โครงข่ายประสาทเทียมแบบเท็กซ์มายนิ่งไฟล์อาร์ทแม็พ

TEXT SIMPLIFIED ARTMAP NEURAL NETWORK

พรพกา คุณาธิวุฒิ

PUANGPAKA KUNASIT

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาเทคโนโลยีสารสนเทศ

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2548

ISBN 974-15-2010-7

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

TEXT SIMPLIFIED ARTMAP NEURAL NETWORK

T060905

พวงผกา คุณาสิทธิ์

PUANGPAKA KUNASIT

เลขหมู่.....
เลขทะเบียน.....60905
วัน,เดือน,ปี.....6 ก.ค. 2549

.b.....
.i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
บัณฑิตวิทยาลัย
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2548

ISBN 974-15-2010-7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TEXT SIMPLIFIED ARTMAP NEURAL NETWORK

PUANGPAKA KUNASIT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2005

ISBN 974-15-2010-7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2005

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ
นักศึกษา	นางพวงผกา คุณาสิทธิ์
รหัสประจำตัว	45066009
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
พ.ศ.	2548
อาจารย์ผู้ควบคุมวิทยานิพนธ์	ผศ.ดร.วรพจน์ กรีสุระเดช

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้ นำเสนอโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ โดยมีวัตถุประสงค์เพื่อประยุกต์การทำงานของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ ซึ่งมีการเรียนรู้แบบมีการชี้นำและรับข้อมูลที่มีค่าเป็นตัวเลขได้เท่านั้น ให้สามารถรับข้อมูลที่เป็นข้อความได้โดยตรง โดยได้นำแนวความคิดของการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็คมาประยุกต์ใช้ นอกจากนี้ยังได้นำหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี ที่เรียนรู้แบบไม่มีการชี้นำมาใช้ ซึ่งโมเดลนี้สามารถรับข้อมูลที่เป็นค่าเชิงคุณภาพได้โดยตรง โดยไม่ผ่านขั้นตอนการแปลงข้อมูลจากค่าเชิงคุณภาพไปเป็นค่าเชิงตัวเลข เนื่องจากการแปลงข้อมูลเป็นค่าเชิงตัวเลขทำให้สูญเสียเวลาในขั้นตอนการแปลงข้อมูล การทำงานของโครงข่ายประสาทเทียมที่นำเสนอนี้ สามารถจำแนกประเภทเอกสารได้อย่างถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Text Simplified ARTMAP Neural Network
Student	Mrs.Puangpaka Kunasit
Student ID.	45066009
Degree	Master of Science
Programme	Information Technology
Year	2005
Thesis Advisor	Asst.Prof.Dr.Worapoj Kreesuradej

ABSTRACT

This thesis proposes text simplified ARTMAP neural network. The proposed neural network, based on the concepts of simplified fuzzy ARTMAP neural network, text adaptive resonance theory neural network, and similarity measure for symbolic objects, works directly on textual information without mapping text data into some representation which has quantitative features. This algorithm can receive qualitative value directly without mapping into numerical value. The proposed neural network can assign class labels to the text data correctly.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้อย่างดี ด้วยการให้คำปรึกษาพร้อมทั้งคำชี้แนะแนวทางการแก้ปัญหาของงานวิจัยตั้งแต่เริ่มต้นจนเสร็จสมบูรณ์จาก ผศ.ดร.วรพจน์ กรีสุระเดช ซึ่งเป็นอาจารย์ผู้ควบคุมวิทยานิพนธ์ ผู้วิจัยรู้สึกซาบซึ้งในความอนุเคราะห์ของท่านและขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณบิดา-มารดา ที่คอยให้กำลังใจและให้คำปรึกษามาโดยตลอด

ขอขอบพระคุณอาจารย์ทินกร คุณาสิทธิ์ ที่คอยให้กำลังใจ ให้คำปรึกษาและสนับสนุนค่าใช้จ่ายและอุปกรณ์การเรียน

ขอขอบพระคุณ รศ.ดร.อาริต ธรรมโน ที่ให้คำแนะนำต่างๆ ที่เป็นประโยชน์ในการทำวิจัย

ขอขอบคุณคณาจารย์คณะเทคโนโลยีสารสนเทศทุกท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ต่างๆ เพื่อนำไปใช้ประโยชน์ในการทำวิจัย

ขอขอบคุณเจ้าหน้าที่คณะเทคโนโลยีสารสนเทศทุกท่าน ที่ให้คอยให้ความสะดวกในการทำงาน

ขอขอบคุณอาจารย์รัชดาภรณ์ อมรชีวิน ที่คอยให้กำลังใจและให้คำปรึกษา

ขอขอบคุณพี่ๆ ปริญญาเอกรุ่นที่ 1 ทุกคนที่ให้กำลังใจและคำแนะนำ

ขอขอบคุณเพื่อนๆ ห้องปฏิบัติการ Data Mining & Data Exploration Laboratory และห้องปฏิบัติการ Computational Intelligence Laboratory ที่คอยให้ความช่วยเหลือ คำแนะนำ และให้กำลังใจซึ่งกันและกันตลอดมา

ขอขอบคุณมหาวิทยาลัยราชภัฏมหาสารคาม ที่ให้ทุนสนับสนุนในการเรียนระดับปริญญาโทแก่ผู้วิจัย

ขอขอบคุณบัณฑิตวิทยาลัย ที่ได้ให้ทุนอุดหนุนการทำวิทยานิพนธ์แก่ผู้วิจัย

อนึ่ง งานวิจัยที่นำเสนอในงานวิทยานิพนธ์ฉบับนี้นั้น ส่วนหนึ่งกระทำภายใต้ห้องปฏิบัติการ Data Mining & Data Exploration Laboratory คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

คุณค่าและประโยชน์จากวิทยานิพนธ์นี้ ผู้วิจัยขอมอบแด่ผู้มีพระคุณทุกท่าน

พวงผกา คุณาสิทธิ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VIII

สารบัญรูป XI

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย 3

1.4 ขอบเขตของการวิจัย 3

1.5 ขั้นตอนการศึกษา 3

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 5

2.1 การเตรียมเอกสาร (Document Preprocessing) 5

2.2 การทำดัชนี (Indexing) 5

2.2.1 Boolean Weighting 6

2.2.2 Word Frequency Weighting 6

2.2.3 TF x IDF-Weighting 6

2.2.4 TFC-Weighting 6

2.2.5 LTC-Weighting 7

2.3 โครงข่ายประสาทเทียม (Artificial Neural Network) 7

2.4 โครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ (Simplified Fuzzy ARTMAP Neural Network) 8

2.5 โครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี (A Text Adaptive Resonance Theory Neural Network) 13

2.6 การเปรียบเทียบความคล้ายคลึงกันของเอกสาร (Similarity Measure) 16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

บทที่ 3 โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ 20

3.1 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ 20

3.1.1 หลักการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์-อาร์ทแม็พ 22

3.1.2 ตัวอย่างการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์-อาร์ทแม็พ 25

บทที่ 4 การทดลองและผลการทดลอง 38

4.1 การวัดประสิทธิภาพการทำงานของโมเดล 38

4.1.1 ตัววัดประสิทธิภาพอัตราความถูกต้อง (Accuracy Rate) 38

4.1.2 ตัววัดประสิทธิภาพ F-Measure 38

4.1.3 ตัววัดประสิทธิภาพ Entropy 40

4.2 ชุดข้อมูลที่ใช้ในการทดลอง 40

4.3 ชุดข้อมูลตัวอักษรที่ใช้ในการทดลองและผลการทดลอง 41

4.3.1 ชุดข้อมูลที่สร้างจากกลุ่มตัวอักษร(Synthesized Alphabet Document) 41

4.3.2 ผลการทดลองของชุดข้อมูลตัวอักษร 43

4.4 ชุดข้อมูล CSTR (CSTR Dataset) ที่ใช้ในการทดลองและผลการทดลอง 44

4.4.1 ชุดข้อมูล CSTR ที่ใช้ในการทดลอง 44

4.4.2 ผลการทดลองของชุดข้อมูล CSTR 46

4.4.3 ผลการเปรียบเทียบกับโมเดลต่างๆ ของชุดข้อมูล CSTR 52

4.5 ชุดข้อมูล Reuters-Top10 (Reuters-Top10 Dataset) ที่ใช้ในการทดลองและผลการทดลอง 54

4.5.1 ชุดข้อมูล Reuters-Top10 ที่ใช้ในการทดลอง 54

4.5.2 ผลการทดลองของชุดข้อมูล Reuters-Top10 55

4.5.3 ผลการเปรียบเทียบกับโมเดลต่างๆ ของชุดข้อมูล Reuter-Top10 60

4.6 ชุดข้อมูล K-dataset ที่ใช้ในการทดลองและผลการทดลอง 61

4.6.1 ชุดข้อมูล K-dataset ที่ใช้ในการทดลอง 61

4.6.2 ผลการทดลองของชุดข้อมูล K-dataset 62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.6.3 ผลการเปรียบเทียบกับโมเดลต่างๆ ของชุดข้อมูล K-dataset 67

4.7 ผลการเปรียบเทียบกับโมเดลต่างๆ ของทุกชุดข้อมูล 68

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 71

5.1 สรุปผลการวิจัย 71

5.2 ข้อเสนอแนะ 72

เอกสารอ้างอิง 74

ภาคผนวก 76

ภาคผนวก ก. แสดงตัวอย่างชุดข้อมูลต่างๆ ที่ใช้ในการทดลอง 77

ก.1 ชุดข้อมูล CSTR 77

ก.1.1 ไฟล์เอกสาร CSTR ในรูปแบบ HTML 77

ก.1.2 ไฟล์เอกสาร CSTR ที่ตัดเอาเฉพาะข้อมูล Title และ Keyword 78

ก.1.3 ข้อมูล Title นำไปตัด Stopwords และ ข้อมูล Keyword นำไปหาคำสำคัญ ด้วยโปรแกรม Copernic Summarizer 79

ก.1.4 นำข้อมูลทั้ง Title และ Keyword ไปหา Word Stemming 79

ก.2 ชุดข้อมูล Reuters-Top10 80

ก.2.1 ไฟล์เอกสาร Reuters-Top10 ในรูปแบบ XML 80

ก.2.2 ไฟล์เอกสาร Reuters-Top10 ที่ตัดเอาเฉพาะข้อมูล Title และ Keyword 81

ก.3 ชุดข้อมูล K-dataset 82

ก.3.1 ไฟล์เอกสาร K-dataset ในรูปแบบ HTML 82

ก.3.2 ไฟล์เอกสาร K-dataset ที่ตัดเอาเฉพาะข้อมูล Title และ Keyword 84

ภาคผนวก ข. แสดงตัวอย่างของคำสำคัญที่ได้จากการทดลองของชุดข้อมูลต่างๆ 85

ข.1 คำสำคัญที่ได้จากชุดข้อมูล CSTR 85

ข.2 คำสำคัญที่ได้จากชุดข้อมูล Reuters-Top10 92

ข.3 คำสำคัญที่ได้จากชุดข้อมูล K-dataset 114

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

ภาคผนวก ค. ผลงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์และได้รับการตีพิมพ์ 158

ประวัติผู้เขียน 175

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
2.1	แสดงตัวอย่างข้อมูลเอกสาร	18
3.1	แสดงตัวอย่างของข้อมูล ที่นำมาใช้ในการเรียนรู้ของโครงข่ายประสาทเทียมนี้	26
4.1	แสดงจำนวนสมาชิกแต่ละ Class ของข้อมูลตัวอักษร ที่นำมาใช้ในการเรียนรู้	42
4.2	แสดงตัวอย่างของข้อมูลตัวอักษรที่สร้างจากข้อมูลแสดงลักษณะเฉพาะ รูปที่ 4.2 และรูปที่ 4.3 ที่นำมาใช้ในการเรียนรู้	43
4.3	แสดงจำนวนชุดข้อมูลตัวอักษรที่นำมาใช้ทดสอบความถูกต้องของโมเดล	43
4.4	แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูลตัวอักษร	44
4.5	แสดงจำนวนข้อมูล CSTR แต่ละกลุ่มเอกสาร	45
4.6	แสดงจำนวนข้อมูล CSTR ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร	45
4.6	(ต่อ) แสดงจำนวนข้อมูล CSTR ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร	46
4.7	แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล CSTR ทุกชุด	46
4.8	แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ทุกชุด	47
4.9	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 1	47
4.9	(ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 1	48
4.10	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 2	48
4.11	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 3	48
4.11	(ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 3	49
4.12	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 4	49
4.13	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 5	49
4.14	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 6	50
4.15	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 7	50
4.16	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 8	50
4.16	(ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 8	51
4.17	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 9	51
4.18	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 10	51

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่		หน้า
4.19	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 11	52
4.20	แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดล ของชุดข้อมูล CSTR	53
4.21	แสดงจำนวนข้อมูล Reuters-Top10 แต่ละกลุ่มเอกสาร	54
4.22	แสดงจำนวนข้อมูล Reuters-Top10 ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร	55
4.23	แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล Reuters-Top10 ทุกชุด	55
4.23	(ต่อ) แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล Reuters-Top10 ทุกชุด	56
4.24	แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ทุกชุด	56
4.25	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 1	56
4.25	(ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 1	57
4.26	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 2	57
4.27	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 3	58
4.28	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 4	58
4.28	(ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 4	59
4.29	แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 5	59
4.30	แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดล ของชุดข้อมูล Reuters-Top10	60
4.31	แสดงจำนวนข้อมูล K-dataset ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร	61
4.31	(ต่อ) แสดงจำนวนข้อมูล K-dataset ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร	62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.32 แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล K-dataset ทุกชุด 62

4.33 แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดลของชุดข้อมูล K-dataset ทุกชุด 63

