0	0	0	0
well	6	0	0	0	0	0	0
my	7	5	0	0	0	0	0
over	14	0	0	0	0	0	0
where	7	0	3	0	0	0	0
important	6	5	3	0	0	0	0
he	7	0	3	0	0	0	0
between	7	0	3	0	0	0	0
us	7	0	3	0	0	0	0
health	7	0	3	0	0	0	0
game	7	0	3	0	0	0	0
had	7	0	3	0	0	0	0
after	7	0	3	0	0	0	0
life	7	5	3	0	0	0	0
during	7	0	3	0	0	0	0
meat	7	0	3	0	0	0	0
number	7	0	3	0	0	0	0
start	4	0	3	0	0	0	0
once	7	0	3	0	0	0	0
human	7	3	3	0	0	0	0
without	7	0	3	0	0	0	0
general	7	0	3	0	0	0	0
specific	8	0	3	0	0	0	0
popular	6	0	3	0	0	0	0
job	7	5	3	0	0	0	1
thank	6	0	3	0	0	0	1
method	7	5	3	0	0	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

data	7	0	3	0	0	0	1
type	7	0	4	0	0	0	1
home	6	3	4	0	1	0	1
tonight	7	0	4	0	1	0	1
always	7	0	4	0	1	0	1
bird	6	5	4	0	1	0	1
literature	7	0	4	0	1	0	1
create	6	5	4	0	1	0	1
control	8	0	4	0	1	0	1
basic	7	0	4	0	1	0	1
company	7	0	4	0	1	0	1
size	7	0	4	0	1	0	1
television	7	0	4	0	1	0	1
card	7	0	4	0	1	0	1
particular	8	3	3	0	1	0	1
either	7	0	3	0	1	0	1
line	7	0	3	0	1	0	1
risk	14	0	3	0	1	0	1
word	7	0	3	0	1	0	1
simply	7	0	3	0	1	0	1
name	7	0	3	0	1	0	1
generally	7	0	3	0	1	0	1
amount	7	0	3	0	1	0	1
sense	8	0	3	0	1	0	1
instead	7	0	3	0	1	0	1
show	7	0	3	0	1	0	1
develop	6	5	3	4	1	0	1
boss	8	5	3	4	1	0	1
last	8	4	3	4	1	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

sport	7	4	3	4	1	0	1
fun	6	4	3	4	1	0	1
term	7	4	3	4	1	0	1
increase	6	4	3	4	1	0	1
sound	7	4	3	4	1	0	1
individual	7	4	3	4	1	0	1
kind	6	4	3	4	1	0	1
buy	7	0	3	4	1	0	1
later	7	0	3	4	1	0	1
potential	6	4	3	4	1	0	1
almost	7	0	3	4	1	0	1
management	7	4	3	4	1	0	1
reason	7	0	3	4	1	0	1
above	7	0	3	4	1	0	1
site	7	0	3	4	1	0	1
following	5	0	3	0	1	0	1
case	7	0	3	0	1	0	1
TRUE	6	0	3	0	0	0	1
bad	14	0	3	0	0	0	1
excuse	7	0	3	0	0	0	1
result	7	0	3	0	0	0	1
building	7	0	3	0	0	4	1
clear	7	0	3	0	0	4	1
live	8	0	3	0	0	4	1
store	7	0	4	0	0	4	1
directly	8	3	4	0	0	4	1
old	10	4	4	0	0	4	1
side	7	0	4	0	0	4	1
basis	7	0	4	0	0	4	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

primary	7	0	4	0	0	4	1
army	9	3	4	0	0	4	1
rule	7	0	4	0	0	4	1
bring	7	0	4	0	0	4	1
improve	6	5	4	0	0	4	1
produce	5	0	4	0	0	4	1
earth	7	0	4	0	0	4	1
month	7	0	4	0	0	0	1
truth	5	0	4	0	0	0	1
creative	6	3	3	0	0	0	1
marketing	7	0	3	0	0	0	1
rock	9	0	3	0	0	0	1
university	7	0	3	0	0	0	1
car	7	0	3	0	0	0	1
drive	7	0	3	0	0	0	1
late	13	1	3	0	0	0	1
support	6	4	3	0	0	0	1
entire	7	0	3	0	0	0	1
income	6	0	3	0	0	0	1
proper	7	0	3	0	0	0	1
run	7	5	3	0	0	0	1
user	7	0	3	0	0	0	1
combination	5	0	3	0	0	0	1
easily	6	0	3	0	0	0	1
half	7	0	3	0	0	0	1
outside	7	0	3	0	0	0	1
bus	7	0	3	0	0	0	1
fire	14	0	3	0	0	0	1
position	7	0	3	0	0	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

stress	14	0	3	4	0	0	1
beautiful	6	0	3	4	0	0	1
complete	7	0	3	4	0	0	1
night	7	0	3	4	0	0	1
successful	6	5	3	4	0	0	1
cycle	7	0	4	4	0	0	1
expensive	10	3	4	4	0	0	1
nation	7	0	4	4	0	0	1
review	7	0	4	4	0	0	1
screen	7	0	4	4	0	0	1
visit	7	0	4	4	0	0	1
advertising	7	0	4	4	0	0	1
avoid	10	0	4	4	0	0	1
cover	7	0	4	4	0	0	1
prepare	6	0	4	4	0	0	1
share	7	0	4	4	0	0	1
bit	7	0	4	0	0	0	1
education	7	1	3	0	0	0	1
moment	7	0	3	0	0	0	1
tool	7	0	3	0	0	0	1
address	7	0	3	0	0	0	1
career	6	5	3	0	0	0	1
excellent	6	0	3	0	0	0	1
morning	7	0	3	0	0	0	1
property	7	0	3	0	0	0	1
student	7	0	3	0	0	0	1
anything	6	0	3	0	0	0	1
credit	7	0	3	0	0	0	1
entertainmen	7	0	3	0	0	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

t							
lift	7	3	2	0	0	0	1
overall	7	0	3	0	0	0	0
responsible	7	0	3	0	0	0	0
unit	7	0	3	0	0	0	0
category	7	0	3	0	2	0	0
cute	6	0	3	0	2	0	0
finally	9	3	3	0	2	0	0
link	7	0	3	0	2	0	0
post	7	0	2	0	2	0	0
voice	7	0	2	0	2	0	0
capital	6	0	2	0	2	0	0
despite	7	0	2	0	2	0	0
gain	7	0	2	0	2	0	0
newspaper	7	0	2	0	2	0	0
relationship	7	0	2	0	2	0	0
soon	7	2	2	0	2	0	0
wonderful	6	0	2	0	2	0	0
couple	7	0	2	0	0	0	0
ensure	7	0	2	0	0	0	0
finding	7	0	2	0	0	0	0
highly	7	4	2	0	0	0	0
lake	7	0	2	0	0	0	0
message	7	0	3	0	0	0	0
relatively	7	0	3	0	0	0	0
spot	7	0	3	0	0	0	0
track	7	0	3	0	0	0	0
appear	7	0	3	0	0	0	0
click	7	0	3	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

desire	7	0	3	0	0	0	0
fairly	9	1	3	0	0	0	0
housing	7	0	3	0	0	0	0
involve	7	0	3	0	0	0	0
notice	7	0	3	0	0	0	0
sell	7	0	3	0	0	0	0
wall	7	0	3	0	0	0	0
base	7	0	3	0	0	0	0
critical	7	0	3	0	0	0	0
expression	7	0	3	0	0	0	0
immediate	7	0	3	0	0	0	0
pair	7	0	3	0	0	0	0
remain	7	0	3	0	0	0	0
secret	9	0	3	0	0	0	0
staff	7	0	3	0	0	0	0
tough	7	0	3	0	0	0	0
whose	7	0	3	0	0	0	0
author	7	0	3	0	0	0	0
city	7	0	2	0	0	2	0
discount	7	0	2	0	0	2	0
file	7	0	2	0	0	2	0
grandmother	6	0	2	0	0	2	0
lesson	7	3	2	0	0	2	0
perspective	7	0	2	0	0	2	0
recommend	7	0	2	0	0	2	0
secure	7	0	2	0	0	2	0
stick	7	0	2	0	0	2	0
touch	7	0	2	0	0	2	0
bowl	7	0	2	0	0	2	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

character	7	0	2	0	0	2	0
cry	14	0	2	0	0	2	0
evidence	7	0	2	0	0	2	0
generate	6	0	2	0	0	2	0
mostly	7	0	2	0	0	0	0
obtain	7	0	2	0	0	0	0
park	7	0	3	0	0	0	0
quarter	7	0	3	0	0	0	0
sort	7	0	3	0	0	0	0
baby	5	0	3	0	0	0	0
college	5	0	3	0	0	0	0
criticism	14	0	3	0	0	0	0
dish	7	0	3	0	0	0	0
essentially	7	0	3	0	0	0	0
fine	4	0	3	0	0	0	0
happen	7	0	3	0	0	0	0
master	6	0	3	0	2	0	0
patience	8	0	3	0	2	0	0
remote	7	0	2	0	2	0	0
strength	6	0	2	0	2	0	0
treat	7	0	2	0	2	0	0
administrati on	7	0	2	0	2	0	0
beat	7	0	2	0	2	0	0
director	7	0	2	0	2	0	0
hungry	12	0	2	0	2	0	0
log	7	0	2	0	2	0	0
net	7	0	2	0	2	0	0
principle	7	0	2	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

rarely	9	0	2	0	2	0	0
sale	7	0	2	0	2	0	0
similarly	7	0	2	0	0	0	0
storage	7	0	2	0	0	0	0
wave	7	0	2	0	0	0	0
basically	7	0	2	0	0	0	0
commission	7	0	2	0	0	0	0
contribute	6	0	2	0	0	0	0
draw	7	0	2	0	0	0	0
equal	7	0	3	0	0	0	0
frequent	7	0	3	0	0	0	0
join	7	0	3	0	0	0	0
membership	7	0	3	0	0	0	0
possession	7	0	3	0	0	0	0
react	7	0	3	0	0	0	0
sensitive	10	0	3	0	0	0	0
stuff	7	0	3	0	0	0	0
union	7	0	3	0	0	0	0
attack	14	0	3	0	0	0	0
claim	7	0	3	0	0	0	0
currency	7	0	3	0	0	0	0
drag	7	0	3	0	0	0	0
emphasize	7	0	3	0	0	0	0
essay	7	0	3	0	5	0	0
finance	7	0	3	0	5	0	0
interaction	7	0	5	0	5	0	0
minor	7	0	3	0	5	0	0
operate	7	0	3	0	5	0	0
preference	7	0	3	0	5	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

reflect	7	0	3	0	5	0	0
send	7	0	3	0	5	0	0
storm	7	0	3	0	5	0	0
virus	7	0	3	0	5	0	0
afraid	14	0	3	0	5	0	0
assume	7	0	3	0	5	0	0
catch	7	0	3	0	5	0	0
cream	7	0	3	0	5	0	0
detailed	7	0	3	0	5	0	0
election	7	0	3	0	5	0	0
FALSE	7	0	3	0	5	0	0
guidance	7	0	3	0	0	0	0
kid	6	0	3	0	0	0	0
obvious	7	0	3	0	0	0	0
pleasure	6	0	3	0	0	0	0
score	7	0	3	0	0	0	0
shop	7	0	3	0	0	0	0
suggestion	7	0	3	0	0	0	0
totally	7	0	3	0	0	0	0
wise	4	0	3	0	0	0	0
awareness	7	0	3	0	0	0	0
bread	7	0	3	0	0	0	0
climate	7	0	3	0	0	0	0
comparison	7	0	3	0	0	0	0
court	7	0	3	0	0	0	0
door	7	0	3	0	0	0	0
employee	7	0	3	0	0	0	0
finger	7	0	3	0	0	0	0
height	7	0	3	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