4.34 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 1 64

4.35 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 2 64

4.35 (ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 2 65

4.36 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 3 65

4.36 (ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 3 66

4.37 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 4 66

4.37 (ต่อ) แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 4 67

4.38 แสดงผลการเปรียบเทียบการทำงานแต่ละ โมเดล ของชุดข้อมูล K-dataset 68

4.39 แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดลของทุกชุดข้อมูล 69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
2.1	แสดงโครงสร้างของเซลล์ประสาท	8
2.2	สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ	9
2.3	อัลกอริทึมในการเรียนรู้ของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ	11
2.3	(ต่อ) อัลกอริทึมในการเรียนรู้ของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่-อาร์ทแม็พ	12
2.3	(ต่อ) อัลกอริทึมในการเรียนรู้ของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่-อาร์ทแม็พ	13
2.4	สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี	14
2.5	อัลกอริทึมการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี	14
2.5	(ต่อ) อัลกอริทึมการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพ-เรโซแน้นเทียรี	15
2.5	(ต่อ) อัลกอริทึมการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพ-เรโซแน้นเทียรี	16
3.1	สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ	20
3.2	อัลกอริทึมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ	23
3.2	(ต่อ) อัลกอริทึมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ	24
3.2	(ต่อ) อัลกอริทึมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ	25
3.3	แสดงกลุ่มของคำที่นำไปสร้างข้อมูลในคุณสมบัติชื่อเรื่อง	25
3.4	แสดงกลุ่มของคำที่นำไปสร้างข้อมูลในคุณสมบัติคำสำคัญ	26
3.5	แสดงค่า Top-down Weight ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พหลังจากการเรียนรู้	37
4.1	แสดงส่วนที่ใช้วัดค่า Precision และ Recall ในการร้องขอสารสนเทศ	39
4.2	แสดงกลุ่มของตัวอักษรที่นำไปสร้างข้อมูลในคุณสมบัติชื่อเรื่อง	42
4.3	แสดงแสดงกลุ่มของตัวอักษรที่นำไปสร้างข้อมูลในคุณสมบัติคำสำคัญ	42
4.4	กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของชุดข้อมูล CSTR	53
4.5	กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของชุดข้อมูล Reuters-Top10	60
4.6	กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของชุดข้อมูล K-dataset	68
4.7	กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของทุกชุดข้อมูล	69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันงานทุกอย่างจำเป็นต้องใช้ข้อมูล เมื่อวิทยาการต่างๆ มีความเจริญก้าวหน้ามากขึ้น ข้อมูลที่ใช้กับงานต่างๆ เหล่านี้จึงมีการเปลี่ยนแปลงไปเช่นเดียวกันและข้อมูลยังมีหลากหลายรูปแบบ เช่น ข้อมูลอิเล็กทรอนิกส์ ซึ่งมีปริมาณเพิ่มขึ้นอย่างรวดเร็ว จากที่เคยเก็บข้อมูลแบบแฟ้มเอกสาร หรือการจดบันทึก เปลี่ยนมาเป็นเก็บข้อมูลลงในสื่ออิเล็กทรอนิกส์ ซึ่งข้อมูลเหล่านี้มีขนาดใหญ่ขึ้นเรื่อยๆ อย่างรวดเร็ว ในปัจจุบันมีข้อมูลจำนวนมากที่เป็นข้อมูลประเภทข้อความ (Text) เช่น ข้อมูลงานวิจัย ข้อมูลข่าว หนังสืออิเล็กทรอนิกส์ ซึ่งส่วนใหญ่จะเป็นข้อมูลแบบกึ่งโครงสร้างหรือโครงสร้างไม่ชัดเจน (Semi-structured Data) โครงสร้างนี้ประกอบด้วย เขตข้อมูลที่เป็นโครงสร้าง (Structure Data Field) และเขตข้อมูลที่ไม่เป็นโครงสร้าง (Unstructure Data Field) เช่น ข้อมูลหนังสือ ประกอบด้วย ชื่อเรื่องหนังสือ ชื่อผู้แต่ง วันที่ในการจัดพิมพ์ เป็นข้อมูลแบบมีโครงสร้าง ส่วนเนื้อหาในหนังสือ เป็นข้อมูลแบบไม่มีโครงสร้าง ข้อมูลเหล่านี้จะถูกจัดเก็บไว้ตามแหล่งเก็บต่าง ๆ เช่น Text Database เป็นต้น การจะสกัดความรู้ที่ถูกซ่อนอยู่ในข้อมูลที่มีโครงสร้างไม่ชัดเจนและข้อมูลมีจำนวนมาก มาใช้ประโยชน์ตามความต้องการ คือการทำเท็กไมนิ่ง (Text Mining) ซึ่งวิธีการในการทำเท็กไมนิ่ง มี หลายวิธี แต่ที่นิยมในปัจจุบัน ได้แก่ การจัดกลุ่มเอกสาร (Text Clustering) และ การจำแนกประเภทเอกสาร (Text Classification หรือ Text Categorization)

การจำแนกประเภทเอกสาร มีหลายวิธีดังนี้ 1. วิธีเค-เนียเรส เนเบอร์ (K-Nearest Neighbor : KNN) หลักการคือ คำนวณหาระยะห่างระหว่างข้อมูลใหม่ที่ต้องการจำแนก กับข้อมูลในกลุ่มตัวอย่าง กล่าวคือเมื่อต้องการทราบว่าเอกสารที่ต้องการทดสอบ อยู่ในกลุ่มเอกสารประเภทใด โดยทำการค้นหาเอกสารในกลุ่มเรียนรู้ ที่มีความใกล้เคียงกับเอกสารที่ต้องการทดสอบมากที่สุด เป็นจำนวน K ชุด และตรวจสอบว่าเอกสารกลุ่มเรียนรู้ ที่ใกล้เคียงกับเอกสารทดสอบ เป็นเอกสารประเภทใด ให้ตัดสินว่าเอกสารทดสอบเป็นเอกสารประเภทนั้น 2. วิธีนาอีฟ เบย์ (Naive Bayes : NB) ซึ่งวิธีนี้อาศัยหลักการทางสถิติมาทำการคำนวณ สามารถคาดเดาการเป็นสมาชิกของ Class ได้ การจำแนกประเภทเอกสารวิธีนี้ อาศัยพื้นฐานมาจากทฤษฎีของ Bayes 3. วิธีใช้โครงข่ายประสาทเทียม (Neural Networks : NN) วิธีนี้จะมีการเรียนรู้จากชุดข้อมูลสำหรับเรียนรู้ และมีการคำนวณค่าระหว่างชุดข้อมูลเข้า และ Weight ที่เชื่อมต่อระหว่างโหนด ซึ่งค่าที่ได้นำไปเปรียบเทียบกับค่า Threshold ที่กำหนดไว้ ถ้าผ่านค่า Threshold จะมีการปรุงปรุงค่า Weight เมื่อเสร็จสิ้นการเรียนรู้จะมีการทดสอบการทำงาน ด้วยชุดข้อมูลทดสอบ เพื่อทดสอบการทำงานของระบบว่าถูกต้องหรือไม่ ทั้ง 3 วิธีดังกล่าว เป็นวิธีในการจำแนกประเภทเอกสาร ที่เป็นที่รู้จักกันดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในงานวิจัยนี้ จะกล่าวถึงวิธีการจำแนกประเภทเอกสารด้วยโครงข่ายประสาทเทียม โดยทั่วไปจะมีการรับข้อมูลเข้าเป็นแบบตัวเลข ซึ่งการที่จะรับข้อมูลเข้าสู่โครงข่ายประสาทเทียม จะต้องทำการแปลงค่าให้เป็นค่าเชิงตัวเลขก่อน โครงข่ายประสาทเทียมที่นำมาใช้ในการจำแนกประเภทเอกสาร จะเป็นประเภทที่มีการเรียนรู้แบบมีการชี้นำ (Supervised Learning) ซึ่งที่เป็นที่รู้จักกันดี คือ Backpropagation Neural Network แต่การรับข้อมูลเข้า จะรับเป็นค่าเชิงตัวเลข ซึ่งจะต้องทำการแปลงค่าเสียก่อน

การแปลงค่าของข้อมูลที่เป็นข้อความ ให้อยู่ในลักษณะของค่าเชิงตัวเลขเพื่อให้สามารถนำข้อมูลเข้าสู่โมเดล (Model) ได้นั้น จะมีขั้นตอนการแปลงค่า เรียกว่า การแทนเอกสาร (Document Representation) ซึ่งจะต้องแปลงค่าข้อมูลที่เป็นข้อความให้อยู่ในรูปแบบตัวเลขที่สามารถประมวลผลด้วยโมเดลทั่วไปได้ วิธีการหนึ่งที่ใช้อย่างแพร่หลาย คือ วิธีการ Vector Space Model นำเสนอโดย Gerard Salton [1] มีลักษณะข้อมูลเป็น Matrix Array 2 Dimensions แต่การแทนเอกสารโดยวิธีการนี้จะทำให้สูญเสียเวลาในขั้นตอนการแปลงข้อมูล เนื่องจากจะต้องเลือกคำสำคัญที่จะใช้เป็นตัวแทนของเอกสารให้เหมาะสมเสียก่อน และยังเกิดปัญหา High Dimensions เนื่องจากคำสำคัญที่ใช้เป็นตัวแทนของเอกสารแต่ละเอกสารมีจำนวนคำมาก จำนวนของ Dimension จะมากตามจำนวนของคำที่ใช้ ซึ่งมีผลกับการทำงานของโมเดลที่จะต้องใช้เวลาในการคำนวณมากขึ้น

ในปัจจุบันได้มีการออกแบบอัลกอริทึมสำหรับการจำแนกกลุ่มเอกสาร ที่มีการจัดการกับข้อมูลประเภทข้อความหรือลักษณะข้อมูลที่ไม่ใช่ตัวเลขได้โดยตรง โดยไม่ต้องทำการแปลงค่าจากค่าเชิงคุณภาพ (Qualitative Value) ไปเป็นค่าเชิงตัวเลข (Numerical Value)

สำหรับโครงข่ายประสาทเทียมที่สามารถรับข้อมูลที่เป็นข้อความได้โดยตรง ได้พัฒนาจากโครงข่ายประสาทเทียมเดิม ได้แก่ Text Processing Kohonen Neural Network [2] พัฒนาจาก Kohonen Neural Network, Text Processing Competitive Learning Neural Network [3] พัฒนาจาก Competitive Learning Neural Network และ Text Adaptive Resonance Theory Neural Network [4] พัฒนาจาก Adaptive Resonance Theory Neural Network ซึ่งโครงข่ายประสาทเทียมใหม่เหล่านี้สามารถแบ่งกลุ่มข้อความได้โดยตรง โดยไม่มีการแปลงรูปแบบข้อความซึ่งเป็นค่าเชิงคุณภาพไปเป็นค่าเชิงตัวเลข และโมเดลเหล่านี้ได้นำแนวความคิดของการวัดความคล้ายคลึงกัน (Similarity Measure) และวัดความแตกต่างกัน (Dissimilarity Measure) ของซิมโบลิคอ็อบเจ็ค (Symbolic Objects) มาประยุกต์ใช้ โดยการวัดความคล้ายคลึงกัน จะวัดที่ความคล้ายคลึงกันมากที่สุด และการวัดความแตกต่างกัน จะวัดที่ความแตกต่างกันน้อยที่สุด [5], [6] และ [7]

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

งานวิจัยนี้ มีวัตถุประสงค์เพื่อประยุกต์การทำงานของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ ซึ่งมีการเรียนรู้แบบมีการชี้นำ (Supervised Learning) และสามารถรับข้อมูลที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มีค่าเป็นตัวเลขได้เท่านั้น ให้สามารถรับข้อมูลที่เป็นข้อความได้โดยตรง โดยได้นำแนวความคิดของการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็คมาประยุกต์ใช้ และยังได้นำหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี [4] ที่เรียนรู้ด้วยตนเองหรือเรียนรู้แบบไม่มีการชี้นำ (Unsupervised Learning) มาใช้ ซึ่งโมเดลนี้สามารถรับข้อมูลที่เป็นค่าเชิงคุณภาพ (Qualitative Value) ได้โดยตรง โดยไม่ผ่านขั้นตอนการแปลงข้อมูลจากค่าเชิงคุณภาพไปเป็นค่าเชิงตัวเลข (Numerical Value) ซึ่งจะประหยัดเวลาในจัดการกับข้อมูลก่อนนำเข้าสู่โมเดล และเพื่อให้โมเดลนี้สามารถจำแนกกลุ่มเอกสารโดยรับค่าที่เป็นข้อความได้โดยตรงอย่างมีประสิทธิภาพ สามารถทำงานได้ถูกต้องเป็นที่ยอมรับได้

1.3 ทฤษฎีหรือแนวคิดที่ใช้ในการวิจัย

งานวิจัยนี้ได้นำทฤษฎีและแนวความคิดต่างๆ มาใช้ คือการประยุกต์ใช้โครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พซึ่งเป็นโครงข่ายประสาทเทียมที่มีการเรียนรู้แบบมีการชี้นำ โดยมีคุณสมบัติที่ดี 2 ข้อคือ 1. ไม่ลืมสิ่งที่เคยเรียนรู้มาก่อนหน้าหรือสิ่งที่เคยเรียนรู้มาแล้วไม่เลือนหายไป (Stability) และ 2.สามารถเรียนรู้สิ่งใหม่ได้โดยไม่ลืมสิ่งที่เคยเรียนรู้มาก่อนหน้าหรือรักษาความรู้เก่าให้คงอยู่ (Plasticity) และได้นำแนวความคิดของการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็คมาประยุกต์ใช้ รวมทั้งหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้น-เทียรีมาใช้

1.4 ขอบเขตของการวิจัย

งานวิจัยที่นำเสนอนี้ เป็นโมเดลที่ประยุกต์ใช้โครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ ในการจำแนกประเภทเอกสาร ที่จะรับชุดข้อมูลเข้า (Input Pattern) ที่เป็นข้อความภาษาอังกฤษเท่านั้น ซึ่งจะรับข้อความโดยตรง โดยไม่ผ่านการแปลงค่าจากค่าเชิงคุณภาพให้เป็นค่าเชิงตัวเลข โมเดลนี้ยังได้นำแนวความคิดของการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็ค และหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรีมาประยุกต์ใช้