implement	7	0	3	0	0	0	0
layer	7	0	3	0	0	0	0
manager	7	0	5	0	2	0	0
mention	7	0	5	0	0	0	0
obviously	7	0	5	0	0	0	0
physically	7	0	5	0	0	0	0
respect	7	0	5	0	0	0	0
settle	7	0	5	0	0	0	0
swimming	7	0	5	0	0	0	0
unhappy	12	0	5	0	0	0	0
adult	7	0	5	0	0	0	0
attract	7	0	5	0	0	0	0
boring	10	0	5	0	0	0	0
code	7	0	5	0	0	0	0
dimension	7	0	5	0	0	0	0
ease	7	0	5	0	0	0	0
escape	8	0	5	0	0	0	0
extent	7	0	5	0	0	0	0
gather	7	0	5	0	0	0	0
homework	7	0	5	0	0	0	0
leader	7	0	5	0	0	0	0
mountain	7	0	5	0	0	0	0
outcome	7	0	5	0	2	0	0
pause	7	0	2	0	2	0	0
promotion	7	0	2	0	2	0	0
refrigerator	7	0	2	0	2	0	0
rough	7	0	2	0	2	0	0
session	7	0	2	0	2	0	0
stomach	7	0	2	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

succeed	6	0	2	0	2	0	0
tennis	7	0	2	0	2	0	0
vacation	7	0	2	0	2	0	0
acquire	7	0	2	0	2	0	0
anyway	7	0	2	0	2	0	0
associate	7	0	2	0	0	0	0
blow	7	0	2	0	0	0	0
brother	6	0	2	0	0	0	0
chip	7	0	2	0	0	0	0
coffee	7	0	2	0	0	0	0
deliberately	7	0	5	0	0	0	0
dust	7	0	5	0	0	0	0
floor	7	0	5	0	0	0	0
hair	7	0	5	0	0	0	0
incorporate	7	0	5	0	0	0	0
justify	7	0	5	0	0	0	0
lay	7	0	5	0	0	0	0
measurement	7	0	5	0	0	0	0
nasty	7	0	5	0	0	0	0
organize	7	0	5	0	0	0	0
poetry	7	0	5	0	0	0	0
pound	7	0	5	0	0	0	0
relation	7	0	5	0	0	0	0
retain	7	0	5	0	0	0	0
senior	7	0	5	0	0	0	0
signature	7	0	5	0	0	0	0
southern	7	0	5	0	0	0	0
swim	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

tooth	7	0	5	0	0	0	0
unfortunatel y	14	0	5	0	0	0	0
wash	7	0	5	0	0	0	0
accident	7	0	5	0	0	0	0
arrival	7	0	5	0	0	0	0
baseball	7	0	5	0	0	0	0
brick	7	0	5	0	0	0	0
clue	7	0	5	0	0	0	0
concerning	7	0	5	0	0	0	0
crash	7	0	5	0	0	0	0
devil	7	0	0	0	0	0	0
explanation	7	0	0	0	0	0	0
forth	7	0	0	0	0	0	0
girl	7	0	0	0	0	0	0
herself	7	3	3	0	2	0	0
injury	7	3	0	0	2	0	0
lunch	7	3	0	0	2	0	0
monitor	7	3	0	0	2	0	0
panic	7	3	0	0	2	0	0
plane	7	3	0	0	2	0	0
qualify	7	3	0	0	2	0	0
replacement	7	3	0	0	2	0	0
rid	7	3	0	0	2	0	0
shock	7	3	0	0	2	0	0
spare	7	3	0	0	2	0	0
suspect	7	3	0	0	2	0	0
transition	7	3	0	0	2	0	0
usual	7	3	0	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

welcome	6	3	0	0	2	0	0
abroad	7	3	0	0	2	0	0
attach	7	0	0	0	2	0	0
blind	7	0	0	0	0	0	0
buyer	7	0	0	0	0	0	0
chest	7	0	0	0	0	0	0
concentrate	7	0	0	0	0	0	0
cookie	7	0	0	0	0	0	0
desk	7	0	0	0	0	0	0
establish	7	0	0	0	0	0	0
garbage	7	0	0	0	0	0	0
honest	6	0	0	0	0	0	0
of	7	0	0	0	0	0	0
is	7	0	0	0	0	0	0
or	7	0	0	0	0	0	0
your	7	0	0	0	0	0	0
an	7	0	0	0	0	0	0
from	7	0	0	0	0	0	0
if	7	0	0	0	0	0	0
which	7	0	0	0	0	0	0
also	7	0	0	0	0	0	0
most	7	0	0	0	0	0	0
was	7	0	0	0	0	0	0
use	7	0	0	0	0	0	0
who	7	0	0	0	0	0	0
through	7	0	0	0	0	0	0
its	7	0	3	0	0	0	0
just	7	0	3	0	0	0	0
water	7	0	3	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

any	7	0	3	0	0	0	0
art	7	0	3	0	0	0	0
first	7	0	3	0	0	0	0
world	7	0	3	0	0	0	0
much	7	0	3	0	0	0	0
family	6	0	3	0	0	0	0
look	7	0	3	0	0	0	0
go	7	0	3	0	0	0	0
great	7	0	3	0	0	0	0
same	7	0	3	0	0	0	0
might	7	0	3	0	0	0	0
hi	7	0	3	0	0	0	0
computer	7	0	3	0	0	0	0
since	7	0	3	0	0	0	0
learn	6	0	0	0	0	0	0
air	7	0	0	0	0	0	0
public	7	0	0	0	0	0	0
year	7	0	0	0	0	0	0
available	7	0	0	0	0	0	0
both	7	0	0	0	0	0	0
come	7	0	0	0	3	0	0
process	7	0	0	0	3	0	0
enough	7	0	0	0	3	0	0
small	7	0	0	0	3	0	0
music	6	0	0	0	3	0	0
book	7	0	0	0	3	0	0
never	7	0	5	0	3	0	0
feel	7	0	5	0	3	0	0
whether	7	0	5	0	3	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

certain	7	0	5	0	3	0	0
law	7	0	5	0	3	0	0
body	7	0	5	0	3	0	0
guide	7	0	5	0	3	0	0
sometimes	7	0	5	0	3	0	0
simple	7	0	5	0	3	0	0
knowledge	7	0	5	0	3	0	0
course	7	0	5	0	3	0	0
known	7	0	5	0	0	0	0
away	7	0	5	0	0	0	0
three	7	0	5	0	0	0	0
difficult	9	0	5	0	0	0	0
real	7	0	5	0	0	0	0
library	7	0	5	0	0	0	0
product	7	0	5	0	0	0	0
several	7	0	5	0	0	0	0
fat	8	0	5	0	0	0	0
today	7	0	5	0	0	0	0
personal	7	0	5	0	0	0	0
historical	7	0	5	0	0	0	0
level	7	0	5	0	0	0	0
service	7	0	5	0	0	0	0
least	7	0	5	0	0	0	0
society	7	0	5	0	0	0	0
second	7	0	5	0	0	0	0
short	7	0	5	0	0	0	0
media	7	0	5	0	0	0	0
thing	7	0	5	0	0	0	0
house	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

test	7	3	5	0	3	0	0
oven	7	3	5	0	3	0	0
again	7	3	5	0	3	0	0
matter	7	3	5	0	3	0	0
quality	7	3	5	0	3	0	0
development	7	3	5	0	3	0	0
main	7	3	5	0	3	0	0
professional	7	3	5	0	3	0	0
especially	7	3	5	0	3	0	0
open	7	3	5	0	3	0	0
travel	7	3	5	0	3	0	0
according	7	3	5	0	3	0	0
alternative	7	3	5	0	3	0	0
image	7	3	5	0	3	0	0
cause	7	3	5	0	3	0	0
whole	7	3	5	0	3	0	0
boat	7	3	5	0	3	0	0
grow	7	3	5	0	3	0	0
section	7	0	5	0	0	0	0
mouse	7	0	5	0	0	0	0
dry	7	0	5	0	0	0	0
nothing	7	0	5	0	0	0	0
tax	7	0	5	0	0	0	0
full	7	0	5	0	0	0	0
policy	7	0	5	0	0	0	0
subject	7	0	5	0	0	0	0
boyfriend	7	0	5	0	0	0	0
space	7	0	5	0	0	0	0
camera	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

similar	7	0	5	0	0	0	0
chance	6	0	5	0	0	0	0
man	7	0	5	0	0	0	0
search	7	0	5	3	0	0	0
else	7	0	5	3	0	0	0
present	7	0	5	3	0	0	0
upset	7	0	5	3	0	0	0
design	7	0	5	3	0	0	0
material	7	0	5	3	0	0	0
salt	7	0	5	3	0	0	0
writing	7	0	5	3	0	0	0
cost	7	0	5	3	0	0	0
exist	7	0	5	3	0	0	0
news	7	0	5	3	0	0	0
tend	7	0	5	3	0	0	0
fishing	7	0	5	3	0	0	0
marriage	7	0	5	3	0	0	0
related	7	0	5	3	0	0	0
speed	7	0	5	3	0	0	0
war	7	0	5	0	0	0	0
complex	7	0	5	0	0	0	0
effective	7	0	5	0	0	0	0
inside	7	0	5	0	0	0	0
philosophy	7	0	5	0	0	0	0
decide	7	0	5	0	0	0	0
identify	7	0	5	0	0	0	0
pressure	7	0	5	0	0	0	0
teacher	7	0	5	0	0	0	0
benefit	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

continue	7	0	5	0	0	0	0
protect	7	0	5	0	0	0	0
unless	7	0	5	0	0	0	0
disease	7	0	5	0	0	0	0
face	7	0	5	0	0	0	0
negative	7	0	5	0	0	0	0
road	7	0	5	0	0	0	0
soup	7	0	5	0	0	0	0
visual	7	0	5	0	0	0	0
affect	7	0	5	0	0	0	0
ball	7	0	5	0	0	0	0
discipline	7	0	5	0	0	0	0
quick	7	0	5	0	0	0	0
success	6	0	5	0	0	0	0
black	8	0	5	0	0	0	0
gift	6	0	5	0	0	0	0
painting	7	0	5	0	0	0	0
walk	7	0	5	0	0	0	0
attention	7	0	5	0	0	0	0
culture	7	0	5	0	0	0	0
extra	7	0	5	0	3	0	0
pick	7	0	5	0	3	0	0
receive	7	0	5	0	3	0	0
table	7	0	5	0	3	0	0
classic	7	0	5	0	3	0	0
currently	7	0	5	0	3	0	0
final	7	0	5	0	3	0	0
mix	7	0	5	0	3	0	0
population	7	0	5	0	3	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

separate	7	0	5	0	3	0	0
western	7	0	5	0	3	0	0
cigarette	7	0	5	0	3	0	0
date	7	0	5	0	3	0	0
fly	4	0	5	0	3	0	0
official	7	0	5	0	3	0	0
recent	7	0	5	0	3	0	0
willing	7	0	5	0	0	0	0
challenge	7	0	5	0	0	0	0
driver	7	0	5	0	0	0	0
him	7	0	5	0	0	0	0
nice	6	0	5	0	0	0	0
rich	6	0	5	0	0	0	0
specifically	6	0	5	0	0	0	0
young	7	0	5	0	0	0	0
daily	7	0	5	0	0	0	0
exit	7	0	5	0	0	0	0
front	7	0	5	0	0	0	0
immediately	7	0	5	0	0	0	0
lead	7	0	5	0	0	0	0
phone	7	0	5	0	0	0	0
scene	7	0	5	0	0	0	0
summer	7	0	5	0	0	0	0
valuable	7	0	5	0	0	0	0
appearance	7	0	5	0	0	0	0
concept	7	0	5	0	0	0	0
discussion	7	0	5	0	0	0	0
fixed	7	0	5	0	0	0	0
huge	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