1.5 ขั้นตอนของการศึกษา

ในขั้นตอนของการศึกษานี้ ได้แสดงลำดับการทำงานตั้งแต่เริ่มต้นจนถึงสิ้นสุดการทำงานดังรายละเอียดต่อไปนี้

1.5.1 ศึกษาทฤษฎีและงานวิจัยจากเอกสารบทความต่าง ๆ ที่เกี่ยวข้องกับการทำงานวิจัย

1.5.2 กำหนดหัวข้อ วัตถุประสงค์ และขอบเขตการทำงานวิจัย

1.5.3 วิเคราะห์อัลกอริทึมและออกแบบโมเดลใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.5.4 พัฒนาโปรแกรม โดยใช้ซอฟท์แวร์ MATLAB รุ่น 6.5 พร้อมทั้งแก้ไขข้อผิดพลาด และทดสอบการทำงานของโมเดลกับข้อมูลที่กำหนดขึ้นเอง

1.5.5 เตรียมข้อมูลที่ใช้งานจริง เพื่อนำมาทดสอบการทำงานโมเดล

1.5.6 ทดลองกับข้อมูลที่ใช้งานจริง พร้อมทั้งวัดประสิทธิภาพการทำงานของโมเดล

1.5.7 รวบรวมผลการทดลองจากการทำงานของโมเดล

1.5.8 วิเคราะห์และสรุปผลการทดลอง

1.5.9 เรียบเรียงเอกสารประกอบวิทยานิพนธ์

การทำเท็กไมนิ่ง เพื่อให้ได้ความรู้ที่เป็นประโยชน์จากข้อมูลประเภทข้อความที่มีอยู่มากมายในปัจจุบันมาใช้งาน มีวิธีที่นิยม 2 รูปแบบ คือ การจัดกลุ่มเอกสาร และการจำแนกประเภทเอกสาร งานวิจัยนี้ ได้ทำการจำแนกกลุ่มเอกสารโดยใช้โครงข่ายประสาทเทียม โดยประยุกต์ใช้หลักการทำงานของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ ซึ่งมีการเรียนรู้แบบมีการชี้นำและสามารถรับข้อมูลที่มีค่าเป็นตัวเลขได้เท่านั้น ให้สามารถรับข้อมูลที่เป็นข้อความได้โดยตรง โดยไม่ผ่านขั้นตอนการแปลงข้อมูลจากค่าเชิงคุณภาพไปเป็นค่าเชิงตัวเลข อีกทั้งได้นำแนวความคิดของการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็คและหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรีมาประยุกต์ใช้

ในบทที่ 2 จะกล่าวถึง ทฤษฎีต่างๆ ที่เกี่ยวข้องกับการทำงานวิจัยที่นำเสนอนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

ในงานด้านการค้นคืนสารสนเทศ (Information Retrieval) และงานด้านการทำเท็กไมนิ่ง (Text Mining) จะมีการนำเอกสารต่างๆ จากแหล่งเก็บข้อมูลเอกสารไปใช้งานตามที่ต้องการ ซึ่งเอกสารเหล่านี้มีการเก็บอยู่ในรูปแบบแฟ้มข้อมูลต่างๆ เช่น HTML, XML และ SGML แต่ก่อนจะนำเอกสารเหล่านี้ไปใช้งาน จะต้องมีการเตรียมเอกสารให้อยู่ในรูปแบบที่สามารถนำไปใช้งานได้เสียก่อน ซึ่งมีรายละเอียดดังนี้

2.1 การเตรียมเอกสาร (Document Preprocessing)

ขั้นตอนแรกของการจำแนกประเภทเอกสาร (Text Categorization) คือการเตรียมเอกสารให้อยู่ในรูปแบบที่จะสามารถนำเข้าสู่โมเดลของการเรียนรู้ได้ ซึ่งจะมีการดำเนินการเปลี่ยนรูปแบบเอกสาร (Text Transformations หรือ Text Operations) [8] 5 ลักษณะ ดังต่อไปนี้

ลักษณะที่ 1 การวิเคราะห์คำ (Lexical Analysis) พิจารณาการแยกคำ ตัวเลข รวมทั้งเครื่องหมายวรรคตอนต่างๆ

ลักษณะที่ 2 กำจัด Stopwords (Elimination of Stopwords) คือตัดคำที่มีอำนาจในการแบ่งแยกได้น้อย เช่น a, an, the ออกไป

ลักษณะที่ 3 ทำให้อยู่ในรูปแบบของรากคำหรือ Word Stemming คือตัดส่วนที่เป็น Affixes (Prefixes และ Suffixes) ของคำออกไป

ลักษณะที่ 4 เลือก Index Terms คือการหาคำสำคัญที่ใช้แทนเอกสารที่เหมาะสม

ลักษณะที่ 5 การสร้าง Thesaurus คือการหาศัพท์สัมพันธ์

2.2 การทำดัชนี (Indexing)

การแทนเอกสาร (Document Representation) โดยทั่วไปที่ใช้กันอย่างแพร่หลาย คือวิธี Vector Space Model (SMART) [9] ซึ่งวิธีการแทนเอกสารนี้ เอกสารจะแทนโดย Vector ของคำสำคัญ ซึ่งเอกสาร 1 เอกสาร จะแปลงให้อยู่ในรูปโครงสร้างของ Matrix A เมื่อสมาชิกแต่ละตัว คือคำสำคัญที่ปรากฏในเอกสารดังสมการที่ (2.1)

เมื่อ a_{ik} คือ Weight ของคำสำคัญ i ในเอกสาร k โดยที่ Matrix A มีจำนวนแถวเท่ากับ M ซึ่งคือจำนวนของคำสำคัญที่ใช้นั่นเอง ทำให้ M มีขนาดใหญ่เนื่องจากมีจำนวนคำสำคัญของแต่ละเอกสารจำนวนมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$A = (a_{ik}), \tag{2.1}$$

วิธีการในการกำหนด Weight ของ a_{ik} ของคำสำคัญ i ในเอกสาร k มีหลายวิธีดังจะได้กล่าวต่อไป เมื่อกำหนดให้ f_{ik} คือความถี่ของคำสำคัญ i ปรากฏในเอกสาร k ส่วน N คือจำนวนเอกสารทั้งหมด M คือจำนวนของคำสำคัญที่ใช้หลังจากตัด Stopwords และตัด Word Stemming แล้ว และ n_i คือ จำนวนครั้งของคำสำคัญ i ที่ปรากฏในทั้ง Collection วิธีการต่างๆ ในการกำหนด Weighting มีดังนี้

2.2.1 Boolean Weighting

เป็นวิธีการที่ง่ายที่สุด วิธีนี้สามารถบอกได้เพียงว่ามีคำสำคัญนั้นปรากฏในเอกสารหรือไม่ คือแทนค่า Weight ด้วย 0 และ 1 โดยจะแทนค่าเป็น 1 ถ้ามีคำสำคัญนั้นปรากฏในเอกสาร และแทนค่าเป็น 0 เมื่อไม่มีคำสำคัญนั้นปรากฏในเอกสาร ดังสมการที่ (2.2)

$$a_{ik} = \begin{Bmatrix} 1 & if\ f_{ik} > 0 \\ 0 & otherwise \end{Bmatrix} \tag{2.2}$$

2.2.2 Term Frequency Weighting

วิธีการนี้สามารถบอกความถี่ของคำสำคัญที่ปรากฏในเอกสารได้ จึงมีประสิทธิภาพดีกว่าแบบ Boolean Weighting เป็นวิธีการที่หาความถี่ของคำสำคัญที่ปรากฏในเอกสารว่ามีคำสำคัญนั้นปรากฏในเอกสารกี่ครั้ง โดยยังคงใช้ 0 แทนความหมายว่าไม่มีคำสำคัญนั้นปรากฏในเอกสาร ดังสมการที่ (2.3)

$$a_{ik} = f_{ik} \tag{2.3}$$

2.2.3 TF x IDF-Weighting

วิธีการนี้เป็นที่รู้จักกันดีในการคำนวณค่าน้ำหนักของคำสำคัญคือ วิธี TF x IDF-Weighting ซึ่งเป็นวิธีการที่หาความถี่ของคำสำคัญที่ปรากฏในเอกสารร่วมกับการหาน้ำหนักของคำสำคัญที่ปรากฏในเอกสาร เนื่องจากคำสำคัญแต่ละคำมีความสำคัญแตกต่างกัน ดังสมการที่ (2.4)

$$a_{ik} = f_{ik} * \log\left(\frac{N}{n_i}\right) \tag{2.4}$$

2.2.4 TFC-Weighting

วิธีการนี้จะคล้ายกับ TF x IDF-Weighting แต่มีการ Normalization ดังสมการที่ (2.5)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$a_{ik} = \frac{f_{ik} * \log\left(\frac{N}{n_i}\right)}{\sqrt{\sum_{j=1}^{M}\left[f_{jk} * \log\left(\frac{N}{n_j}\right)\right]^2}} \qquad (2.5)$$

2.2.5 LTC-Weighting

วิธีการนี้แตกต่างจากวิธี TFC-Weighting เล็กน้อย คือใช้ logarithm ของความถี่คำสำคัญแทนความถี่ของคำสำคัญแบบเดิม ซึ่งจะทำให้ลดความถี่ที่แตกต่างกันมากได้

$$a_{ik} = \frac{\log(f_{ik} + 1.0) * \log\left(\frac{N}{n_i}\right)}{\sqrt{\sum_{j=1}^{M}\left[\log(f_{jk} + 1.0) * \log\left(\frac{N}{n_j}\right)\right]^2}} \qquad (2.6)$$

วิธีการทั้งหมดที่กล่าวมาเป็นวิธีการในการกำหนด Weighting ของคำสำคัญที่ใช้เป็นตัวแทนของเอกสาร

2.3 โครงข่ายประสาทเทียม (Artificial Neural Network)

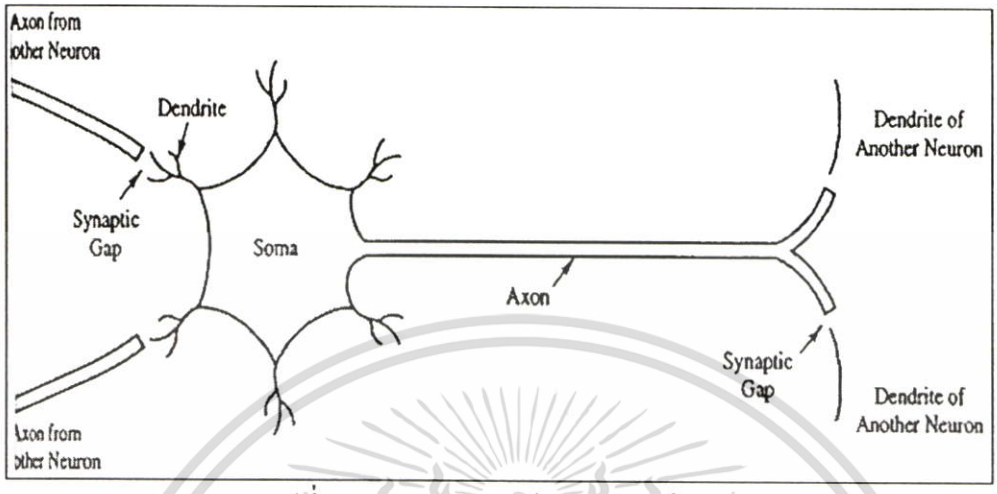
โครงข่ายประสาทเทียมเป็นการจำลองลักษณะการทำงานของเซลล์ประสาทของมนุษย์ ซึ่งเซลล์ประสาทในทางชีววิทยา (Biological Neuron) มีส่วนประกอบหลักๆ 3 ส่วน [10] คือ 1.เด็นไดรท์ (Dendrites) 2.โซม่า (Soma) และ 3.แอ็กซอน (Axon) ส่วนของเด็นไดรท์จะได้รับสัญญาณไฟฟ้าจากนิวรอน (Neuron) อื่นที่ส่งมาให้ การส่งสัญญาณจะส่งข้ามส่วนที่เรียกว่า Synaptic Gap โดยกระบวนการทางเคมี โซม่าหรือตัวเซลล์ (Cell Body) จะมีกระบวนการทำงานภายใน โดยมีตัวส่งสัญญาณทางเคมี ซึ่งการทำงานของตัวส่งสัญญาณทางเคมีจะเปลี่ยนแปลงสัญญาณที่รับเข้ามา โดยการรวบรวมสัญญาณที่รับเข้ามา เมื่อได้รับสัญญาณเข้าจนเพียงพอแล้วเซลล์จะมีการทำงาน (Fire) แล้วจะส่งสัญญาณไปตามแอ็กซอน เพื่อส่งไปยังเซลล์อื่นต่อไป ดังรูปที่ 2.1

โครงข่ายประสาทเทียม ที่จำลองลักษณะการทำงานของเซลล์ประสาทของมนุษย์ มีส่วนของการรับข้อมูล เปรียบเสมือนกับการทำงานของเด็นไดรท์ ส่วนของการประมวลผลข้อมูลเปรียบเสมือนกับโซม่าหรือตัวเซลล์ และส่วนของการแสดงผลลัพธ์ เปรียบเสมือนกับแอ็กซอน

โครงข่ายประสาทเทียมในกลุ่มของ ARTMAP ทุกตัว จะมีคุณสมบัติเด่นหลัก ๆ อยู่ 2 ข้อ คือ 1. ไม่ลืมสิ่งที่เคยเรียนรู้มาก่อนหน้าหรือสิ่งที่เคยเรียนรู้มาแล้วไม่เลือนหายไป (Stability) และ 2.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถเรียนรู้สิ่งใหม่ได้โดยไม่ลืมสิ่งที่เคยเรียนรู้มาก่อนหน้าหรือรักษาความรู้เก่าให้คงอยู่ (Plasticity)



รูปที่ 2.1 แสดงโครงสร้างของเซลล์ประสาท [10]

2.4 โครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ (Simplified Fuzzy ARTMAP Neural Network)

โครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ พัฒนามาจากโครงข่ายประสาทเทียมแบบฟัซซี่อาร์ทแม็พ (Fuzzy ARTMAP Neural Network) โดย Tom Kasuba [11] ได้ตัดสถาปัตยกรรมที่ซ้ำซ้อนของโครงข่ายประสาทเทียมแบบฟัซซี่อาร์ทแม็พ ในส่วนของ ART_b ออกไป คงเหลือไว้เฉพาะส่วนของ ART_a และส่วนของ Map Field เท่านั้น ซึ่งทำให้โครงข่ายประสาทเทียมนี้ทำงานได้เร็วขึ้น

สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ [11] มีจำนวนชั้นการทำงานทั้งหมด 4 ชั้น (Layer) คือ 1. Raw Input Layer 2. Input Layer 3. Output Category Layer และ 4. Category Layer ดังรูปที่ 2.2

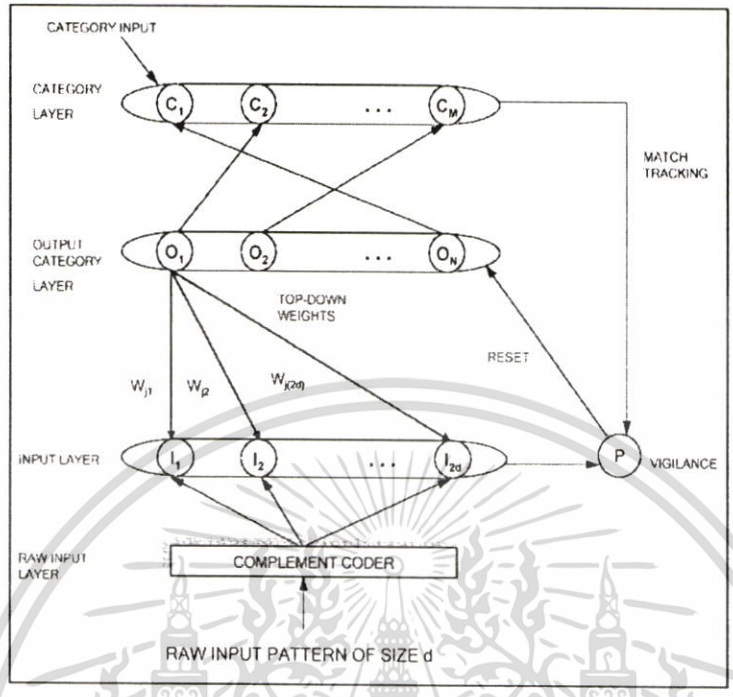
ที่ Raw Input Layer มีการนำข้อมูลดิบที่มีค่าอยู่ระหว่าง 0 ถึง 1 มาทำ Complement Coding ตามสมการที่ (2.7)

$$\overline{a_i} = 1 - a_i \tag{2.7}$$

ดังนั้นชุดข้อมูลเข้า (Input Pattern) ที่ทำ Complement Coding แล้วจะมีขนาดเป็น 2 เท่าดังสมการที่ (2.8)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$I = (a, \overline{a}) = (a_1, ..., a_d, \overline{a_1}, ..., \overline{a_d}) \tag{2.8}$$



รูปที่ 2.2 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ [11]

แล้วส่งค่าของชุดข้อมูลเข้า ไปยัง Input Layer พร้อมกับส่งค่า Category Input มายัง Category Laye

ที่ Input Layer จะมีการส่งค่าของชุดข้อมูลเข้าไปยัง Output Category Layer ซึ่งระหว่าง 2 ชั้นนี้จะมีชุด Weight เชื่อมต่อแบบถึงกันหมดระหว่างโหนดอินพุท (Input Node) และโหนดเอาท์พุท (Output Node) เรียกว่า Top-down Weight ที่ Output Category Layer จะมีจำนวนโหนดเอาท์พุทได้ไม่จำกัดและจะเพิ่มจำนวนขึ้นเรื่อยๆ ระหว่างการเรียนรู้ ช่วงการเรียนรู้จะส่งค่าข้อมูลเข้า มาเปรียบเทียบกับชุด Weight ที่เชื่อมต่อไปยังโหนดเอาท์พุท โดยการคำนวณค่าตาม Activation Function ตามสมการที่ (2.9)

$$T_j(I) = \frac{|I \wedge W_j|}{\alpha + |W_j|} \tag{2.9}$$

ซึ่งจะมีการคำนวณค่าทุกโหนดเอาท์พุท จนครบทุกโหนดเพื่อตัดสินว่าจะเก็บค่าของชุดข้อมูลเข้าไว้ที่โหนดเอาท์พุทใด แล้วเลือกโหนดเอาท์พุทที่มีค่ามากที่สุดเป็นโหนดที่ชนะตามสมการที่ (2.10)

$$Winner = \max(T_j) \tag{2.10}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้าโหนดเอาท์พุทที่ค่ามากที่สุด มีมากกว่า 1 โหนด ให้เลือกโหนดที่มีค่า j น้อยที่สุด หลังจากเลือกโหนดที่ชนะได้แล้ว ให้ทำการตรวจสอบว่า ค่า Weight ที่เก็บอยู่ในโหนดเอาท์พุทนั้น มีความคล้ายคลึงกันกับชุดข้อมูลเข้า ที่รับเข้ามา มากพอหรือไม่ โดยนำโหนดเอาท์พุทที่ชนะมาทำการคำนวณตาม Match Function ตามสมการที่ (2.11)

$$\frac{|I \wedge W_j|}{|I|} \tag{2.11}$$

ซึ่งสามารถแปลง Match Function ได้ดังนี้

$$\frac{|I \wedge W_j|}{d} \tag{2.12}$$

โดยที่

$$|I| = |(a, \bar{a})| = \sum_{i=1}^{d} a_i + (d - \sum_{i=1}^{d} a_i) = d \tag{2.13}$$

หลังจากคำนวณค่าตามสมการ Match Function ได้แล้ว นำค่าที่ได้มาเปรียบเทียบกับค่าค่าวิจิแลนซ์ (Vigilance) ตามสมการที่ (2.14)

$$\frac{|I \wedge W_j|}{d} \geq \rho \tag{2.14}$$

ค่าวิจิแลนซ์ ซึ่งแทนด้วย ρ เป็นค่าที่ควบคุมการจำแนกกลุ่มใน Output Category Layer จะมีค่าอยู่ระหว่าง 0 ถึง 1 ค่านี้จะกำหนดเริ่มแรกเป็น Base-line Vigilance (แทนด้วย $\bar{\rho}$) เพราะจะมีการเพิ่มขึ้นเรื่อยๆ ในระหว่างการเรียนรู้ การที่กำหนดค่าวิจิแลนซ์ไว้สูง แสดงว่าชุดข้อมูลเข้าต้องคล้ายคลึงกันมากถึงจะจำแนกไว้ในกลุ่มเดียวกันได้ ซึ่งจะทำให้โหนดเอาท์พุทมีขนาดเล็กและจำนวนโหนดมาก แต่ถ้ากำหนดค่าวิจิแลนซ์ไว้ต่ำ ชุดข้อมูลเข้าไม่ต้องคล้ายคลึงกันมากก็สามารถจำแนกไว้ในกลุ่มเดียวกันได้ ทำให้โหนดเอาท์พุทมีขนาดใหญ่และจำนวนโหนดน้อย

ถ้าค่าของโหนดเอาท์พุทที่ชนะมีค่าจากการคำนวณมากกว่าหรือเท่ากับค่าวิจิแลนซ์ แสดงว่าระดับความคล้ายคลึงกันของการจำแนกกลุ่มข้อมูลมากพอ สามารถจัดให้ชุดข้อมูลเข้าชุดนั้นเป็นสมาชิกของโหนดเอาท์พุทนั้นได้

ถ้าค่าจากการคำนวณน้อยกว่าค่าวิจิแลนซ์ คือไม่ผ่านค่าวิจิแลนซ์ เกิด Mismatch Reset แสดงว่าค่า Weight ที่เก็บอยู่ในโหนดเอาท์พุทนั้น ไม่มีความคล้ายคลึงกันกับชุดข้อมูลเข้ามากพอ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ให้ยับยั้งโหนดนั้น แล้วหาโหนดที่มีค่ามากที่สุดโหนดถัดไป ทดสอบค่าตามสมการ Match Function จนกว่าจะได้ค่ามากกว่าหรือเท่ากับค่าวิจิแลนซ์ คือผ่านค่าวิจิแลนซ์ แล้วดำเนินการตามขั้นตอนถัดไป

ถ้าไม่มีโหนดใดผ่านค่าวิจิแลนซ์ ให้สร้างโหนดเอาท์พุทใหม่ แต่ถ้าผ่านค่าวิจิแลนซ์ ให้ตรวจสอบต่อไปว่า Class ที่ตอบ ซึ่งเก็บอยู่ในโหนดเอาท์พุทนั้น ตรงกับ Category Input ที่รับเข้ามาหรือไม่ ถ้าไม่ตรงกัน จะมีการทำงานของ Match Tracking

ส่วนของ Match Tracking ทำหน้าที่ควบคุมความถูกต้องในการเรียนรู้ว่า ชุดข้อมูลเข้าถูกจำแนกไว้ในโหนดเอาท์พุทที่ชี้ไปยังคำตอบใน Category Layer ตรงกับ Category Input ของชุดข้อมูลเข้าที่กำลังเรียนรู้อยู่หรือไม่ ถ้าคำตอบไม่ตรงกันแสดงว่าชุดข้อมูลเข้าชุดนี้ไม่เหมาะสมที่จะจำแนกอยู่ในโหนดเอาท์พุทนี้ ถ้าเหตุการณ์นี้เกิดขึ้น ระบบจะกำหนดให้โหนดเอาท์พุทที่ชนะในขณะนั้นมีค่าเป็น 0 หรือค่าน้อยๆ เพื่อไม่ให้ระบบกลับมาเลือกโหนดเอาท์พุทนี้อีก และเพิ่มค่าวิจิแลนซ์ขึ้นเล็กน้อย แล้วหาโหนดเอาท์พุทที่ชนะโหนดถัดไป และตรวจสอบตามขั้นตอนเดิมจนกว่าโหนดเอาท์พุทที่เลือก มีคำตอบตรงกับ Category Input ของชุดข้อมูลเข้า ถ้าคำตอบตรงกัน ถือว่าการจำแนกกลุ่มสำเร็จ เรียกว่าเกิด Resonance ทำการปรับปรุงค่า Weight ตามสมการที่ (2.15)

$$W_j^{new} = \beta(I \wedge W^{old}{}_j) + (1-\beta)W^{old}{}_j \qquad (2.15)$$

ถ้าเป็นแบบ Fast Learning ให้ปรับปรุงค่า Weight ตามสมการที่ (2.16)

$$W_j^{new} = (I \wedge W^{old}{}_j) \qquad (2.16)$$

เมื่อปรับปรุงค่า Weight เสร็จแล้ว แสดงว่าระบบได้เรียนรู้ชุดข้อมูลเข้าชุดนั้น เรียบร้อยแล้ว ให้รับข้อมูลเข้าชุดถัดไป จนกว่าจะพบเงื่อนไขในการจบการทำงาน

สำหรับอัลกอริทึมในการเรียนรู้ ของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ท-แม็พ ตามแนวความคิดของ Tom Kasuba [9] แสดงรายละเอียดไว้ ดังรูปที่ 2.3

```
FOR  Some number of training epochs DO
    FOR_EACH  labeled training pattern DO
        IF_1  this category has never been seen before THEN
```

รูปที่ 2.3 อัลกอริทึมในการเรียนรู้ของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ [11]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Create an output node to encode this training pattern.

Set the new output node's top-down weights to the complement coded input pattern.

Set the category of the new output node to that of the input pattern.

ELSE

Set the vigilance to the baseline vigilance.

Present the input pattern and category to the network.

Evaluate the activation function for all output nodes.

Sort the output nodes in descending order according to activation.

FOR i = 1 TO the number of output nodes DO

Select the ith highest activated output node (the current winner).

IF_2 there is a mismatch reset ($| I \wedge W_j | / d \geq \rho$) THEN

Suppress activation of the current winning output node.

ELSE

IF_3 winning output node encodes the same category as the input pattern THEN

Update the winners top-down weights $W^{new} = | I \wedge W^{old} |$.

Set the indicator that there was no mismatch.

BREAK_FROM_FOR number of active output nodes loop.

ELSE

Set vigilance to the match value of the winning node plus a small epsilon.

Suppress activation of the current winning output node.

Indicate that there is a category mismatch.