leading	7	0	5	0	0	0	0
primarily	7	0	5	0	0	0	0
slow	8	0	5	0	0	0	0
woman	7	0	5	0	0	0	0
blood	10	0	5	0	0	0	0
damage	14	0	5	0	0	0	0
feeling	7	0	5	0	0	0	0
importance	6	0	5	0	0	0	0
payment	7	0	5	0	0	0	0
represent	7	0	5	0	3	0	0
situation	7	0	5	0	3	0	0
statement	7	0	5	0	3	0	0
ultimately	7	0	5	0	3	0	0
widely	7	0	5	0	3	0	0
aware	7	0	5	0	3	0	0
complicated	7	0	5	0	3	0	0
display	7	0	5	0	3	0	0
flow	7	0	5	0	3	0	0
ground	7	0	5	0	3	0	0
minute	7	0	5	0	3	0	0
phase	7	0	5	0	3	0	0
reference	7	0	5	0	3	0	0
seriously	10	0	5	0	3	0	0
studio	7	0	5	0	3	0	0
trouble	14	0	5	0	3	0	0
bridge	7	0	5	0	3	0	0
chemical	7	0	5	0	3	0	0
dangerous	14	0	5	0	3	0	0
extreme	7	0	5	0	3	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

imagination	7	0	5	0	3	0	0
myself	7	0	5	0	3	0	0
occasionally	7	0	5	0	3	0	0
passion	7	0	5	0	3	0	0
resource	7	0	5	0	3	0	0
weight	7	0	5	0	3	0	0
background	7	0	5	0	3	0	0
communicat e	7	0	5	0	3	0	0
debt	7	0	5	0	3	0	0
dramatic	7	0	5	0	3	0	0
exact	7	0	5	0	3	0	0
formal	7	0	5	0	3	0	0
indicate	7	0	5	0	3	0	0
memory	7	0	5	0	3	0	0
pitch	7	0	5	0	3	0	0
secretary	7	0	5	0	3	0	0
suggest	7	0	5	0	3	0	0
trip	7	0	5	0	3	0	0
appeal	7	0	5	0	3	0	0
burn	7	0	5	0	0	0	0
equally	7	0	5	0	0	0	0
ideal	7	0	5	0	0	0	0
lost	14	0	5	0	0	0	0
party	7	0	0	0	0	0	0
print	7	0	0	0	0	0	0
recommenda tion	7	0	0	0	0	0	0
season	7	0	0	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

sleep	7	0	0	0	0	0	0
street	7	0	5	0	0	0	0
advance	7	0	5	0	0	0	0
belt	7	0	5	0	0	0	0
complaint	7	0	5	0	0	0	0
copy	7	0	5	0	0	0	0
drop	7	0	5	0	0	0	0
everybody	7	0	5	0	0	0	0
highway	7	0	5	0	0	0	0
kill	14	0	5	0	0	0	0
merely	7	0	5	0	0	0	0
preparation	7	0	5	0	0	0	0
recognize	7	0	5	0	0	0	0
sit	7	0	5	0	0	0	0
sufficient	7	0	5	0	0	0	0
unusual	12	0	5	0	0	0	0
blue	8	0	5	0	0	0	0
confidence	7	0	5	0	0	0	0
daughter	7	0	5	0	0	0	0
dream	6	0	5	0	0	0	0
employment	7	0	5	0	0	0	0
existing	7	0	5	0	0	0	0
gently	7	0	5	0	0	0	0
juice	7	0	5	0	0	0	0
mixed	7	0	5	0	0	0	0
originally	7	0	5	0	0	0	0
previous	7	0	5	0	0	0	0
region	7	0	5	0	0	0	0
significantly	6	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

substance	7	0	5	0	0	0	0
wear	7	0	5	0	0	0	0
afternoon	7	0	5	0	0	0	0
bat	7	0	5	0	0	0	0
chain	7	0	5	0	0	0	0
crew	7	0	5	0	0	0	0
device	7	0	5	0	0	0	0
engage	7	0	5	0	0	0	0
feed	7	0	5	0	0	0	0
hotel	7	0	5	0	0	0	0
mark	7	0	5	0	0	0	0
ourselves	7	0	5	0	0	0	0
priority	7	0	5	0	0	0	0
screw	7	0	5	0	0	0	0
shower	7	0	5	0	0	0	0
suit	7	0	5	0	0	0	0
twice	7	0	5	0	0	0	0
wish	7	0	5	0	0	0	0
band	7	0	5	0	0	0	0
calendar	7	0	5	0	0	0	0
coat	7	0	5	0	0	0	0
confusion	7	0	5	0	3	0	0
cup	7	0	5	0	3	0	0
east	7	0	5	0	3	0	0
employer	7	0	5	0	3	0	0
garage	7	0	5	0	3	0	0
himself	7	0	5	0	3	0	0
initial	7	0	5	0	3	0	0
leadership	7	0	5	0	3	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

manner	7	0	5	0	3	0	0
narrow	8	0	5	0	3	0	0
operation	7	0	5	0	3	0	0
profile	7	0	5	0	3	0	0
rice	7	0	5	0	3	0	0
smell	7	0	5	0	3	0	0
telephone	7	0	5	0	3	0	0
wild	7	0	5	0	3	0	0
aggressive	7	0	5	0	3	0	0
bag	7	0	5	0	0	0	0
bother	7	0	5	0	0	0	0
cousin	7	0	5	0	0	0	0
disaster	7	0	5	0	0	0	0
eastern	7	0	5	0	0	0	0
evening	7	0	5	0	0	0	0
farm	7	0	5	0	0	0	0
grade	7	0	5	0	0	0	0
horror	7	0	5	0	0	0	0
loan	7	0	5	0	0	0	0
nail	7	0	5	0	0	0	0
overcome	7	0	5	0	0	0	0
permission	7	0	5	0	0	0	0
proof	7	0	5	0	0	0	0
relief	7	0	5	0	0	0	0
sad	7	0	5	0	0	0	0
shoulder	7	0	0	0	0	0	0
strange	8	0	0	3	0	0	0
successfully	6	0	0	3	0	0	0
throw	7	0	0	3	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

virtually	7	0	0	3	0	0	0
adapt	7	0	0	3	0	0	0
argue	7	0	0	3	0	0	0
automatic	7	0	0	3	0	0	0
bonus	7	0	0	3	0	0	0
buddy	7	0	0	3	0	0	0
church	7	0	0	3	0	0	0
confirm	7	0	0	3	0	0	0
dinner	7	0	0	3	0	0	0
employ	7	0	0	3	0	0	0
former	7	0	0	3	0	0	0
hardly	7	0	0	3	0	0	0
initiative	7	0	0	3	0	0	0
knife	7	0	0	3	0	0	0
league	7	0	0	3	0	0	0
mess	7	0	0	3	0	2	0
native	7	0	5	0	0	0	0
ought	7	0	3	0	0	0	0
police	7	0	3	0	0	0	0
procedure	7	0	3	0	0	0	0
relieve	7	0	3	0	0	0	0
royal	7	0	3	0	0	0	0
shame	12	0	3	0	0	0	0
significance	6	0	3	0	0	0	0
split	7	0	3	0	0	0	0
tackle	7	0	3	0	0	0	0
town	7	0	3	0	0	0	0
upper	7	0	3	0	0	0	0
waste	10	0	3	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

airport	7	0	3	0	0	0	0
assist	7	0	3	0	0	0	0
bell	7	0	3	0	0	0	0
calculate	7	0	3	0	0	0	0
collar	7	0	5	0	0	0	0
conference	7	0	5	0	0	0	0
database	7	0	5	0	0	0	0
diet	7	0	5	0	0	0	0
extend	7	0	5	0	0	0	0
friendly	6	0	5	0	0	0	0
glove	7	0	5	0	0	0	0
hide	7	0	5	0	0	0	0
instruction	7	0	5	0	0	0	0
maintenance	7	0	5	0	0	0	0
mortgage	7	0	5	0	0	0	0
peak	7	0	5	0	0	0	0
poem	7	0	5	0	0	0	0
quote	7	0	5	0	0	0	0
resolve	7	0	5	0	0	0	0
river	7	0	5	0	0	0	0
sink	7	0	5	0	0	0	0
speech	7	0	5	0	0	0	0
sweet	7	0	5	0	0	0	0
twist	7	0	5	0	0	0	0
village	7	0	5	0	0	0	0
winner	7	0	5	0	0	0	0
alarm	7	0	5	0	0	0	0
behave	7	0	5	0	0	0	0
bottle	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

cable	7	0	5	0	0	0	0
chocolate	7	0	5	0	0	0	0
concert	7	0	5	0	0	0	0
counter	7	0	5	0	0	0	0
dirty	9	0	5	0	0	0	0
establishment	7	0	5	0	0	0	0
grand	6	0	5	0	0	0	0
honey	6	0	5	0	0	0	0
and	7	0	5	0	0	0	0
you	7	0	5	0	0	0	0
it	7	0	5	0	0	0	0
with	7	0	5	0	0	0	0
by	7	0	5	0	0	0	0
I	7	0	5	0	0	0	0
some	7	0	5	0	0	0	0
when	7	0	5	0	0	0	0
how	7	0	5	0	0	0	0
people	7	0	5	0	2	0	0
we	7	0	5	0	2	0	0
into	7	0	5	0	2	0	0
them	7	0	5	0	2	0	0
get	7	0	5	0	2	0	0
no	7	0	5	0	2	0	0
only	7	0	5	0	2	0	0
been	7	0	5	0	2	0	0
history	7	0	5	0	2	0	0
know	7	0	5	0	2	0	0
would	7	0	5	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

information	7	0	5	0	2	0	0
take	7	0	5	0	2	0	0
those	7	0	5	0	2	0	0
government	7	0	5	0	2	0	0
own	7	0	5	0	2	0	0
his	7	0	5	0	2	0	0
study	7	0	5	0	2	0	0
think	7	0	5	0	2	0	0
right	7	0	5	0	2	0	0
best	7	0	5	0	2	0	0
could	7	0	5	0	2	0	0
around	7	0	5	0	2	0	0
day	7	0	5	0	0	0	0
read	7	0	5	0	0	0	0
every	7	0	5	0	0	0	0
down	7	0	5	0	0	0	0
local	7	0	5	0	0	0	0
me	7	0	5	0	0	0	0
she	7	0	5	0	0	0	0
long	7	0	5	0	0	0	0
though	7	0	5	0	0	0	0
person	7	0	5	0	0	0	0
early	7	0	5	0	0	0	0
less	7	0	5	0	0	0	0
high	7	0	5	0	0	0	0
food	7	0	5	0	0	0	0
economy	7	0	5	0	0	0	0
put	7	0	5	0	0	0	0
common	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

provide	7	0	5	0	0	0	0
problem	7	0	5	0	0	0	0
software	7	0	5	0	0	0	0
power	7	0	5	0	0	0	0
economics	7	0	5	0	0	0	0
love	6	0	5	0	0	0	0
big	7	0	5	0	0	0	0
understand	7	0	5	0	0	0	0
including	7	2	5	0	0	0	0
science	7	2	5	0	0	0	0
likely	7	2	5	0	0	0	0
care	7	2	5	0	0	0	0
someone	7	2	5	0	0	0	0
force	7	2	5	0	0	0	0
training	7	2	5	0	0	0	0
school	7	2	5	0	0	0	0
investment	7	2	5	0	0	0	0
order	7	2	5	0	0	0	0
area	7	2	5	0	0	0	0
natural	7	2	5	0	0	0	0
try	7	2	5	0	0	0	0
useful	7	2	5	0	0	0	0
story	7	2	5	0	0	0	0
mental	7	2	5	0	0	1	0
actually	7	2	5	0	0	1	0
let	7	2	1	0	1	1	0
within	7	0	1	0	1	1	0
quite	7	0	1	0	1	1	0
community	7	0	1	0	1	1	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

safety	7	0	1	0	1	1	0
soil	7	0	1	0	1	1	0
guard	7	0	1	0	1	1	0
offer	7	0	1	0	1	1	0
rather	7	0	1	0	1	1	0
garden	7	0	1	0	1	1	0
player	7	0	1	0	1	1	0
variety	7	0	1	0	1	1	0
cook	7	0	1	0	1	1	0
demand	7	0	1	0	1	1	0
quickly	7	0	1	0	1	1	0
coast	7	0	1	0	1	0	0
action	7	0	1	0	1	0	0
country	7	0	1	0	0	0	0
movie	7	0	1	0	0	0	0
across	7	0	5	0	0	0	0
allow	7	0	5	0	0	0	0
easy	6	0	5	0	0	0	0
period	7	0	5	0	0	0	0
analysis	7	0	5	0	0	0	0
involved	7	0	5	0	0	0	0
political	7	0	5	0	0	0	0
supply	7	0	5	0	0	0	0
deal	7	0	5	0	0	0	0
strategy	7	0	5	0	0	0	0
fall	7	0	5	0	0	0	0
stock	7	0	5	0	0	0	0
environment	7	0	5	0	0	0	0
model	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

source	7	0	5	0	0	0	0
healthy	7	0	5	0	0	0	0
program	7	0	5	0	0	0	0
begin	7	0	5	0	0	0	0
feature	7	0	5	0	0	0	0
medical	7	0	5	0	0	0	0
tell	7	0	5	0	0	0	0
act	7	0	5	0	0	0	0
department	7	0	5	0	0	0	0
federal	7	0	5	0	0	0	0
object	7	0	5	0	0	0	0
thus	7	0	5	0	2	0	0
fit	7	0	5	0	2	0	0
note	7	0	5	0	2	0	0
remove	7	0	5	0	2	0	0
strong	7	0	5	0	2	0	0
actual	7	0	5	0	2	0	0
content	7	0	5	0	2	0	0
eventually	7	0	5	0	2	0	0
meaning	7	0	5	0	2	0	0
regular	7	0	5	0	2	0	0
exchange	7	0	5	0	2	0	0
independent	7	0	5	0	2	0	0
reach	7	0	5	0	2	0	0
watch	7	0	5	0	2	0	0
box	7	0	5	0	2	0	0
frame	7	0	5	0	2	0	0
require	7	0	5	0	2	0	0
active	7	0	5	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

disk	7	0	5	0	0	0	0
interested	7	0	5	0	0	0	0
occur	7	0	5	0	0	0	0
role	7	0	5	0	0	0	0
stay	7	0	5	0	0	0	0
write	7	0	5	0	0	0	0
ago	7	0	5	0	0	0	0
behind	7	0	5	0	0	0	0
location	7	0	5	0	0	0	0
ready	7	0	5	0	0	0	0
addition	7	0	5	0	0	0	0
bottom	7	0	5	0	0	0	0
impact	7	0	5	0	0	0	0
politics	7	0	5	0	0	0	0
white	7	0	5	0	0	0	0
average	7	0	5	0	0	0	0
decision	7	0	5	0	0	0	0
intelligent	6	0	5	0	0	0	0
poor	7	0	5	0	0	0	0
seem	7	0	5	0	0	0	0
task	8	0	5	0	0	0	0
competition	7	0	5	0	0	0	0
discuss	7	0	5	2	0	0	0
happy	6	0	5	2	0	0	0
network	7	0	5	2	0	0	0
president	7	0	5	2	0	0	0
square	7	0	5	2	0	0	0
yes	7	0	5	2	0	0	0
concern	7	0	5	2	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

effect	7	0	5	2	0	0	0
follow	7	0	5	2	0	0	0
opportunity	7	0	5	2	0	0	0
refer	7	0	5	2	0	0	0
born	7	0	5	2	0	0	0
comfortable	7	0	5	2	0	0	0
flat	7	0	5	2	0	0	0
length	7	0	5	2	0	0	0
prefer	7	0	5	2	0	0	0
save	7	0	5	2	0	0	0
stand	7	0	5	2	0	0	0
ahead	7	0	5	2	0	0	0
dealer	7	0	5	0	0	0	0
expect	7	0	5	0	0	0	0
function	7	0	5	0	0	0	0
impossible	6	0	5	0	0	0	0
listen	7	0	5	0	0	0	0
plant	7	0	5	0	0	0	0
serious	9	0	5	0	0	0	0
taste	7	0	5	0	0	0	0
whatever	7	0	5	0	0	0	0
association	7	0	5	0	0	0	0
correct	7	0	5	0	0	0	0
explain	7	0	5	0	0	0	0
foot	7	0	5	0	0	0	0
inflation	7	0	5	0	0	0	0
lose	9	0	5	0	0	0	0
rain	7	0	5	0	0	0	0
technical	7	2	2	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

advice	7	2	2	0	2	0	0
clean	7	2	2	0	2	0	0
distance	7	2	2	0	2	0	0
finish	10	2	2	0	2	0	0
normal	7	2	2	0	2	0	0
plus	7	2	2	0	2	0	0
responsibility	7	2	2	0	2	0	0
skill	7	2	2	0	2	0	0
sugar	6	2	2	0	2	0	0
wait	7	2	2	0	2	0	0
animal	7	2	2	0	2	0	0
brown	7	2	2	0	2	0	0
county	7	2	2	0	2	0	0
educational	7	2	2	0	2	0	0
forget	7	2	2	0	2	0	0
heart	6	2	2	0	2	0	0
near	7	0	5	0	2	0	0
photo	7	0	5	0	2	0	0
register	7	0	5	0	2	0	0
shoot	7	0	5	0	0	0	0
thin	7	0	5	0	0	0	0
vary	7	0	5	0	0	0	0
campaign	7	0	5	0	0	0	0
club	7	0	5	0	0	0	0
depression	7	2	2	0	0	0	0
fan	7	2	2	0	2	0	0
letter	7	2	2	0	2	0	0
naturally	7	2	2	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

option	7	2	2	0	2	0	0
percentage	7	2	2	0	2	0	0
select	7	2	2	0	2	0	0
accept	7	2	2	0	2	0	0
carefully	7	2	2	0	2	0	0
complain	7	2	2	0	2	0	0
depend	7	2	2	0	2	0	0
eat	7	2	2	0	2	0	0
factor	7	2	2	0	2	0	0
forward	7	2	2	0	2	0	0
joint	7	2	2	0	2	0	0
muscle	7	2	2	0	2	0	0
possibly	7	2	2	0	2	0	0
slightly	7	0	5	0	0	0	0
survive	7	0	5	0	0	0	0
vast	7	0	5	0	0	0	0
appreciate	7	0	5	0	0	0	0
chart	7	0	5	0	0	0	0
foreign	7	0	5	0	2	2	0
imagine	6	0	5	0	2	2	0
manage	7	0	5	0	2	2	0
personality	7	0	2	0	2	2	0
psychologic al	7	0	2	0	2	2	0
regularly	7	0	2	0	2	2	0
selection	7	0	2	0	2	2	0
smooth	7	0	2	0	2	2	0
suitable	7	0	2	0	2	2	0
alcohol	7	0	2	0	2	2	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

bench	7	0	2	0	2	2	0
connect	7	0	2	0	2	2	0
dark	7	0	2	0	2	2	0
effectively	7	0	2	0	2	2	0
expand	7	0	2	0	2	2	0
hire	7	0	2	0	2	2	0
literally	7	0	2	0	2	2	0
minimum	7	0	2	0	2	2	0
progress	7	0	2	0	2	2	0
relax	6	0	5	0	2	2	0
south	7	0	5	0	0	0	0
tap	7	0	5	0	0	0	0
win	7	0	5	0	0	0	0
borrow	7	0	5	0	0	0	0
consistent	7	0	5	0	0	0	0
degree	7	0	5	0	0	0	0
drink	7	0	5	0	0	0	0
enable	7	0	5	0	0	0	0
famous	7	0	5	0	0	0	0
guess	7	0	5	0	0	0	0
limit	7	0	5	0	0	0	0
mixture	7	0	5	0	0	0	0
peace	7	0	5	0	0	0	0
pull	7	0	5	0	0	0	0
republic	7	0	5	0	0	0	0
soft	7	0	5	0	0	0	0
team	7	0	5	0	0	0	0
weird	7	0	5	0	0	0	0
amazing	6	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

beach	7	0	5	0	0	0	0
classroom	7	0	5	0	0	0	0
dead	7	0	5	0	0	0	0
difficulty	14	0	5	0	0	0	0
engine	7	0	5	0	0	0	0
football	7	0	5	0	0	0	0
impress	7	0	5	0	0	0	0
match	7	0	5	0	0	0	0
owner	7	0	5	0	0	0	0
protection	7	0	5	0	0	0	0
seek	7	0	5	0	0	0	0
sing	7	0	5	0	0	0	0
tension	7	0	5	0	0	0	0
variation	7	0	5	0	0	0	0
agent	7	0	5	0	0	0	0
bath	7	0	5	0	0	0	0
candidate	7	0	5	0	0	0	0
collect	7	0	5	0	0	0	0
construction	7	0	5	0	0	0	0
dig	7	0	5	0	0	0	0
elevator	7	0	5	0	0	0	0
equivalent	7	0	5	0	0	0	0
guarantee	7	0	5	0	0	0	0
hole	7	0	5	0	0	0	0
intend	7	0	5	0	0	0	0
lecture	7	0	5	0	0	0	0
march	7	0	5	0	0	0	0
nearby	7	0	5	0	0	0	0
parking	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

proud	7	0	5	0	0	0	0
routine	7	0	5	0	0	0	0
somehow	7	0	5	0	0	0	0
tie	7	0	5	0	0	0	0
winter	7	0	5	0	0	0	0
airline	7	0	5	0	0	0	0
battle	7	0	5	0	0	0	0
brief	7	0	5	0	0	0	0
crazy	7	0	5	0	0	0	0
distinct	7	0	5	0	0	0	0
editor	7	0	5	0	0	0	0
excitement	7	0	5	0	0	0	0
feedback	7	0	5	0	0	0	0
guitar	7	0	5	0	0	0	0
horse	7	0	5	0	0	0	0
logical	7	0	5	0	0	0	0
noise	7	0	5	0	0	0	0
owe	7	0	5	0	0	0	0
phrase	7	0	5	0	0	0	0
race	7	0	5	0	0	0	0
repair	7	0	5	0	0	0	0
sand	7	0	5	0	0	0	0
sick	7	0	5	0	0	0	0
strict	7	0	5	0	0	0	0
suddenly	7	0	5	0	0	0	0
tourist	7	0	5	0	0	0	0
west	7	0	5	0	0	0	0
adjust	7	0	5	0	0	0	0
arise	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

automaticall y	7	0	5	0	0	0	0
border	7	0	5	2	0	0	0
bunch	7	0	5	2	0	0	0
civil	7	0	5	2	0	0	0
cross	7	0	5	2	0	0	0
document	7	0	5	2	0	0	0
encouraging	7	0	5	2	0	0	0
god	7	0	5	2	0	0	0
hearing	7	0	5	2	0	0	0
iron	7	0	5	2	0	0	0
lab	7	0	5	2	0	0	0
loud	7	0	5	2	0	0	0
mobile	7	0	5	2	0	0	0
opening	7	0	5	2	0	0	0
parent	7	0	5	2	0	0	0
pool	7	0	5	2	0	0	0
queen	7	0	5	2	0	0	0
request	7	0	5	2	0	0	0
salary	7	0	5	2	0	0	0
shelter	7	0	5	2	0	0	0
silver	7	0	5	0	0	0	0
strain	7	0	5	0	0	0	0
tank	7	0	5	0	0	0	0
train	7	0	5	0	0	0	0
vehicle	7	0	5	0	0	0	0
wife	7	0	5	0	0	0	0
alive	7	0	5	0	0	0	0
assumption	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