END_IF_3

END_IF_2

END_FOR the number of output nodes

IF_4 there was a mismatch or a category mismatch THEN

รูปที่ 2.3 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                Create an output node to encode this training pattern.
                Set the new node's top-down weights to the complement coded
                input pattern.
                Set the category of the new output node to that of the input pattern.
            END_IF_4
        END_IF_1
    END_FOR_EACH training pattern
END_FOR some number of training epochs
```

รูปที่ 2.3 (ต่อ)

2.5 โครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี (A Text Adaptive Resonance Theory Neural Network)

โครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี [4] พัฒนามาจาก Adaptive Resonance Theory Neural Network ซึ่งได้ออกแบบให้รับข้อมูลที่เป็นข้อความได้โดยตรง โดยไม่มีการแปลงค่าข้อความให้เป็นค่าตัวเลข เป็นการเรียนรู้ด้วยตัวเองหรือไม่มีการชี้นำ สถาปัตยกรรมของโครงข่ายประสาทเทียมนี้ประกอบด้วยชั้นการทำงาน 3 ชั้น คือ 1. Input Layer ($F_{1(a)}$) 2. Interface Layer ($F_{1(b)}$) และ 3. Output Layer (F_2) ดังรูปที่ 2.4

ระหว่าง Interface Layer และ Output Layer มีชุด Weights 2 ชุดคือ Bottom-up Weight และ Top-down Weight ซึ่งจะเชื่อมต่อระหว่าง 2 Layer โดยที่ Weights ทั้ง 2 ชุด จะเก็บค่าใน Weight เป็นค่าเชิงคุณภาพและค่าแสดงความเป็นสมาชิก (Degree) ของ Weight เพื่อให้โครงข่ายประสาทเทียม สามารถหาความคล้ายคลึงของชุดข้อมูลเข้าที่เป็นค่าเชิงคุณภาพได้ ดังสมการที่ (2.17) และ (2.18)

$$b_{ij} = \{(A_{1ij}, e_{1ij}), (A_{2ij}, e_{2ij}), ..., (A_{pij}, e_{pij})\} \tag{2.17}$$

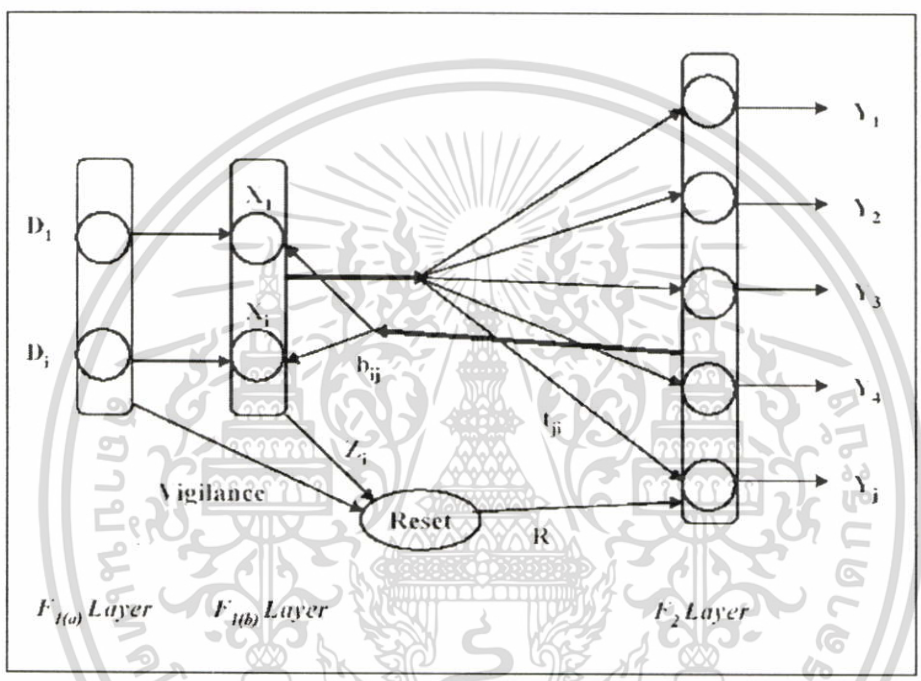
$$t_{ji} = \{(B_{1ji}, e_{1ji}), (B_{2ji}, e_{2ji}), ..., (B_{pji}, e_{pji})\} \tag{2.18}$$

กำหนดให้ A_{pij} คือค่าเชิงคุณภาพของ Bottom-up Weight และ e_{pij} คือค่าแสดงความเป็นสมาชิกของ A_{pij}

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่า e_{pij} ของ A_{pij} ในโครงข่ายประสาทเทียมนี้ มีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งค่านี้ เป็นค่าที่ให้ระดับความเป็นสมาชิกของ A_{pij} กับชุดข้อมูลเข้า ถ้า e_{pij} ของ A_{pij} มีค่าเท่ากับ 0 ให้ความหมายว่าค่าเชิงคุณภาพ A_{pij} ไม่ได้เป็นส่วนหนึ่งของชุดข้อมูลเข้าลำดับที่ i ถ้า e_{pij} ของ A_{pij} มีค่าเท่ากับ 1 ให้ความหมายว่าค่าเชิงคุณภาพ A_{pij} เป็นสมาชิกของชุดข้อมูลเข้าลำดับที่ i อย่างสมบูรณ์

ส่วนของ Top-down Weight มีลักษณะของการเก็บค่าใน Weight เหมือนกับ Bottom-up Weight



รูปที่ 2.4 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี [4]

อัลกอริทึมการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี [4] มีรายละเอียด ดังรูปที่ 2.5

ขั้นตอนที่ 0: กำหนดค่าเริ่มต้นให้กับ Bottom-up Weight และ Top-down Weight ในแต่ละโหนดของโครงข่ายประสาทเทียม ค่าเริ่มต้นเหล่านี้อาจได้จากการสุ่มเลือกจากข้อมูลที่นำมาใช้ในกระบวนการเรียนรู้ และกำหนดค่า Vigilance Parameter, $\rho = (0,1]$

รูปที่ 2.5 อัลกอริทึมการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี [4]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 1: เมื่อยังไม่ตรงเงื่อนไขในการหยุด ให้ทำขั้นตอนที่ 2-9

ขั้นตอนที่ 2: เลือกข้อมูลเข้าทั้งหมด แล้วทำการ Transpose

$$Doc = (D_1, D_2, D_3, ..., D_d)' \tag{2.19}$$

ทำขั้นตอนที่ 3-8 ต่อไป

ขั้นตอนที่ 3: กำหนดข้อมูลเข้า D จาก $F_{1(a)}$ ให้กับตัวแปร X ใน $F_{1(b)}$

$$X_i = D_i \tag{2.20}$$

ขั้นตอนที่ 4: คำนวณเปรียบเทียบข้อมูลเข้า X กับแต่ละโหนดของโครงข่ายประสาทเทียมจนครบทุกหนด

$$Y_j = \sum_{k=1}^{d} \sum_{n=1}^{p} S(X_k, A_{nkj}).e_{nkj} \tag{2.21}$$

กำหนดให้

p คือจำนวนของ Bottom-up Weight

d คือจำนวนของ Feature Values

ขั้นตอนที่ 5: เปรียบเทียบหาโหนด J ของโครงข่ายประสาทเทียมที่มีค่ามากที่สุด

ขั้นตอนที่ 6: เปรียบเทียบข้อมูลเข้า X กับ Top-down Weight ที่เชื่อมต่อกับ Winning Node J

$$Z = \frac{S(X_i, t_{Ji}) - (0.5 * no_feature)}{(2 * no_feature) - (0.5 * no_feature)} \tag{2.22}$$

ขั้นตอนที่ 7: ทดสอบเงื่อนไขการ Reset Mechanism

ถ้า $Z < \rho$, ให้ Y_J = -1 (ยับยั้งโหนด J), ทำขั้นตอนที่ 5 อีกครั้ง

ในกรณี ทุกโหนดเอาท์พุทถูกยับยั้ง ให้สร้างโหนดเอาท์พุทใหม่

รูปที่ 2.5 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้า $Z \geq \rho$, ให้ทำขั้นตอนที่ 8

ขั้นตอนที่ 8: ปรับปรุงค่า Bottom-up Weight และ Top-down Weight ของโหนด J

$$b_{iJ}^{(new)} = b_{iJ}^{(old)} \cap X$$

$$e_{niJ}^{(new)} = \begin{cases} f(e_{niJ}^{(old)} + \eta) & \text{if} \quad A_{nij} \in b_{iJ} \cap X \quad , \\ f(e_{niJ}^{(old)} - \eta) & \text{if} \quad A_{nij} \notin b_{iJ} \cap X \quad , \\ 5 * \eta \quad ; & \text{otherwise} \end{cases} \tag{2.23}$$

$$t_{Ji}^{(new)} = t_{Ji}^{(old)} \cap X$$

$$e_{nJi}^{(new)} = \begin{cases} f(e_{nJi}^{(old)} + \eta) & \text{if} \quad B_{nji} \in t_{Ji} \cap X \quad , \\ f(e_{nJi}^{(old)} - \eta) & \text{if} \quad B_{nji} \notin t_{Ji} \cap X \quad , \\ 5 * \eta \quad ; & \text{otherwise} \end{cases} \tag{2.24}$$

กำหนดให้ $f(.)$ นิยามดังนี้

$$f(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x < 0 \\ 1 & \text{if } x > 1 \end{cases} \tag{2.25}$$

ขั้นตอนที่ 9: ทดสอบเงื่อนไขการหยุด เงื่อนไขการหยุดสามารถกำหนดโดย:

9.1 Weight ไม่มีการเปลี่ยนแปลง

9.2 ไม่มีการรีเซต

9.3 จำนวนรอบของการเรียนรู้

รูปที่ 2.5 (ต่อ)

2.6 การเปรียบเทียบความคล้ายคลึงกันของเอกสาร (Similarity Measure)

การเปรียบเทียบความคล้ายคลึงของเอกสาร [6] นั้น จะต้องผ่านขั้นตอนการแทนเอกสารต่างๆ (Document Representation) ให้อยู่ในรูปแบบที่สามารถนำมาใช้ในขั้นตอนการเปรียบเทียบได้ เอกสารต่างๆ ประกอบด้วยคุณสมบัติต่างๆ ซึ่งสามารถแทนให้อยู่ในรูปแบบของ Cartesian Product ได้ดังนี้ [7]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

$$Doc = D_1 \times D_2 \times D_3 \times ... \times D_d \quad (2.26)$$

d คือจำนวนคุณสมบัติของเอกสาร ซึ่งมีลักษณะเป็นข้อความ

ตัวอย่าง คุณสมบัติ (Features) ของเอกสาร ในข้อมูลข่าว Reuters-21578 [12] ซึ่งเป็นชุดข้อมูลที่ใช้สำหรับทดสอบโมเดลด้านการจำแนกประเภทเอกสาร ประกอบด้วยคุณสมบัติต่างๆ เช่น Title, Author และ Keyword เป็นต้น ในที่นี้เลือกใช้ 2 คุณสมบัติ คือคุณสมบัติชื่อเรื่อง (Title Feature) และคุณสมบัติคำสำคัญ (Keyword Feature) โดยสามารถแทนให้อยู่ในรูปแบบของ Cartesian Product ได้ดังนี้

$$Doc = Title \times Keyword \quad (2.27)$$

เมื่อ Title คือ คำที่ใช้อธิบายชื่อเรื่องของเอกสาร

Keyword คือ คำที่ใช้เป็นตัวแทนของเอกสาร หรือคำสำคัญของเอกสาร

การเปรียบเทียบความคล้ายคลึงของเอกสาร ระหว่างเอกสาร A และเอกสาร B ตามแนวความคิดของ El-Sonbaty [7] สามารถเขียนเป็นสมการได้ดังนี้

$$S(A,B) = \sum_{k=1}^{d} S(A_k, B_k) \quad (2.28)$$

ในการเปรียบเทียบความคล้ายคลึงของ $S(A_k, B_k)$ มีการเปรียบเทียบ 2 ส่วนย่อยคือ ส่วนที่เป็นขนาดของคุณสมบัติ เรียกว่า Span $S_s(A,B)$ และส่วนที่เป็นเนื้อหาของคุณสมบัติ เรียกว่า Content $S_c(A,B)$ ซึ่งนิยามสมการไว้ดังนี้ [6]

ส่วนของการเปรียบเทียบความคล้ายคลึงของ Span มีนิยามตามสมการ (2.29)

$$S_s(A_k, B_k) = \frac{(l_a + l_b)}{2.l_s} \quad (2.29)$$

ส่วนของการเปรียบเทียบความคล้ายคลึงของ Content มีนิยามตามสมการ (2.30)

$$S_c(A_k, B_k) = \frac{inters}{l_s} \quad (2.30)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

60905

นิยามของสัญลักษณ์ในสมการมีดังนี้

l_a = จำนวนสมาชิกทั้งหมดในคุณสมบัติของเอกสาร A

l_b = จำนวนสมาชิกทั้งหมดในคุณสมบัติของเอกสาร B

$inters$ = จำนวนของสมาชิกทั้งหมดที่ Intersection กันระหว่างคุณสมบัติของเอกสาร A และคุณสมบัติของเอกสาร B

l_s = จำนวนของสมาชิกทั้งหมดที่ Union กันระหว่างคุณสมบัติของเอกสาร A และ คุณสมบัติของเอกสาร B

เมื่อเปรียบเทียบความคล้ายคลึงครบทั้งสองส่วนย่อยแล้ว ให้ทำการรวมค่าผลลัพธ์จากสองส่วนย่อยเข้าด้วยดังสมการ Net Similarity ดังนี้

$$S(A_k, B_k) = S_s(A_k, B_k) + S_c(A_k, B_k) \qquad (2.31)$$

ค่าความคล้ายคลึงของ Net Similarity นี้สามารถมีค่าได้ระหว่าง [0.5,2] ตัวอย่างคุณสมบัติของเอกสาร A และ B แสดงในตารางที่ 2.1

ตารางที่ 2.1 แสดงตัวอย่างข้อมูลเอกสาร

เอกสาร	คุณสมบัติชื่อเรื่อง (Title)	คุณสมบัติคำสำคัญ (Keyword)
A	a,d,e,f	h,i,j
B	b,c,d	f,g,h,i

ตัวอย่างที่ 2.1 การคำนวณหาค่าความคล้ายคลึงกันระหว่างเอกสาร A และ เอกสาร B

$S(A_{Title}, B_{Title}) = S_s(A_{Title}, B_{Title}) + S_c(A_{Title}, B_{Title})$

$S(A_{Title}, B_{Title})$ = (S_s({a,d,e,f}, {b,c,d}) + S_c({a,d,e,f}, {b,c,d}))