bike	7	0	5	0	0	0	0
chair	7	0	5	0	0	0	0
comment	7	0	5	0	0	0	0
consult	7	0	5	0	0	0	0
deliver	7	0	5	0	0	0	0
enthusiasm	7	0	5	0	0	0	0
farmer	7	0	5	0	0	0	0
fuel	7	0	5	0	0	0	0
grab	7	0	5	0	0	0	0
historian	7	0	5	0	0	0	0
investigate	7	0	5	0	0	0	0
manufacture	7	0	5	0	0	0	0
r							
negotiate	7	0	5	0	0	0	0
perception	7	0	5	0	0	0	0
presence	7	0	5	0	0	2	0
realistic	7	0	5	0	0	2	0
retire	7	0	5	0	0	2	0
roll	7	0	5	0	0	2	0
slip	7	0	5	0	0	2	0
spite	7	0	5	2	2	2	0
swing	7	0	5	2	2	2	0
ugly	7	0	5	2	2	2	0
warning	7	0	5	2	2	2	0
worker	7	0	5	2	2	2	0
anxious	7	0	5	2	2	2	0
bend	7	0	5	2	2	2	0
brave	7	0	5	2	2	2	0
calm	7	0	5	2	2	2	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

clerk	7	0	5	2	2	2	0
conclusion	7	0	5	2	2	2	0
courage	6	0	5	2	2	2	0
disagree	7	0	5	2	2	2	0
estimate	7	0	5	2	2	2	0
grandfather	7	0	5	2	2	2	0
ignore	10	0	5	2	2	2	0
are	7	0	5	2	2	2	0
as	7	0	5	2	2	0	0
can	7	0	5	2	2	0	0
not	7	0	5	2	2	0	0
they	7	0	5	2	2	0	0
there	7	0	5	2	2	0	0
one	7	0	5	2	2	0	0
many	7	0	5	2	2	0	0
other	7	0	5	2	2	0	0
these	7	0	5	2	2	0	0
than	7	0	5	0	0	0	0
make	7	0	5	0	0	0	0
work	7	0	5	0	0	0	0
our	7	0	5	0	0	0	0
see	7	0	5	0	0	0	0
need	7	0	5	0	0	0	0
often	7	0	5	0	0	0	0
were	7	0	5	0	0	0	0
money	7	0	5	0	0	0	0
map	7	0	5	0	0	0	0
two	7	0	5	0	0	0	0
example	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

before	7	0	5	0	0	0	0
however	7	0	5	0	0	0	0
being	7	0	5	0	0	0	0
why	7	0	5	0	0	0	0
free	7	0	5	0	0	0	0
still	7	0	5	0	0	0	0
must	7	0	5	0	0	0	0
does	7	2	5	0	0	0	0
usually	7	2	5	0	0	0	0
place	7	2	5	0	0	0	0
keep	7	2	5	0	0	0	0
field	7	2	5	0	0	0	0
give	7	2	5	0	0	0	0
sure	7	2	5	0	0	0	0
back	7	2	5	0	0	0	0
heat	7	2	5	0	0	0	0
lot	7	2	5	0	0	0	0
experience	7	2	5	0	0	0	0
really	7	2	5	0	0	0	0
reading	7	2	5	0	0	0	0
play	7	2	5	0	0	0	0
off	7	2	5	0	0	0	0
understandin g	7	2	5	0	0	0	0
little	7	2	5	0	0	0	0
under	7	2	5	0	0	0	0
market	7	2	5	0	0	0	0
change	7	2	5	0	0	0	0
say	7	2	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

state	7	0	5	0	0	0	0
radio	7	0	5	0	0	0	0
hard	7	0	5	0	0	0	0
past	7	0	5	0	0	0	0
internet	7	0	5	0	0	0	0
various	7	0	5	0	0	0	0
list	7	0	5	0	0	0	0
trade	7	0	5	0	0	0	0
nature	7	0	5	0	0	0	0
group	7	0	5	0	0	0	0
temperature	7	0	5	0	0	0	0
key	7	0	5	0	0	0	0
until	7	0	5	0	0	0	0
top	7	0	5	0	0	0	0
left	7	0	5	0	0	0	0
practice	7	0	5	0	0	0	0
cut	7	0	5	0	0	0	0
physical	7	0	5	0	0	0	0
check	7	0	5	0	0	0	0
web	7	0	5	0	2	0	0
call	7	0	5	0	2	0	0
move	7	0	5	0	2	0	0
against	7	0	5	0	2	0	0
page	7	2	2	0	2	0	0
along	7	2	2	0	2	0	0
scared	7	2	2	0	2	0	0
definition	7	2	2	0	2	0	0
turn	7	2	2	0	2	0	0
ask	7	2	2	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

hold	7	2	2	0	2	0	0
oil	7	2	2	0	2	0	0
access	7	2	2	0	2	0	0
international	7	2	2	0	2	0	0
range	7	2	2	0	2	0	0
video	7	2	2	0	2	0	0
determine	7	2	2	0	2	0	0
ever	7	2	2	0	2	0	0
special	7	2	2	0	2	0	0
probably	7	2	2	0	2	0	0
age	7	2	2	0	2	0	0
dance	7	2	2	0	2	0	0
organization	7	0	5	0	2	0	0
already	7	0	5	0	0	0	0
cash	7	0	5	0	0	0	1
emotional	6	0	5	0	0	0	1
physics	7	0	5	0	0	0	1
cold	7	0	5	0	0	0	1
itself	7	0	5	0	0	0	1
purchase	7	0	5	0	1	0	1
therefore	7	0	5	0	1	0	1
direction	7	0	5	0	1	0	1
technology	7	0	5	0	1	0	1
freedom	6	0	5	0	1	0	1
weather	7	0	5	0	1	0	1
everyone	7	0	5	0	1	0	1
necessary	6	0	5	0	1	0	1
beginning	7	0	5	0	1	0	1
instance	7	0	5	0	1	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

spend	7	0	5	0	1	0	1
chicken	7	0	5	0	1	0	1
financial	7	0	5	0	1	0	1
purpose	7	0	5	0	1	0	1
themselves	7	0	5	0	1	0	1
article	7	0	5	0	1	0	1
difference	7	0	5	0	1	0	1
goal	7	0	5	0	1	0	1
scale	7	0	5	0	1	0	1
audience	7	0	5	0	1	0	1
glad	7	0	5	0	1	0	1
perform	7	0	5	0	1	0	1
rent	7	0	5	0	1	0	1
style	7	0	5	0	1	0	1
appropriate	7	0	5	0	1	0	1
craft	7	0	5	0	1	0	1
exactly	7	0	5	0	1	0	1
medicine	7	0	5	0	1	0	1
reserve	7	0	5	0	1	0	1
eye	7	0	5	0	1	0	1
leave	7	0	5	0	1	0	1
rest	7	0	5	0	1	0	1
wide	7	0	5	0	1	0	1
charge	7	0	5	0	1	0	1
issue	7	0	5	0	1	0	1
significant	7	0	5	0	1	0	1
break	7	0	5	0	1	0	1
electrical	7	0	2	2	2	0	1
item	7	0	2	2	2	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

paint	7	0	2	2	2	0	1
room	7	0	2	2	2	0	1
structure	7	0	2	2	2	0	1
wrong	14	0	2	2	2	0	1
anyone	7	0	2	2	2	0	1
certainly	7	0	2	2	2	0	1
medium	7	0	2	2	2	0	1
report	7	0	2	2	2	0	1
apartment	7	0	2	2	2	0	1
build	7	0	2	2	2	0	1
machine	7	0	2	2	2	0	1
shape	7	0	2	2	2	0	1
wind	7	0	5	0	0	0	1
believe	7	0	5	0	0	0	1
direct	7	0	5	0	0	0	1
interesting	7	0	5	0	0	0	1
pot	7	0	5	0	0	0	1
shopping	7	0	5	0	0	0	1
unique	7	0	5	0	0	0	1
condition	7	0	5	0	0	0	1
distribution	7	0	5	0	0	0	1
hope	7	0	5	0	0	0	1
north	7	0	5	0	0	0	1
private	7	0	5	0	0	0	1
stop	7	0	5	0	0	0	1
alone	7	0	5	0	0	0	1
contain	7	0	5	0	0	0	1
extremely	7	0	5	0	0	0	1
helpful	6	0	5	0	0	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

perfect	7	0	5	0	0	0	1
solve	7	0	5	0	0	0	1
bright	7	0	5	0	0	0	1
constantly	7	0	5	0	0	0	1
flight	7	0	5	0	0	0	1
magazine	7	0	5	0	0	0	1
prevent	7	0	5	0	0	0	1
self	7	0	5	0	0	0	1
teaching	7	0	5	1	0	0	1
brush	7	0	5	1	2	2	1
debate	7	0	5	1	2	2	1
experienced	7	0	5	1	2	2	1
heavy	9	0	2	1	2	2	1
invest	7	0	2	1	2	2	1
living	7	0	2	1	2	2	1
plastic	7	0	2	1	2	2	1
slowly	12	0	2	1	2	2	1
theme	7	0	2	1	2	2	1
wing	7	1	2	1	2	2	1
brain	7	1	2	1	2	2	1
customer	7	1	2	1	2	2	1
explore	7	1	2	1	2	2	1
gas	7	1	2	1	2	2	1
influence	7	1	2	1	2	2	1
meet	7	1	2	1	2	2	1
rare	7	1	2	1	2	2	1
typical	7	1	2	1	2	2	1
afford	7	1	2	1	0	0	1
competitive	7	1	2	1	0	0	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

effort	7	1	5	0	0	0	1
fresh	7	1	5	0	0	0	1
opinion	7	1	5	0	0	0	1
press	7	1	5	0	0	0	1
ride	7	1	5	0	0	0	1
spread	7	1	5	0	0	0	1
target	7	1	5	0	0	0	1
wealth	7	1	5	0	0	0	1
application	7	1	5	0	0	0	1
budget	7	1	5	0	0	0	1
deep	7	1	5	0	0	0	1
environmental	7	1	5	0	0	0	1
foundation	7	1	5	0	0	0	1
hit	7	1	5	0	0	0	1
objective	7	1	5	0	0	0	1
recently	7	1	5	0	0	0	1
relevant	7	1	5	0	0	0	1
sky	7	1	5	0	0	0	1
title	7	1	5	0	0	0	1
accurate	7	1	5	0	0	0	1
cancel	14	1	5	2	2	2	1
collection	7	1	5	2	2	2	1
dump	7	1	5	2	2	2	1
frequently	7	1	5	2	2	2	1
lock	7	1	5	2	2	2	1
nearly	7	1	5	2	2	2	1
organized	7	1	5	2	2	2	1
plenty	7	1	5	2	2	2	1