=[((4+3) / (2*6)) + (1 / 6)]

= [(7 / 12) + (1 / 6)]

= [0.583 + 0.167]

= 0.75

จากการหาค่าความคล้ายคลึงกันระหว่างเอกสาร A และเอกสาร B

ค่า Net Similarity ของคุณสมบัติชื่อเรื่อง เท่ากับ 0.75

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$S(A_{Keyword}, B_{Keyword}) = S_s(A_{Keyword}, B_{Keyword}) + S_c(A_{Keyword}, B_{Keyword})$$

$$S(A_{Keyword}, B_{Keyword}) = (S_s(\{h,i,j\}, \{f,g,h,i\}) + S_c(\{h,i,j\}, \{f,g,h,i\}))$$

$$= [((3+4) / (2*5)) + (2 / 5)]$$

$$= [(7 / 10) + (2 / 5)]$$

$$= [0.7 + 0.4]$$

$$= 1.1$$

จากการหาค่าความคล้ายคลึงกันระหว่างเอกสาร A และเอกสาร B
ค่า Net Similarity ของคุณสมบัติคำสำคัญ เท่ากับ 1.1

หาค่า Net Similarity ระหว่างเอกสาร A และเอกสาร B

$S(A, B) = 0.75+1.1 = 1.85$

ดังนั้นค่า Net Similarity ระหว่างเอกสาร A และเอกสาร B จากการเปรียบเทียบด้วย 2 คุณสมบัติ เท่ากับ 1.85 ซึ่งจะนำค่าที่ได้ไปพิจารณาต่อไป

ทฤษฎีต่างๆ ที่เกี่ยวข้องเหล่านี้ จะนำไปประยุกต์เป็นโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ (Text Simplified ARTMAP Neural Network) ซึ่งมีรายละเอียดในบทที่ 3

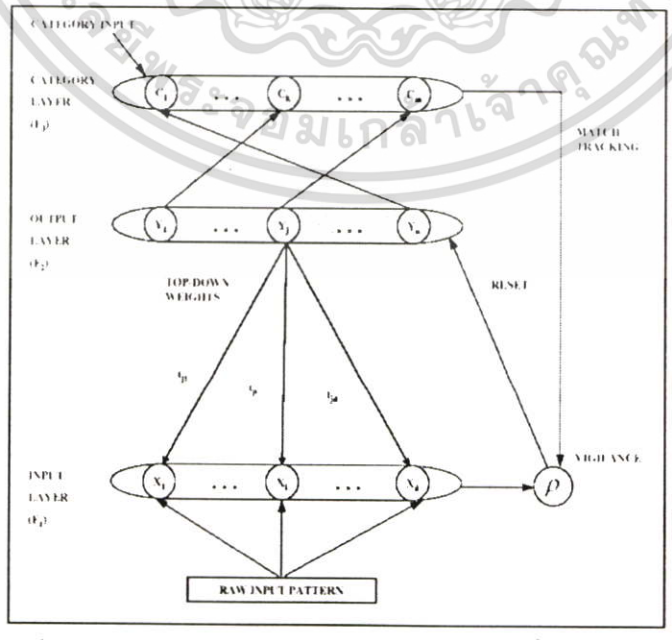
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ (Text Simplified ARTMAP Neural Network) เป็นการประยุกต์ใช้หลักการทำงานของโครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ (Simplified Fuzzy ARTMAP Neural Network) ซึ่งมีการเรียนรู้แบบมีการชี้นำ (Supervised Learning) ได้เพิ่มเติมเกี่ยวกับคุณสมบัติของค่าข้อมูลที่ใช้ เป็นค่าข้อมูลเชิงคุณภาพ (Qualitative Value) ซึ่งทำให้โมเดลนี้สามารถรับชุดข้อมูลเข้า ที่เป็นข้อมูลเชิงคุณภาพคือข้อความ (Text) ได้โดยตรง โดยไม่ต้องผ่านกระบวนการแปลงค่าข้อมูลเชิงคุณภาพไปเป็นค่าข้อมูลเชิงตัวเลข (Numerical Value) ซึ่งทำให้สูญเสียเวลาในขั้นตอนการแปลงค่าข้อมูล อีกทั้งยังนำแนวความคิดเกี่ยวกับการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็ค (Symbolic Objects) และหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี (Text Adaptive Resonance Theory Neural Network) มาประยุกต์ใช้งานกับโครงข่ายประสาทเทียมนี้

3.1 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ประกอบด้วยชั้นการทำงาน 3 ชั้น (Layer) คือ 1. ชั้นอินพุท (Input Layer หรือ F_1) 2. ชั้นเอาท์พุท (Output Layer หรือ F_2) และ 3. ชั้นแคททิกอรี่ (Category Layer หรือ F_3) ดังรูปที่ 3.1



รูปที่ 3.1 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนของชั้นอินพุท จะมีการรับชุดข้อมูลเข้า (Input Pattern) ที่เป็นข้อความ ซึ่งนิยามดังสมการที่ (3.1)

$$D = (D_1, D_2, ..., D_d) \tag{3.1}$$

โดยที่ d คือ จำนวนคุณสมบัติของ D

ระหว่างชั้นอินพุทและชั้นเอาท์พุท มีชุด Weights เรียกว่า Top-down Weight ซึ่งจะเชื่อมต่ออย่างสมบูรณ์และส่งข้อมูลระหว่างโหนดอินพุทที่ i ในชั้นอินพุท กับโหนดเอาท์พุทที่ j ในชั้นเอาท์พุท โดยที่ Top-down Weight จะเก็บค่าใน Weight เป็นค่าเชิงคุณภาพและค่าแสดงความเป็นสมาชิก (Degree) ของ Weight เพื่อให้โครงข่ายประสาทเทียม สามารถหาความคล้ายคลึงของชุดข้อมูลเข้าที่เป็นค่าเชิงคุณภาพได้ [5] ดังสมการที่ (3.2)

$$t_{ji} = \{(A_{1ji}, e_{1ji}), (A_{2ji}, e_{2ji}), ..., (A_{pji}, e_{pji})\} \tag{3.2}$$

กำหนดให้ A_{pji} คือค่าเชิงคุณภาพของ Top-down Weight และ e_{pji} คือค่าแสดงความเป็นสมาชิกของ A_{pji}

ค่า e_{pji} ของ A_{pji} ในโครงข่ายประสาทเทียมนี้ มีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งค่านี้ เป็นค่าที่ให้ระดับความเป็นสมาชิกของ A_{pji} กับชุดข้อมูลเข้า ถ้า e_{pji} ของ A_{pji} มีค่าเท่ากับ 0 ให้ความหมายว่าค่าเชิงคุณภาพ A_{pji} ไม่ได้เป็นส่วนหนึ่งของชุดข้อมูลเข้าลำดับที่ i ถ้า e_{pji} ของ A_{pji} มีค่าเท่ากับ 1 ให้ความหมายว่าค่าเชิงคุณภาพ A_{pji} เป็นสมาชิกของชุดข้อมูลเข้าลำดับที่ i อย่างสมบูรณ์

ในชั้นเอาท์พุท ประกอบด้วยชุดข้อมูลเข้าที่ถูกจำแนกไว้ด้วยกันเป็นกลุ่มหรือโหนดเอาท์พุท (Output Node) ซึ่งจำนวนโหนดเอาท์พุทมีได้ไม่จำกัด และจะเพิ่มจำนวนขึ้นเรื่อยๆ ในระหว่างการเรียนรู้ โหนดเอาท์พุทเหล่านี้จะเชื่อมต่อกับโหนดอินพุทในชั้นอินพุท ด้วยชุด Weight ดังกล่าว ในช่วงการเรียนรู้ที่ชั้นนี้ จะหาโหนดเอาท์พุทที่ชนะ โดยใช้หลักการหาค่าความคล้ายคลึงกันระหว่างชุดข้อมูลเข้า กับค่าเชิงคุณภาพของ Weight หลังจากนั้นจะนำโหนดเอาท์พุทที่ชนะมาคำนวณหาค่าความคล้ายคลึงกันระหว่าง Weight กับชุดข้อมูลเข้า ซึ่งค่านี้จะต้องมากกว่าหรือเท่ากับค่าวิจิแลนซ์ (Vigilance) แทนด้วย ρ ซึ่งเป็นค่าที่ควบคุมการจำแนกกลุ่มในชั้นเอาท์พุท ในที่นี้จะมีค่าอยู่ระหว่าง 0.25 ถึง 1 ค่านี้จะกำหนดเริ่มแรกเป็น Base-line Vigilance (แทนด้วย $\overline{\rho}$) เพราะจะมีการเพิ่มขึ้นเรื่อยๆ ในระหว่างการเรียนรู้ การที่กำหนดค่าวิจิแลนซ์ไว้สูง แสดงว่าชุดข้อมูลเข้า ต้องคล้ายคลึงกันมากถึงจะจำแนกไว้ในกลุ่มเดียวกันได้ ซึ่งจะทำให้โหนดเอาท์พุทมีขนาดเล็กและมีจำนวนโหนดมาก แต่ถ้ากำหนดค่าวิจิแลนซ์ไว้ต่ำ ชุดข้อมูลเข้าไม่ต้องคล้ายคลึงกันมากก็สามารถจำแนกไว้ในกลุ่มเดียวกันได้ ทำให้โหนดเอาท์พุทมีขนาดใหญ่และมีจำนวนโหนดน้อย ถ้าค่าของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โหนดเอาท์พุทที่ชนะ มีค่ามากกว่าหรือเท่ากับค่าวิจิแลนซ์ แสดงว่าระดับความคล้ายคลึงกันของการจำแนกกลุ่มข้อมูลมากพอ สามารถจัดให้ชุดข้อมูลเข้าชุดนั้นเป็นสมาชิกในโหนดเอาท์พุทนั้นได้

ส่วนของ Match Tracking ทำหน้าที่ควบคุมความถูกต้องในการเรียนรู้ว่า ชุดข้อมูลเข้าถูกจำแนกไว้ในโหนดเอาท์พุทที่ชี้ไปยังคำตอบในชั้นแคททิกอรี่ ตรงกับ Category Input ของชุดข้อมูลเข้า ที่กำลังเรียนรู้อยู่หรือไม่ ถ้าคำตอบไม่ตรงกันแสดงว่าชุดข้อมูลเข้าชุดนี้ไม่เหมาะสมที่จะจำแนกอยู่ในโหนดเอาท์พุทนี้ ถ้าเหตุการณ์นี้เกิดขึ้น ระบบจะกำหนดให้โหนดเอาท์พุทที่ชนะในขณะนั้นมีค่าเป็น 0 หรือค่าน้อยๆ เพื่อไม่ให้ระบบกลับมาเลือกโหนดเอาท์พุทนั้นอีก และเพิ่มค่าวิจิแลนซ์ขึ้นเล็กน้อย แล้วหาโหนดเอาท์พุทที่ชนะโหนดถัดไป และตรวจสอบตามขั้นตอนเดิมจนกว่าโหนดเอาท์พุทที่เลือก มีคำตอบตรงกับ Category Input ของชุดข้อมูลเข้า ถ้าคำตอบตรงกัน ถือว่าการจำแนกกลุ่มสำเร็จ เรียกว่าเกิด Resonance ทำการปรับปรุงค่า Top-down Weight เมื่อปรับปรุงเสร็จแสดงว่าระบบได้เรียนรู้ชุดข้อมูลเข้าชุดนั้นเรียบร้อยแล้ว

ที่ชั้นแคททิกอรี่ มีการส่ง Category Input พร้อมๆ กับชุดข้อมูลเข้า และในชั้นนี้จะบอกว่าคำตอบคือ Class ใด โดยการชี้คำตอบมาจากโหนดเอาท์พุท ในชั้นเอาท์พุท

3.1.1 หลักการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ มีการรับชุดข้อมูลเข้า และ Category Input ไปทำการคำนวณตามสมการ Activation Function คือ $Y_j = (\sum_{i=1}^{d} S(X_i, t_{ji}))/(\sum_{i=1}^{d} S(t_{ji}, t_{ji}))$ ของแต่ละโหนดเอาท์พุท เพื่อตัดสินว่าจะเก็บชุดข้อมูลเข้าไว้ที่โหนดเอาท์พุทใด แล้วเลือกโหนดเอาท์พุทที่ชนะเพื่อตรวจสอบว่าการจำแนกกลุ่มเหมาะสมหรือไม่ โดยมีหลักการคือ เลือกโหนดเอาท์พุทที่มีค่ามากที่สุด ถ้ามีโหนดเอาท์พุทที่ค่ามากที่สุดเท่ากัน ให้เลือกโดยเรียงจากซ้ายไปขวาคือ $Y_j = \max\{Y_j : j = 1..n\}$ เมื่อเลือกโหนดเอาท์พุทที่ชนะได้แล้ว นำมาคำนวณตามสมการ Match Function คือ $V = (\sum_{i=1}^{d} S(X_i, t_{ji}))/(\sum_{i=1}^{d} S(X_i, X_i))$ เพื่อตรวจสอบว่าค่า Weight ที่เก็บอยู่ในโหนดเอาท์พุทนั้น มีความคล้ายคลึงกันกับชุดข้อมูลเข้ามากพอหรือไม่ นำค่าที่ได้มาเปรียบเทียบกับค่าวิจิแลนซ์ ตามสมการ $V \geq \rho$ ถ้าน้อยกว่าค่าวิจิแลนซ์คือไม่ผ่านค่าวิจิแลนซ์ แสดงว่าค่า Weight ที่เก็บอยู่ในโหนดเอาท์พุทนั้น มีความคล้ายคลึงกันกับชุดข้อมูลเข้าไม่มากพอ ให้ยับยั้งโหนดเอาท์พุทนั้น แล้วหาโหนดเอาท์พุทที่มีค่ามากที่สุดโหนดถัดไปคำนวณค่าตามสมการ Match Function แล้วเปรียบเทียบกับค่าวิจิแลนซ์ จนกว่าจะมากกว่าหรือเท่ากับค่าวิจิแลนซ์ คือผ่านค่าวิจิแลนซ์ ถ้าไม่มีโหนดเอาท์พุทใดผ่านค่าวิจิแลนซ์ ให้สร้างโหนดเอาท์พุทใหม่ แต่ถ้าผ่านค่าวิจิแลนซ์ ให้ตรวจสอบว่า Class ที่ตอบซึ่งเก็บอยู่ในโหนดเอาท์พุทนั้นตรงกับ Category Input ที่รับเข้ามาหรือไม่ ถ้าไม่ตรงกัน จะมีการทำงานของ Match Tracking โดยจะเพิ่มค่าวิจิแลนซ์ เพื่อให้ค่าการ Match เพิ่มขึ้นเล็กน้อยตามสมการ $\rho = V + \varepsilon$ แล้วยับยั้ง (Reset)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โหนดนั้น หลังจากนั้นเลือกโหนดเอาท์พุทที่มีค่ามากที่สุดโหนดถัดไป คำนวณค่าตามสมการ Match Function แล้วเปรียบเทียบกับค่าวิจิแลนซ์อีกครั้ง ถ้าผ่านค่าวิจิแลนซ์ ให้ตรวจสอบว่า Class ที่ตอบตรงกันหรือไม่ ถ้าตรงกัน เกิด Resonance ทำการปรับปรุงค่า Top-down Weight ตามสมการต่อไปนี้

$$t_{Ji}^{(new)} = t_{Ji}^{(old)} \cap X$$

$$e_{nJi}^{(new)} = \begin{cases} f(e_{nJi}^{(old)} + \beta) & if \quad A_{nji} \in t_{Ji} \cap X, \\ f(e_{nJi}^{(old)} - \beta) & if \quad A_{nji} \notin t_{Ji} \cap X, \\ \beta_0 & otherwise \end{cases}$$