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

setting	7	1	5	2	2	2	1
ad	7	0	5	2	2	2	1
carry	7	0	5	2	2	2	1
conflict	7	0	5	2	2	2	1
description	7	0	5	2	2	2	1
efficient	7	0	5	2	2	2	1
fair	7	0	5	2	2	2	1
fruit	7	0	5	2	2	2	1
jump	7	0	5	0	0	0	1
opposite	7	0	5	0	0	3	1
powerful	6	0	5	0	0	2	1
solution	7	0	5	2	0	2	1
total	7	0	3	2	0	2	1
vegetable	7	3	3	2	0	2	1
aspect	7	3	3	2	0	2	1
compare	7	3	3	3	0	2	1
gear	7	3	3	3	0	2	1
kitchen	7	3	3	3	0	2	1
mother	7	3	3	3	0	2	1
personally	7	3	3	3	0	2	1
psychology	7	3	3	3	0	2	1
relative	7	3	3	3	0	2	1
severe	7	3	3	3	0	2	0
somewhere	7	3	3	3	0	2	0
tree	7	3	3	3	0	2	0
anywhere	7	0	3	3	0	2	0
closed	7	0	3	3	0	2	0
consist	7	0	3	3	0	2	0
differ	7	0	5	3	0	2	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

emphasis	7	0	5	3	0	2	0
firm	7	0	5	3	0	2	0
initially	7	0	5	3	0	0	0
loss	14	0	5	2	0	0	0
numerous	7	2	5	0	0	0	0
project	7	2	5	0	0	0	0
replace	7	2	5	0	0	0	0
status	7	2	5	0	0	0	0
ticket	7	2	5	0	0	0	0
agreement	7	2	5	0	0	0	0
breakfast	7	3	5	0	0	0	0
constant	7	3	5	0	0	0	0
doctor	7	3	5	0	3	0	0
duty	9	3	5	0	3	0	0
engineering	7	3	5	0	3	0	0
father	7	3	5	0	3	0	0
hopefully	7	3	5	0	3	0	0
luck	7	3	5	0	3	0	0
mouth	7	3	5	0	3	0	0
pipe	7	3	5	0	3	0	0
pure	7	3	5	0	3	0	0
roughly	7	3	5	0	3	0	0
solid	7	2	5	0	3	0	0
tradition	7	0	5	0	3	0	0
wonder	7	0	5	0	3	0	0
annual	7	0	5	0	3	0	0
blank	7	0	5	0	3	0	0
consideration	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

delivery	7	0	3	0	0	0	3
doubt	7	0	3	0	0	0	3
enhance	7	0	3	0	0	0	3
forever	7	0	3	0	0	0	3
install	7	0	3	0	0	0	3
mission	7	0	3	0	0	0	3
pain	7	0	3	0	0	0	3
repeat	7	0	3	0	0	0	3
sex	9	0	3	0	0	0	3
slide	7	0	3	0	0	0	3
thick	7	0	3	0	0	0	3
whereas	7	0	3	0	0	0	3
anxiety	7	0	3	0	0	0	3
block	7	0	3	0	0	0	3
cap	7	0	3	0	0	0	3
combine	7	0	3	0	0	0	3
contest	7	0	3	0	0	0	3
district	7	0	3	0	0	0	2
elsewhere	7	0	3	0	0	2	2
everywhere	7	0	4	0	0	2	2
guest	7	0	3	0	0	2	2
hook	7	0	3	3	0	2	2
introduce	7	0	3	3	0	2	2
lie	7	0	3	3	0	2	2
married	6	3	3	3	3	2	2
neither	7	3	3	3	3	2	2
partner	7	3	3	3	3	2	2
recording	7	3	3	3	3	2	2
sample	7	3	3	3	3	2	2

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

spiritual	7	3	3	3	3	2	2
tip	7	3	3	3	3	2	2
absolutely	7	3	3	3	3	2	2
apart	7	3	3	3	3	2	2
bed	7	3	3	3	3	2	2
cake	7	3	3	3	3	2	2
curve	7	3	3	3	3	2	2
distribute	7	3	5	3	3	2	2
efficiency	7	3	5	0	0	2	2
expose	7	3	5	0	0	2	2
fight	7	3	5	0	0	0	2
hate	7	0	4	0	4	0	2
host	7	0	4	0	4	0	2
mistake	7	0	4	0	4	0	0
none	7	0	4	0	4	0	0
package	7	0	4	0	4	0	0
presentation	7	0	4	0	4	0	0
reasonable	7	0	4	0	4	0	0
resolution	7	0	4	0	4	0	0
scratch	7	0	4	0	4	0	0
singer	7	0	4	0	4	0	0
strike	7	0	4	0	4	0	0
suffer	7	0	4	0	4	0	0
towel	7	0	2	0	4	0	0
wheel	7	0	2	0	4	0	0
administrati ve	7	0	2	0	4	0	0
arm	7	0	2	0	4	0	0
basket	7	0	2	0	4	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

branch	7	0	2	0	0	0	0
cabinet	7	0	2	0	0	0	0
clothes	7	0	2	0	0	0	0
deeply	7	0	2	0	0	0	0
draft	7	0	2	0	0	0	0
expert	7	0	2	0	0	0	0
golf	7	0	4	0	4	0	0
hurt	14	0	4	0	4	0	0
judge	7	0	4	0	4	0	0
landscape	7	0	4	0	4	0	0
mail	7	0	4	0	4	0	0
mode	7	0	4	0	4	0	0
orange	7	0	4	0	4	0	0
pattern	7	0	4	0	4	0	0
possess	7	0	4	0	4	0	0
ratio	7	0	4	0	4	0	0
respond	7	0	4	0	4	0	0
satisfaction	7	0	4	0	4	0	0
shoe	7	0	4	0	0	0	0
somebody	7	0	4	0	0	0	0
struggle	7	0	5	0	0	0	0
terribly	7	0	5	0	0	0	0
trust	6	0	5	0	0	0	0
visible	7	0	5	0	0	0	0
yellow	7	0	5	0	0	0	0
angry	14	0	5	0	0	0	0
bake	7	0	5	0	0	0	0
blame	7	0	5	0	0	0	0
chapter	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

committee	7	0	5	0	0	0	0
conversation	7	0	5	0	0	0	0
dependent	7	0	5	0	0	0	0
error	12	0	5	0	0	0	0
fear	7	0	5	0	4	0	0
funny	5	0	5	0	4	0	0
gross	7	0	5	0	4	0	4
hospital	7	0	5	0	4	0	4
jacket	7	0	5	0	4	0	4
meal	7	0	5	0	4	0	4
nurse	7	0	5	0	4	0	4
permit	7	0	5	0	4	0	4
proposal	7	0	5	0	4	0	4
reception	7	0	5	0	4	0	4
revolution	7	0	5	0	4	0	4
row	7	0	5	0	4	0	4
son	7	0	5	0	4	0	4
spray	7	0	5	0	4	0	4
tea	7	0	5	0	4	0	4
unlikely	12	0	5	0	4	0	4
weekend	7	0	5	0	0	0	4
writer	7	0	5	0	0	0	4
arrive	7	0	5	0	0	0	4
bicycle	7	0	5	0	0	0	0
breath	7	0	5	0	0	0	0
candle	7	0	5	0	0	0	0
cloud	7	0	5	0	0	0	0
contribution	7	0	5	0	0	0	0
curious	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

downtown	7	0	5	0	0	0	0
examination	7	0	5	0	0	0	0
grocery	7	0	5	0	0	0	0
simply	7	0	5	0	0	0	0
a	7	0	5	0	0	0	0
for	7	0	5	0	0	0	0
be	7	0	5	0	0	0	0
have	7	0	5	0	0	0	0
but	7	0	5	0	0	0	0
more	7	0	5	0	0	0	0
what	7	0	5	0	0	0	0
their	7	0	5	0	0	0	0
do	7	0	5	0	0	0	0
time	7	0	5	0	0	0	0
may	7	0	5	0	4	0	0
up	7	0	5	0	4	0	0
because	7	4	5	0	4	0	0
even	7	4	5	0	4	0	0
new	7	4	5	0	4	0	0
used	7	4	5	0	4	0	0
should	7	4	5	0	4	0	0
way	7	4	5	0	4	0	0
then	7	4	5	0	4	0	0
each	7	4	5	0	4	0	0
find	7	4	5	0	4	0	0
want	7	4	5	0	4	0	0
while	7	4	5	0	4	0	0
help	7	4	5	0	4	0	0
business	7	4	5	0	4	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

another	7	4	5	0	4	0	0
few	7	4	5	0	0	0	0
too	7	4	5	0	0	0	0
system	7	4	5	0	0	0	0
her	7	4	5	0	0	0	0
now	7	4	5	0	0	0	0
form	7	0	5	0	0	0	0
become	7	0	5	0	0	0	0
part	7	0	5	0	0	0	0
large	7	0	5	0	0	0	0
fish	7	0	5	0	0	0	0
something	7	0	5	0	0	0	0
better	7	0	5	0	0	0	0
thanks	7	0	5	0	0	0	0
hand	7	0	5	4	0	0	0
include	7	0	5	4	0	0	0
although	7	0	5	4	0	0	0
end	7	0	5	4	0	0	0
able	7	0	5	4	0	0	0
point	7	0	5	4	0	0	0
here	7	0	5	4	0	0	0
theory	7	0	5	4	0	0	0
value	7	0	5	4	0	0	0
set	7	0	5	4	0	0	0
interest	7	0	5	4	0	0	0
next	7	0	5	4	0	0	0
together	7	0	5	4	0	0	0
ability	5	0	5	4	0	0	0
add	7	0	5	4	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

price	7	0	5	4	0	0	0
possible	7	0	5	0	0	0	0
yourself	7	0	5	0	0	0	0
mind	7	0	5	0	0	0	0
consider	7	0	5	0	0	0	0
fact	7	0	5	0	0	0	0
idea	7	0	5	0	0	0	0
united	7	0	5	0	0	0	0
light	7	0	5	0	0	0	0
major	7	0	5	0	0	0	0
current	7	0	5	0	0	0	0
national	7	0	5	0	0	0	0
research	7	0	5	0	0	0	0
hot	7	0	5	0	0	0	0
piece	7	0	5	0	0	0	0
choose	7	0	5	0	0	0	0
activity	7	0	5	0	0	0	0
industry	7	0	5	0	0	0	0
pay	7	0	5	0	0	0	0
far	7	0	5	0	4	0	0
remember	7	0	5	0	4	0	0
answer	7	0	5	0	4	0	0
single	7	0	5	0	4	0	0
focus	7	0	5	0	4	0	0
everything	7	0	5	0	4	0	0
board	7	0	5	0	4	0	0
language	7	0	5	0	4	0	0
picture	7	0	5	0	4	0	0
additional	7	0	5	0	4	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

lower	9	0	5	0	4	0	0
rate	7	0	5	0	4	0	0
week	7	0	5	0	4	0	0
future	7	0	5	0	4	0	0
exercise	7	0	5	0	4	0	0
working	7	0	5	0	4	0	0
security	7	0	5	0	4	0	0
among	7	0	5	0	4	0	0
exam	7	0	5	0	4	0	0
record	7	0	5	0	0	0	0
below	7	0	5	0	0	0	0
class	7	0	5	0	0	0	0
equipment	7	0	5	0	0	0	0
plan	7	0	5	0	0	0	0
commercial	7	0	5	0	0	0	0
low	7	0	5	0	0	0	0
series	7	0	5	0	0	0	0
thought	7	0	5	0	0	0	0
mean	7	0	5	0	0	0	0
worth	7	0	5	0	0	0	0
paper	7	0	5	0	0	0	0
yet	7	0	5	0	0	0	0
figure	7	0	5	0	0	0	0
positive	6	0	5	0	0	0	0
child	7	0	5	0	0	0	0
maintain	7	0	5	0	0	0	0
talk	7	0	5	0	0	0	0
close	7	0	5	0	0	0	0
head	7	0	4	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