เมื่อปรับปรุงค่า Top-down Weight เสร็จแล้ว แสดงว่าระบบได้เรียนรู้ชุดข้อมูลเข้าชุดนั้นเรียบร้อยแล้ว ให้รับชุดข้อมูลเข้าชุดถัดไป จนกว่าจะพบเงื่อนไขในการจบการทำงาน

อัลกอริทึมสำหรับเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ มีรายละเอียด ดังรูปที่ 3.2

Step 0 กำหนดค่าพารามิเตอร์ต่าง ๆ ดังนี้
- ค่า Vigilance Parameter (ρ) อยู่ในช่วง 0.25 -1
- ค่า Learning Rate (β) ให้มีค่าน้อย ๆ
- ค่า Small Value (ε) ให้มีค่าน้อย ๆ

Step 1 ยังไม่ตรงเงื่อนไขในการหยุด (จำนวน Epoch) ทำ Step 2-14

Step 2 เลือกชุดข้อมูลเข้า

$$D = (D_1, D_2, ..., D_d) \qquad (3.3)$$

- ทำ Step 3-13

Step 3 ส่งค่าชุดข้อมูลเข้า ไปยังชั้น F_1 พร้อมทั้งส่งค่า Category Input ไปยังชั้น F_3

Step 4 ถ้าโหนดเอาท์พุทเท่ากับ 0
- ให้สร้างโหนดเอาท์พุทใหม่ เป็นโหนดที่ 1 ในชั้น F_2
- กำหนดให้ Top-down Weight ของโหนดเอาท์พุทใหม่เหมือนกับชุดข้อมูลเข้า และกำหนดค่า Degree เริ่มต้น
- กำหนดให้โหนดเอาท์พุทใหม่นี้เป็นตัวแทนของ Class ที่ชุดข้อมูลเข้าชุดนี้เป็นสมาชิก

Step 5 ถ้าโหนดเอาท์พุทไม่เท่ากับ 0 ทำ Step 6-13

Step 6 กำหนด Vigilance Parameter ให้เป็น Base-line Vigilance ($\overline{\rho}$)

รูปที่ 3.2 อัลกอริทึมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 7 คำนวณค่าตาม Activation Function ของแต่ละโหนดเอาท์พุท จนครบทุกโหนด

$$Y_j = (\sum_{i=1}^{d} S(X_i, t_{ji})) / (\sum_{i=1}^{d} S(t_{ji}, t_{ji})) \quad (3.4)$$

d คือ จำนวนของคุณสมบัติ

Step 8 เปรียบเทียบโหนดเอาท์พุท J ที่มีค่ามากที่สุดเป็นโหนดที่ชนะ (Winning Node)

$$Y_J = \max\{Y_j : j = 1..n\} \quad (3.5)$$

Step 9 คำนวณค่าตาม Match Function ที่เชื่อมต่อกับโหนดเอาท์พุท J

$$V = (\sum_{i=1}^{d} S(X_i, t_{Ji})) / (\sum_{i=1}^{d} S(X_i, X_i)) \quad (3.6)$$

d คือ จำนวนของคุณสมบัติ

Step 10 ทดสอบ Mismatch Reset (Vigilance Criterion)

$$V \geq \rho \quad (3.7)$$

- ถ้า $V < \rho$ ให้ยับยั้งโหนดเอาท์พุท J ทำ Step 8 ใหม่จนกว่า $V \geq \rho$
- ถ้า $V \geq \rho$ ทำ Step 11

Step 11 ตรวจสอบว่า Category หรือ Class ที่ตอบ ตรงกับ Category Input ที่รับเข้ามาหรือไม่

- ถ้าโหนดเอาท์พุทชี้ Class ตรงกับ Category Input
  - ทำ Step 13
  - เซต Indicator เป็น No Mismatch (Match)
- ถ้าโหนดเอาท์พุทชี้ Class ไม่ตรงกับ Category Input
  - ให้เพิ่มวิจิแลนซ์ ของโหนดเอาท์พุท J เล็กน้อยโดย

$$\rho = V + \varepsilon \quad (3.8)$$

  - ยับยั้งโหนดเอาท์พุท J
  - เซต Indicator เป็น Category Mismatch
  - ทำ Step 8-11 ต่อไปเรื่อยๆ จนกว่าจะครบทุกโหนด

Step 12 เมื่อเช็คจนครบทุกโหนดแล้วไม่ผ่าน Step 10 (Mismatch) หรือไม่ผ่าน Step 11 (Category Mismatch)

- ให้สร้าง Output Node ใหม่ ในชั้น F_2
- กำหนดให้ Top-down Weight ของโหนดเอาท์พุทใหม่เหมือนกับชุดข้อมูลเข้า และกำหนดค่า Degree เริ่มต้น
- กำหนดให้โหนดเอาท์พุทใหม่นี้เป็นตัวแทนของ Class ที่ชุดข้อมูลเข้าชุดนี้เป็นสมาชิก

รูปที่ 3.2 (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 13 ปรับปรุงค่า Top-down Weight ของโหนดเอาท์พุท J

$$t_{Ji}^{(new)} = t_{Ji}^{(old)} \cap X$$

$$e_{nJi}^{(new)} = \begin{cases} f(e_{nJi}^{(old)} + \beta) & if \quad A_{nji} \in t_{Ji} \cap X, \\ f(e_{nJi}^{(old)} - \beta) & if \quad A_{nji} \notin t_{Ji} \cap X, \\ \beta_0 & otherwise \end{cases} \tag{3.9}$$

- กำหนดให้ $f(.)$ นิยามดังนี้

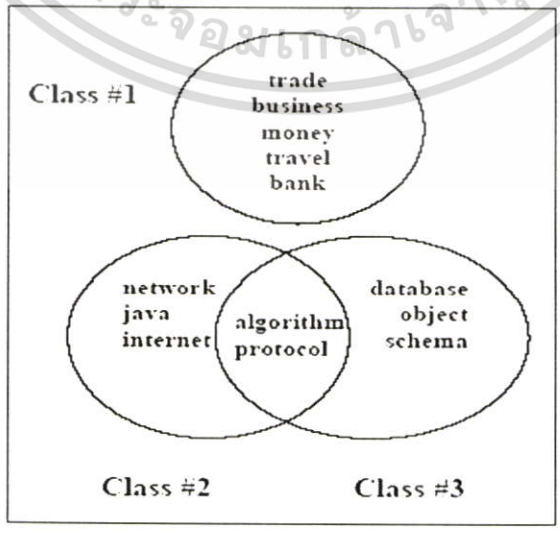
$$f(x) = \begin{cases} x & if \quad 0 \le x \le 1 \\ 0 & if \quad x < 0 \\ 1 & if \quad x > 1 \end{cases} \tag{3.10}$$

Step 14 ทดสอบเงื่อนไขในการหยุด ถ้าครบจำนวนรอบ (Epoch) ให้หยุด

รูปที่ 3.2 (ต่อ)

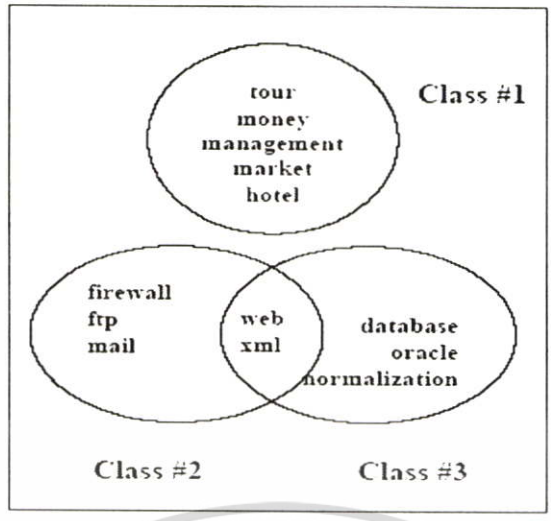
3.1.2 ตัวอย่างการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

ในหัวข้อนี้จะกล่าวถึง ขั้นตอนการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ซึ่งมีชุดข้อมูลที่สร้างจากกลุ่มคำ (Synthesized Text Document) 3 กลุ่ม โดยข้อมูลแต่ละคำเป็นตัวแทนของข้อมูลที่เป็นข้อความ (Text) ซึ่งมีคุณสมบัติชื่อเรื่อง (Title) และ คุณสมบัติคำสำคัญ (Keyword) ดังแสดงตามรูปที่ 3.3 และ รูปที่ 3.4



รูปที่ 3.3 แสดงกลุ่มของคำที่นำไปสร้างข้อมูลในคุณสมบัติชื่อเรื่อง

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การนำไปใช้ประโยชน์ด้านการศึกษาสามารถทำได้ แต่ต้องอ้างอิงถึงเจ้าของเอกสาร ห้ามมิให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 แสดงกลุ่มของคำที่นำไปสร้างข้อมูลในคุณสมบัติคำสำคัญ

กลุ่มคำที่ใช้สร้างข้อมูลเอกสารจากคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ โดยสร้างจากข้อมูลแสดงลักษณะเฉพาะ (Profile) รูปที่ 3.3 และรูปที่ 3.4 มีรายละเอียดของข้อมูลแต่ละ Class บางส่วนตามตารางที่ 3.1

ตารางที่ 3.1 แสดงตัวอย่างของข้อมูล ที่นำมาใช้ในการเรียนรู้ของโครงข่ายประสาทเทียมนี้

เอกสาร	คุณสมบัติชื่อเรื่อง	คุณสมบัติคำสำคัญ	Category Input
D_1	"protocol" "java" "network"	"mail" "web" "ftp"	category 2
D_2	"protocol" "object" "schema"	"xml" "web" "oracle" "database"	category 3
D_3	"internet" "java"	"xml" "web" "mail"	category 2

Step 0 - กำหนด Learning Rate (β) = 0.02

- กำหนด Vigilance Parameter (ρ) = 0.4
- กำหนด Small Value (ε) = 0.02
- กำหนดค่า Degree เริ่มต้น = 0.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 1 ยังไม่ตรงเงื่อนไขในการหยุด ทำ Step 2-14

Step 2 เลือกชุดข้อมูลเข้าลำดับที่ 1

D_{Title} = {"protocol", "java", "network"}

$D_{Keyword}$ = {"mail", "web", "ftp"}

- ทำ Step 3-13

Step 3 ส่งค่าชุดข้อมูลเข้า ลำดับที่ 1 และ Category Input

X_{Title} = {"protocol", "java", "network"}

$X_{Keyword}$ = {"mail", "web", "ftp"}

Category Input = category 2

Step 4 โหนดเอาท์พุท เท่ากับ 0

- สร้างโหนดเอาท์พุทใหม่ โหนดที่ 1 คือ Y_1
- กำหนดให้ Y_1 เป็นตัวแทนของ Class ที่ 2
- ให้ Top-down Weight เท่ากับชุดข้อมูลเข้าชุดนี้ และกำหนดค่า Degree เริ่มต้นให้กับแต่ละค่าของ Top-down Weight ซึ่งค่าสมาชิกใน Top-down Weight เป็นดังนี้

t_{11} = {(protocol, 0.50), (java, 0.50), (network, 0.50)}

t_{12} = {(mail, 0.50), (web, 0.50), (ftp, 0.50)}

Step 14 ตรวจสอบเงื่อนไขในการหยุด ไม่ตรงเงื่อนไขในการหยุด ให้รับชุดข้อมูลเข้าชุดใหม่

Step 1 ยังไม่ตรงเงื่อนไขในการหยุด ทำ Step 2-14

Step 2 เลือกชุดข้อมูลเข้าลำดับที่ 2

D_{Title} = {"protocol", "object", "schema"}

$D_{Keyword}$ = {"xml", "web", "oracle", "database"}

- ทำ Step 3-13

Step 3 ส่งค่าชุดข้อมูลเข้า ลำดับที่ 2 และ Category Input

X_{Title} = {"protocol", "object", "schema"}

$X_{Keyword}$ = {"xml", "web", "oracle", "database" }

Category Input = category 3

Step 4 ไม่มีการทำงาน ทำ Step 5

Step 5 โหนดเอาท์พุทไม่เท่ากับ 0 (เท่ากับ 1) ทำ Step 6-13

Step 6 Vigilance Parameter (ρ) = Base-line Vigilance ($\overline{\rho}$) = 0.4

Step 7 คำนวณค่าตาม Activation Function ของแต่ละโหนดเอาท์พุท จนครบทุกโหนด

$$Y_1 = [S(X_{Title}, t_{11}) + S(X_{Keyword}, t_{12})]/[S(t_{11}, t_{11}) + S(t_{12}, t_{12})]$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พิจารณาสมการที่ (3.4) คำนวณค่าตาม Activation Function ของแต่ละโหนดเอาท์พุท

X_{Title} = {“protocol”, “object”, “schema”}

t_{11} = {“protocol”, “java”, “network”}

$S(X_{Title}, t_{11})$ =

[(S_{span} ({“protocol”, “object”, “schema”}, {“protocol”, “java”, “network”})) +

($S_{content}$ ({“protocol”, “object”, “schema”}, {“protocol”, “java”, “network”}))]

= [((3+3) / (2*5)) + (1 / 5)]

= [(6 / 10) + (1 / 5)]

= [0.6 + 0.2]

= 0.8

$X_{Keyword}$ = {“xml”, “web”, “oracle”, “database”}

t_{12} = {“mail”, “web”, “ftp”}

$S(X_{Keyword}, t_{12})$ =

[(S_{span} ({“xml”, “web”, “oracle”, “database”}, {“mail”, “web”, “ftp”})) +

($S_{content}$ ({“xml”, “web”, “oracle”, “database”}, {“mail”, “web”, “ftp”}))]

= [((4+3) / (2*6)) + (1 / 6)]

= [(7 / 12) + (1 / 6)]

= [0.583 + 0.166]

= 0.749

$S(X, t)$ = 0.8 + 0.749

= 1.549

t_{11} = {“protocol”, “java”, “network”}

$S(t_{11}, t_{11})$ =

[(S_{span} ({“protocol”, “java”, “network”}, {“protocol”, “java”, “network”})) +

($S_{content}$ ({“protocol”, “java”, “network”}, {“protocol”, “java”, “network”}))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

t_{12} = {“mail”, “web”, “ftp”}

$S(t_{12}, t_{12})$ =

เอกสารนี้เป็นสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$[(S_{span}(\{\text{“mail”, “web”, “ftp”}\}, \{\text{“mail”, “web”, “ftp”}\})) +$

$(S_{content}(\{\text{“mail”, “web”, “ftp”}\}, \{\text{“mail”, “web”, “ftp”}\}))]$