question	7	0	4	0	0	0	0
traditional	7	0	4	0	0	0	0
birth	7	0	4	0	0	0	0
dog	7	0	4	0	0	0	0
green	7	0	4	0	0	0	0
sun	7	0	4	0	0	0	0
enjoy	7	0	4	0	0	0	0
growth	7	0	4	0	0	0	0
profit	7	0	4	0	0	0	0
return	7	0	4	0	0	0	0
throughout	7	0	4	0	0	0	0
bank	7	0	4	0	0	0	0
due	7	0	4	0	0	0	0
failure	7	0	4	0	0	0	0
middle	7	0	5	0	0	0	0
standard	7	0	5	0	0	0	0
fast	7	0	5	0	0	0	0
original	7	0	5	0	0	0	0
serve	7	0	5	0	0	0	0
advantage	7	0	5	0	0	0	0
communication	7	0	5	0	0	0	0
limited	7	0	5	0	0	0	0
step	7	0	5	0	0	0	0
chemistry	7	0	5	0	0	0	0
energy	7	0	5	0	0	0	0
metal	7	0	5	0	0	0	0
pregnant	7	0	5	0	0	0	0
safe	7	0	5	0	4	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

view	7	0	5	0	4	0	0
account	7	0	5	0	4	0	0
approach	7	0	5	0	4	0	0
concerned	7	0	5	0	4	0	0
normally	7	0	5	0	4	3	0
rise	7	0	5	0	4	3	0
balance	7	0	5	0	4	3	0
choice	7	0	5	0	4	3	0
math	7	0	5	0	4	3	0
straight	7	0	5	0	4	3	0
achieve	7	0	5	0	4	3	0
beyond	7	2	5	0	4	3	0
event	7	2	5	0	2	3	0
junior	7	2	5	0	2	3	0
pretty	7	2	5	0	2	3	0
sign	7	2	5	0	2	3	0
wood	7	2	5	4	4	0	0
contact	7	2	5	4	4	0	0
egg	7	2	5	4	4	0	0
ice	7	2	5	4	4	0	0
office	7	2	4	4	4	0	0
realize	7	2	4	4	4	0	0
teach	7	2	4	4	4	0	0
attempt	7	2	4	4	4	0	0
context	7	2	4	4	4	0	0
familiar	7	2	4	4	4	0	0
introduction	7	2	4	4	4	0	0
performance	7	2	4	3	4	0	0
star	7	2	4	3	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

broad	7	2	4	3	2	0	0
describe	7	2	4	3	2	0	0
friend	7	2	5	3	2	0	0
maybe	7	2	5	0	2	0	0
properly	7	2	5	0	2	0	0
shot	7	2	5	0	2	0	0
warm	7	2	5	0	0	0	0
cell	7	0	3	0	0	0	0
discover	7	0	3	0	0	0	0
fail	7	0	3	0	0	0	3
hello	5	0	3	0	4	0	3
lack	7	0	3	0	4	0	3
member	7	4	3	0	4	0	3
reduce	7	4	3	0	4	0	3
speak	7	4	3	0	4	0	3
towards	7	4	3	0	4	0	3
worry	7	4	3	0	4	0	3
button	7	4	5	0	4	0	3
death	14	4	5	0	4	0	3
express	7	4	5	0	4	0	3
handle	7	4	5	0	4	0	3
insurance	7	4	3	0	4	0	3
mood	7	4	5	0	4	0	3
release	7	4	5	0	4	0	3
upon	7	4	5	0	4	0	0
agree	7	4	5	0	4	0	0
completely	7	4	5	0	4	0	0
electronic	7	4	5	0	4	0	0
hear	7	0	5	0	4	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

otherwise	7	0	5	0	4	0	0
reality	7	0	5	0	0	0	0
savings	7	0	5	0	3	0	0
spring	7	0	3	0	3	0	0
text	7	0	3	0	3	0	0
whenever	7	0	3	0	3	0	0
apply	7	0	3	0	3	0	0
cheap	8	0	3	0	3	0	0
depth	7	0	3	3	3	0	0
estate	7	0	3	3	3	0	3
global	7	3	3	3	3	0	3
legal	7	3	5	3	3	0	3
officer	7	3	5	3	3	0	3
recipe	7	3	5	3	0	0	3
rely	7	3	5	3	0	0	3
stage	7	3	5	3	0	0	3
topic	7	3	5	3	0	0	3
advanced	7	3	5	3	0	0	3
capable	7	3	5	3	0	0	3
cool	7	3	5	3	0	0	3
edge	7	3	5	3	0	0	3
fully	7	3	5	3	0	0	3
maximum	7	0	5	0	0	0	0
novel	7	0	5	0	0	0	0
pack	7	0	5	0	0	0	0
push	7	0	5	0	0	0	0
skin	7	0	5	0	0	0	0
agency	7	0	5	0	0	0	0
clearly	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

connection	7	0	5	0	0	0	0
die	14	0	5	0	0	0	0
enter	7	0	5	0	0	0	0
fill	7	0	5	0	0	0	0
glass	7	0	5	0	0	0	0
kick	7	0	5	0	0	0	0
pass	7	0	5	1	0	0	0
red	5	0	5	0	0	0	0
somewhat	7	0	5	0	0	0	0
traffic	7	0	5	0	0	0	0
abuse	7	0	5	0	0	0	0
attitude	7	0	5	0	0	0	0
deposit	7	2	5	0	2	0	0
greatly	7	2	5	0	2	0	0
land	7	2	5	0	2	0	0
necessarily	6	2	5	0	2	0	0
practical	7	2	5	0	2	0	0
raise	7	2	5	0	2	0	0
response	7	2	5	0	2	0	0
signal	7	2	5	0	2	0	0
spirit	7	2	5	0	2	0	0
version	7	2	5	0	2	0	0
argument	7	2	1	0	2	0	0
closely	7	2	1	0	2	0	0
contract	7	2	1	0	2	0	0
double	7	2	1	0	2	0	0
encourage	6	2	1	0	2	0	0
fix	7	2	1	0	2	0	0
internal	7	2	1	0	2	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

mainly	7	2	1	0	2	0	0
path	7	2	1	0	2	0	0
prove	7	2	1	0	0	0	0
sea	7	0	5	0	0	0	0
steak	7	0	5	0	0	0	0
tour	7	0	5	0	0	0	0
angle	7	0	5	0	0	0	0
cancer	9	0	5	0	0	0	0
cultural	7	0	5	0	0	0	0
dot	7	0	1	0	0	0	0
earn	7	0	1	0	0	1	0
entry	7	0	1	1	0	1	0
fee	7	0	1	1	0	1	0
hour	7	0	1	1	0	1	0
milk	7	0	1	1	4	1	0
nor	7	4	1	1	4	1	0
please	6	1	1	1	4	1	0
raw	7	1	1	1	4	1	0
seat	7	1	1	1	4	1	0
stable	7	1	1	1	4	1	0
trick	7	1	5	1	4	1	0
actor	7	1	5	1	4	1	0
anticipate	7	1	2	1	4	1	2
busy	7	1	2	1	4	1	2
count	7	1	2	1	4	1	2
detail	7	1	2	0	4	0	2
drama	7	1	2	0	4	0	2
examine	7	1	2	0	4	0	2
gold	7	2	2	0	4	0	2

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

interview	7	2	2	0	2	0	2
nobody	7	2	2	0	2	0	2
participate	7	2	2	0	2	0	2
round	7	2	2	0	2	0	2
sharp	7	2	2	0	2	0	2
strip	7	2	2	0	2	0	2
tone	7	2	2	0	2	0	2
window	7	2	2	0	2	0	2
atmosphere	7	2	2	0	2	0	2
bone	7	2	2	0	2	0	2
careful	7	2	2	0	2	0	2
command	7	2	2	0	2	0	2
corner	7	2	5	0	2	0	2
divide	7	2	5	0	2	0	2
emotion	7	2	5	0	2	0	2
except	7	0	5	0	2	0	2
hang	7	0	5	0	0	0	0
hunt	7	0	5	0	0	0	0
latter	7	0	5	0	0	0	0
mall	7	0	5	0	0	0	0
meeting	7	0	5	0	0	0	0
nose	7	0	5	0	0	0	0
perfectly	7	0	5	0	0	0	0
relate	7	0	5	0	0	0	0
schedule	7	0	5	0	0	0	0
survey	7	0	5	0	0	0	0
transportation	7	0	5	0	0	0	0
acceptable	7	0	5	0	0	0	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

assure	7	0	5	0	0	0	0
bill	7	0	5	0	0	0	0
charity	7	0	5	0	4	0	0
designer	7	4	5	0	4	0	0
dress	7	4	5	0	4	0	0
emergency	7	4	5	0	4	0	0
extension	7	4	5	0	4	0	0
gap	7	4	5	0	4	0	0
holiday	7	4	5	0	4	0	0
husband	7	4	5	0	4	0	0
mom	7	4	5	0	4	0	0
occasion	7	4	5	0	4	0	0
patient	7	0	5	0	4	0	0
prior	7	0	5	1	4	0	0
reflection	7	1	5	1	0	0	0
revenue	7	1	5	1	0	0	0
sentence	7	1	5	1	0	1	0
smoke	7	1	5	1	0	1	0
string	7	1	1	1	0	1	0
surprised	7	1	1	1	0	1	0
truly	7	1	1	1	0	1	0
wine	7	1	1	1	0	1	0
altogether	7	1	1	1	0	1	0
aside	7	1	1	1	0	1	0
bet	7	1	1	1	0	1	0
breast	7	0	1	1	0	1	0
childhood	7	0	1	1	0	1	0
coach	7	0	1	1	0	1	0
definitely	7	0	1	0	0	1	0

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

drawing	7	0	1	0	0	1	0
external	7	0	1	0	0	1	0
habit	7	0	1	0	0	1	0
illegal	7	0	1	0	0	0	0
judgment	7	0	5	0	0	0	0
laugh	7	0	5	0	0	0	0
massive	7	0	5	0	0	0	0
mud	7	0	5	0	0	0	0
ordinary	7	0	5	0	0	0	0
pin	7	0	5	0	0	0	0
possibility	7	0	5	0	0	0	0
readily	7	0	5	0	0	0	0
restaurant	7	0	5	0	0	0	0
sector	7	0	5	0	0	0	0
shut	7	0	5	0	0	0	0
song	7	0	5	0	0	0	0
super	7	0	5	0	0	0	2
tight	7	0	5	0	1	0	2
unfair	7	0	5	0	0	0	2
volume	7	0	2	0	0	0	2
yours	7	0	2	0	0	0	2
appointment	7	0	2	0	0	0	2
bar	7	0	2	0	0	0	2
boy	7	4	2	0	0	0	2
closet	7	4	2	0	2	0	2
compete	7	4	2	0	0	0	2
convert	7	4	2	0	0	0	2
desperate	7	4	2	0	0	0	2
exciting	7	4	2	0	0	0	2

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

fold	7	4	2	0	0	0	0
gate	7	4	2	0	0	0	0
hall	7	4	2	0	0	0	0
ill	7	4	5	0	0	0	2
lucky	6	4	5	0	0	0	2
miss	7	4	5	0	2	0	2
pace	7	4	5	0	2	0	2
pie	7	2	5	0	2	0	2
provided	7	2	5	0	2	0	2
recover	7	2	5	0	2	1	2
reward	7	2	5	0	2	1	2
sandwich	7	2	5	2	2	1	2
sorry	8	2	5	2	2	1	2
surprise	6	1	5	2	2	1	2
till	7	1	5	1	2	1	2
upstairs	7	1	5	2	2	1	2
weigh	7	1	5	2	2	1	2
yard	7	1	5	2	2	1	2
assistance	7	1	5	2	2	1	2
bite	7	1	1	2	2	1	2
briefly	7	1	1	2	2	0	2
celebrate	7	2	1	2	2	0	2
comprehensive	8	2	1	2	2	0	2
convince	7	0	1	0	2	0	2
dad	7	0	1	0	2	0	2
drawer	7	0	1	0	2	0	2
flower	6	0	5	0	2	0	2
harm	9	0	5	0	2	0	2

Appendix A. List of the Most Commonly Used English Words With pH and Theme Value

impression	7	0	5	3	2	0	2
------------	---	---	---	---	---	---	---

Appendix B

Selected C# Code of Possible Creative Connection Function in Poetrypedia

```
<%@ Page Language="C#" AutoEventWireup="true"
CodeBehind="AnalysisbyAspects.aspx.cs" Inherits="PoemS.AnalysisbyAspects1" %>

<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/xhtml">
<head runat="server">
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
    <meta name="viewport" content="width=device-width, initial-scale=1.0,
minimum-scale=1.0, maximum-scale=1.0, user-scalable=0" />
    <title>Poetrypedia</title>

    <!--#include file="Include-Js.aspx" -->
    <!-- Forms -->
    <script type="text/javascript"
src="plugins/uniform/jquery.uniform.min.js"></script> <!-- Styled radio and
checkboxes -->
    <script type="text/javascript" src="plugins/select2/select2.min.js"></script> <!--
Styled select boxes -->

    <!-- Form Validation -->
    <script type="text/javascript"
src="plugins/validation/jquery.validate.min.js"></script>
    <script type="text/javascript"
src="plugins/validation/additional-methods.min.js"></script>
```

```
<!-- DataTables -->
<script type="text/javascript"
src="plugins/datatables/jquery.dataTables.min.js"></script>
<script type="text/javascript"
src="plugins/datatables/tabletools/TableTools.min.js"></script> <!-- optional -->
<script type="text/javascript"
src="plugins/datatables/colvis/ColVis.min.js"></script> <!-- optional -->
<script type="text/javascript"
src="plugins/datatables/DT_bootstrap.js"></script>

<!-- Noty -->
<script type="text/javascript" src="plugins/noty/jquery.noty.js"></script>
<script type="text/javascript" src="plugins/noty/layouts/top.js"></script>
<script type="text/javascript" src="plugins/noty/themes/default.js"></script>

</head>

<body>

<!-- Header -->
<header class="header navbar navbar-fixed-top" role="banner">

<!--#include file="Include-Top.aspx" -->

</header> <!-- /.header -->

<div id="container">
<div id="sidebar" class="sidebar-fixed">
<div id="sidebar-content">

<!--#include file="Include-Menu.aspx" -->
```

```

<div id="content">
  <div class="container">
    <!-- Breadcrumbs line -->
    <div class="crumbs">
      <ul id="breadcrumbs" class="breadcrumb">
        <li>
          <i class="icon-home"></i>
          <a href="index.aspx">Home</a>
        </li>
        <li class="current">
          <a href="Appreciation.aspx"
title="">Appreciation</a>
        </li>
        <li class="current">
          <a href="PossibleCreativeConnections.aspx"
title="">Possible Creative Connections</a>
        </li>
      </ul>

    </div>
    <!-- /Breadcrumbs line -->

    <!--=== Page Header ===-->
    <div class="page-header">
      <div class="page-title">
        <h3>Possible Creative Connections</h3>
        <span> This module provides some relative songs to the
input poem.</span>
      </div>

    </div>
  </div>

```

```

<!-- /Page Header -->

<!--==== Page Content =====>
<button style="display:none" id="btnWarning" class="btn btn-warning
btn-notification" data-layout="top" data-type="warning"
data-text="<strong>Calculating</strong><br/>Please wait
patiently.....">Information</button>

<!--==== Title =====>
<div class="row">
<div class="col-md-12">
<h5 class="widget-title">
<i class="icon-book"></i> Select a poem
</h5>
</div>
</div>
<!-- /Title -->

<div class="row">

<!--#include file="Include-smalltable.aspx" -->

<div class="col-md-6">
<div class="widget box">
<div class="widget-header">
<h4><i class="icon-hand-right"></i>Input Poem</h4>
</div>
<div class="widget-content">

<form id="form8" class="form-horizontal row-border"
action="PossibleCreativeConnectionsResults.aspx" runat="server" method="post">
<div class="form-group">

```

```

        <label class="col-md-2
control-label">Title:</label>
        <div class="col-md-8">
            <asp:TextBox id="Title3" runat="server"
class="form-control"></asp:TextBox>
        </div>
    </div>
    <div class="form-group">
        <label class="col-md-2
control-label">Poet:</label>
        <div class="col-md-6">
            <asp:TextBox id="Poet3" runat="server"
class="form-control"></asp:TextBox>
        </div>
    </div>
    <div class="form-group">
        <label class="col-md-2
control-label">Verse:</label>
        <div class="col-md-10">
            <asp:TextBox id="Verse3" runat="server"
rows="10" placeholder="Select a poem or input the content please" name="textarea"
class="form-control" TextMode="MultiLine"></asp:TextBox>
        </div>
    </div>

    <div class="form-actions">
        <input type="button" onclick="validate()"
value="Get Related Songs" class="btn btn-primary pull-right btn-block"></input>

    </div>

</form>

```

```

        </div>
    </div>
</div>
</div>

<script language="javascript">

    $(document).ready(function () {
        $('.btn-notification').click(function () {
            var self = $(this);

            noty({
                text: self.data('text'),
                type: self.data('type'),
                layout: self.data('layout'),
                timeout: 20000,
                modal: self.data('modal'),
                buttons: (self.data('type') != 'confirm') ? false : [
                    {
                        addClass: 'btn btn-primary', text: 'Ok', onClick:
function ($noty) {
                            $noty.close();
                            noty({ force: true, text: 'You clicked "Ok" button',
type: 'success', layout: self.data('layout') });
                        }
                    }, {
                        addClass: 'btn btn-danger', text: 'Cancel', onClick:
function ($noty) {
                            $noty.close();
                            noty({ force: true, text: 'You clicked "Cancel"
button', type: 'error', layout: self.data('layout') });
                        }
                    }
                ]
            });
        });
    });
}
}

```

```

        ]
    });

    return false;
});
});

function postPoem(target)
{
    var poemId = target.parents("tr").children("td").attr("alt");
    var title = target.parents("tr").children("td:eq(2)").html();
    var poet = target.parents("tr").children("td:eq(1)").html();
    $.get("Fetch-Poems.aspx", {id:poemId, poet:poet}, function (result) {
        $("#Title3").val(title);
        $("#Poet3").val(poet);
        $("#Verse3").val(result);
        $("#form8").valid();
    });
}

function validate()
{
    var form = $("#form8");
    if (form.valid()) { $("#btnWarning").click(); form[0].submit(); }
}

</script>

</body>
</html>
namespace PoemS {

```

```
public partial class PossibleCreativeConnections1 {

    /// <summary>
    /// form9 control.
    /// </summary>
    /// <remarks>
    /// Auto-generated field.
    /// To modify move field declaration from designer file to code-behind file.
    /// </remarks>
    protected global::System.Web.UI.HtmlControls.HtmlForm form9;

    /// <summary>
    /// pccresults control.
    /// </summary>
    /// <remarks>
    /// Auto-generated field.
    /// To modify move field declaration from designer file to code-behind file.
    /// </remarks>
    protected global::System.Web.UI.WebControls.TextBox pccresults;
}
}
```

Appendix C

Matlab Code of Rainbow Spectrum

```
map = zeros(1024,3);
map(1,:) = [1 1 1];
map(1024,:) = [1 1 1];
r = round(1024*1.5/32);
y = round(1024*(3+2)/32);
g = round(1024*(3+4+3)/32);
c = round(1024*(3+4+6+2.5)/32);
b = round(1024*(3+4+6+5+6.5)/32);
m = round(1024*(3+4+6+5+13+0.5)/32);
I = [1,r,y,g,c,b,m,1024];
T = [1,1,1,0,0,0,1,1; 1,0,1,1,1,0,0,1; 1,0,0,0,1,1,1,1];
for i=2:8
    map(I(i-1):I(i), :) = [(linspace(T(1,i-1), T(1,i), numel(I(i-1):I(i))))',
(linspace(T(2,i-1), T(2,i), numel(I(i-1):I(i))))', (linspace(T(3,i-1), T(3,i),
numel(I(i-1):I(i))))');
end;
figure(1);
image(1:1024);
colormap(map);
```

Appendix D

Matlab Code of Curves and Shapes

```
% Parabola
x = -2:0.01:2;
k = 7-5.0044;
y = k*x.^2;
figure(1);
plot(x,y,'LineWidth',2,'Color','red')
axis equal;
```

```
% 'Victory'
x = -2:0.01:2;
y = [2:-0.01:0, 0.01:0.01:2];
figure(1);
plot(x,y,'LineWidth',1,'Color','blue')
axis equal;
```

```
% 'Rising'
x = -2:0.01:2;
y = x;
figure(1);
plot(x,y,'LineWidth',1,'Color','blue')
axis equal;
```

```
% 'Sinusoidal'
```

```
x = -2:0.01:2;  
y = sin(x);  
figure(1);  
plot(x,y,'LineWidth',1,'Color','blue')  
axis equal;
```

```
% 'Hyperbola'
```

```
x = 0.2:0.01:5;  
y = 1./x;  
hold on;  
plot(x,y,'linewidth',2,'color','blue');  
axis equal;
```

```
X = -5:0.01:-0.2;
```

```
Y = 1./X;  
plot(X,Y,'linewidth',2,'color','blue');
```

```
% 'Exponential'
```

```
x = -3:0.01:3;  
y = 2.^x;  
figure(1);  
plot(x,y,'LineWidth',1,'Color','blue')  
axis equal;
```

```
% 'Logarithmic'
```

```
x = 0.5:0.01:5;  
y = log(x);  
figure(1);  
plot(x,y,'LineWidth',1,'Color','blue')  
axis equal;
```

```
% 'Helix'
```

```
t = linspace(0,20*pi,1000);
```

```
x = t.*cos(t);
```

```
y = t.*sin(t)
```

```
plot(x,y, 'linewidth',1);
```

```
axis equal;
```

```
% 'Descending'
```

```
x = -2:0.01:2;
```

```
y = -x;
```

```
figure(1);
```

```
plot(x,y,'LineWidth',1,'Color','blue')
```

```
axis equal;
```

```
% 'Downwards Triangle'
```

```
figure(1);
```

```
hold on;
```

```
quiver(0,0, 1/sqrt(5), 2/sqrt(5), sqrt(5), '.b');
```

```
quiver(1,2, -1, 0, 2, '.b');
```

```
quiver(-1,2, 1/sqrt(5), -2/sqrt(5), sqrt(5), '.b');
```

```
axis equal;
```

```
% 'Rhombus'
```

```
figure(1);
```

```
hold on;
```

```
quiver(0,0, 1/sqrt(5), 2/sqrt(5), sqrt(5), '.b');
```

```
quiver(1,2, -1/sqrt(5), 2/sqrt(5), sqrt(5), '.b');
```

```
quiver(0,4, -1/sqrt(5), -2/sqrt(5), sqrt(5), '.b');
```

```
quiver(-1,2, 1/sqrt(5), -2/sqrt(5), sqrt(5), '.b');
```

```
axis equal;
```

```
% 'Square'
figure(1);
hold on;
quiver(0,0, 1, 0, 2, '.b');
quiver(2,0, 0, 1, 2, '.b');
quiver(2,2, -1, 0, 2, '.b');
quiver(0,2, 0, -1, 2, '.b');
axis equal;

% 'Circle'
figure(1);
x = -1:1:1;
y = [sqrt(1-x.^2) ; -sqrt(1-x.^2)];

plot(x,y);
axis equal;
```