$= [((3+3) / (2*3)) + (3 / 3)]$

$= [(6 / 6) + (3 / 3)]$

$= [1 + 1]$

$= 2$

$S(t,t) = 2 + 2$

$= 4$

$Y_1 = S(X,t) / S(t,t) = 1.549 / 4$

$= \mathbf{0.387}$

Step 8 ในขณะนี้มีโหนดเอาท์พุท เพียงโหนดเดียวคือ Y_1 ดังนั้น Y_j (โหนดที่มีค่ามากที่สุด) คือ Y_1

Step 9 คำนวณค่าตาม Match Function ที่เชื่อมต่อกับโหนดเอาท์พุทที่ชนะ คือ Y_1

$V = [S(X_{Title}, t_{11}) + S(X_{Keyword}, t_{12})] / [S(X_{Title}, X_{Title}) + (X_{Keyword}, X_{Keyword})]$

พิจารณาสมการที่ (3.6) คำนวณค่าตาม Match Function

X_{Title} = {“protocol”, “object”, “schema”}

t_{11} = {“protocol”, “java”, “network”}

$S(X_{Title}, t_{11})$ =

$[(S_{span}(\{\text{“protocol”, “object”, “schema”}\}, \{\text{“protocol”, “java”, “network”}\})) +$

$(S_{content}(\{\text{“protocol”, “object”, “schema”}\}, \{\text{“protocol”, “java”, “network”}\}))]$

$= [((3+3) / (2*5)) + (1 / 5)]$

$= [(6 / 10) + (1 / 5)]$

$= [0.6 + 0.2]$

$= 0.8$

$X_{Keyword}$ = {“xml”, “web”, “oracle”, “database”}

t_{12} = {“mail”, “web”, “ftp”}

$S(X_{Keyword}, t_{12})$ =

$[(S_{span}(\{\text{“xml”, “web”, “oracle”, “database”}\}, \{\text{“mail”, “web”, “ftp”}\})) +$

$(S_{content}(\{\text{“xml”, “web”, “oracle”, “database”}\}, \{\text{“mail”, “web”, “ftp”}\}))]$

$= [((4+3) / (2*6)) + (1 / 6)]$

$= [(7 / 12) + (1 / 6)]$

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

= [0.583 + 0.166]

= 0.749

$S(X,t)$ = 0.8 + 0.749

= 1.549

X_{Title} = {“protocol”, “object”, “schema”}

$S(X_{Title}, X_{Title})$ =

[(S_{span} ({“protocol”, “object”, “schema”}, {“protocol”, “object”, “schema”})) +

($S_{content}$ ({“protocol”, “object”, “schema”}, {“protocol”, “object”, “schema”}))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

$X_{Keyword}$ = {“xml”, “web”, “oracle”, “database”}

$S(X_{Keyword}, X_{Keyword})$ =

[(S_{span} ({“xml”, “web”, “oracle”, “database”}, ({“xml”, “web”, “oracle”, “database”})) +

($S_{content}$ ({“xml”, “web”, “oracle”, “database”}, ({“xml”, “web”, “oracle”, “database”}))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

$S(X,X)$ = 2 + 2

= 4

$V = S(X,t) / S(X,X)$ = 1.549 / 4

= 0.387

Step 10 ทดสอบ Mismatch Reset พิจารณาสมการที่ (3.7)

$V < \rho$, 0.387 < 0.4

- ให้ยับยั้งโหนดเอาท์พุท Y_1 โดย Y_1 = -99
- ทำ Step 8 ใหม่จนกว่า $V \geq \rho$ แต่เนื่องจากโหนดเอาท์พุท มีเพียงโหนดเดียวคือ Y_1 จึงไม่ผ่าน Step 10 (Mismatch)
- ทำ Step 12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 12 เช็คจนครบทุกโหนดแล้ว ไม่ผ่าน Step 10 (Mismatch) หรือ Step 11 (Category Mismatch)

- สร้างโหนดเอาท์พุทใหม่ โหนดที่ 2 คือ Y_2
- กำหนดให้ Y_2 เป็นตัวแทนของ Class ที่ 3
- ให้ Top-down Weight เท่ากับชุดข้อมูลเข้าชุดนี้ และกำหนดค่า Degree เริ่มต้นให้กับแต่ละค่าของ Top-down Weight ซึ่งค่าสมาชิกใน Top-down Weight เป็นดังนี้

t_{21} = {(protocol, 0.50), (object, 0.50), (schema, 0.50)}

t_{22} = {(xml, 0.50), (web, 0.50), (oracle, 0.50) , (database, 0.50)}

Step 14 ตรวจสอบเงื่อนไขในการหยุด ไม่ตรงเงื่อนไขในการหยุด ให้รับชุดข้อมูลเข้าชุดใหม่

Step 1 ยังไม่ตรงเงื่อนไขในการหยุด ทำ Step 2-14

Step 2 เลือกชุดข้อมูลเข้าลำดับที่ 3

D_{Title} = {"internet", "java"}

$D_{Keyword}$ = {"xml", "web", "mail" }

- ทำ Step 3-13

Step 3 ส่งค่าชุดข้อมูลเข้า ลำดับที่ 3 และ Category Input

X_{Title} = {"internet", "java"}

$X_{Keyword}$ = {"xml", "web", "mail" }

Category Input = category 2

Step 4 ไม่มีการทำงาน ทำ Step 5

Step 5 โหนดเอาท์พุทไม่เท่ากับ 0 (เท่ากับ = 2) ทำ Step 6-13

Step 6 Vigilance Parameter (ρ) = Base-line Vigilance ($\bar{\rho}$) = 0.4

Step 7 คำนวณค่าตาม Activation Function ของแต่ละโหนดเอาท์พุท จนครบทุกโหนด

$$Y_1 = [S(X_{Title}, t_{11}) + S(X_{Keyword}, t_{12})] / [S(t_{11}, t_{11}) + S(t_{12}, t_{12})]$$

คำนวณค่าตาม Activation Function ของโหนดเอาท์พุท Y_1

X_{Title} = {"internet", "java"}

t_{11} = {"protocol", "java", "network"}

$S(X_{Title}, t_{11})$ =

[(S_{span} ({"internet", "java"}, {"protocol", "java", "network"})) +

($S_{content}$ ({"internet", "java"}, {"protocol", "java", "network"}))]

= [((2+3) / (2*4)) + (1 / 4)]

= [(5 / 8) + (1 / 4)]

= [0.625 + 0.25]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

= 0.875

$X_{Keyword}$ = {“xml”, “web”, “mail” }

t_{12} = {“mail”, “web”, “ftp”}

$S(X_{Keyword}, t_{12})$ =

[(S_{span} ({“xml”, “web”, “mail” }, {“mail”, “web”, “ftp”})) +

($S_{content}$ ({“xml”, “web”, “mail” }, {“mail”, “web”, “ftp”}))]

= [((3+3) / (2*4)) + (2 / 4)]

= [(6 / 8) + (2 / 4)]

= [0.75 + 0.5]

= 1.25

$S(X,t)$ = 0.875 + 1.25

= 2.125

t_{11} = {“protocol”, “java”, “network”}

$S(t_{11}, t_{11})$ =

[(S_{span} ({“protocol”, “java”, “network”}, {“protocol”, “java”, “network”})) +

($S_{content}$ ({“protocol”, “java”, “network”}, {“protocol”, “java”, “network”}))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

t_{12} = {“mail”, “web”, “ftp”}

$S(t_{12}, t_{12})$ =

[(S_{span} ({“mail”, “web”, “ftp”}, {“mail”, “web”, “ftp”})) +

($S_{content}$ ({“mail”, “web”, “ftp”}, {“mail”, “web”, “ftp”}))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

$S(t,t)$ = 2 + 2

= 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$Y_1 = S(X,t) / S(t,t) = 2.125 / 4$

= 0.531

คำนวณค่าตาม Activation Function ของโหนดเอาท์พุท Y_2

X_{Title} = {“internet”, “java”}

t_{21} = {“protocol”, “object”, “schema”}

$S(X_{Title}, t_{21})$ =

[(S_{span} ({“internet”, “java”}, {“protocol”, “object”, “schema”})) +

($S_{content}$ ({“internet”, “java”}, {“protocol”, “object”, “schema”}))]

= [((2+3) / (2*5)) + (0 / 5)]

= [(5 / 10) + (0 / 5)]

= [0.5 + 0]

= 0.5

$X_{Keyword}$ = {“xml”, “web”, “mail” }

t_{22} = {“xml”, “web”, “oracle”, “database” }

$S(X_{Keyword}, t_{22})$ =

[(S_{span} ({“xml”, “web”, “mail” }, {“xml”, “web”, “oracle”, “database” })) +

($S_{content}$ ({“xml”, “web”, “mail” }, {“xml”, “web”, “oracle”, “database” }))]

= [((3+4) / (2*5)) + (2 / 5)]

= [(7 / 10) + (2 / 5)]

= [0.7 + 0.4]

= 1.1

$S(X,t)$ = 0.5 + 1.1

= 1.6

t_{21} = {“protocol”, “object”, “schema”}

$S(t_{21}, t_{21})$ =

[(S_{span} ({“protocol”, “object”, “schema”}, {“protocol”, “object”, “schema”})) +

($S_{content}$ ({“protocol”, “object”, “schema”}, {“protocol”, “object”, “schema”}))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

t_{22} = {“xml”, “web”, “oracle”, “database” }

$S(t_{22}, t_{22})$ =

[(S_{span} ({“xml”, “web”, “oracle”, “database” }, {“xml”, “web”, “oracle”, “database” })) +

($S_{content}$ ({“xml”, “web”, “oracle”, “database” }, {“xml”, “web”, “oracle”, “database” }))]

= [((4+4) / (2*4)) + (4 / 4)]

= [(8 / 8) + (4 / 4)]

= [1 + 1]

= 2

$S(t,t)$ = 2 + 2

= 4

$Y_2 = S(X,t) / S(t,t) = 1.6 / 4$

= 0.4

Step 8 เลือกโหนดเอาท์พุทที่มีค่ามากที่สุด คือ Y_1

Step 9 คำนวณค่าตาม Match Function ที่เชื่อมต่อกับโหนดที่ชนะ คือ Y_1

$V = [S(X_{Title}, t_{11}) + S(X_{Keyword}, t_{12})] / [S(X_{Title}, X_{Title}) + (X_{Keyword}, X_{Keywprd})]$

คำนวณค่าตาม Match Function

X_{Title} = {“internet”, “java”}

t_{11} = {“protocol”, “java”, “network”}

$S(X_{Title}, t_{11})$ =

[(S_{span} ({“internet”, “java”}, {“protocol”, “java”, “network”})) +

($S_{content}$ ({“internet”, “java”}, {“protocol”, “java”, “network”}))]

= [((2+3) / (2*4)) + (1 / 4)]

= [(5 / 8) + (1 / 4)]

= [0.625 + 0.25]

= 0.875

$X_{Keyword}$ = {“xml”, “web”, “mail” }

t_{12} = {“mail”, “web”, “ftp”}

$S(X_{Keyword}, t_{12})$ =

[(S_{span} ({“xml”, “web”, “mail” }, {“mail”, “web”, “ftp”})) +

($S_{content}$ ({“xml”, “web”, “mail” }, {“mail”, “web”, “ftp”}))]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

= [((3+3) / (2*4)) + (2 / 4)]

= [(6 / 8) + (2 / 4)]

= [0.75 + 0.5]

= 1.25

$S(X,t)$ = 0.875 + 1.25

= 2.125

X_{Title} = {“internet”, “java”}

$S(X_{Title}, X_{Title})$ =

[(S_{span} ({“internet”, “java”}, {“internet”, “java”})) +

($S_{content}$ ({“internet”, “java”}, {“internet”, “java”}))]

= [((2+2) / (2*2)) + (2 / 2)]

= [(4 / 4) + (2 / 2)]

= [1 + 1]

= 2

$X_{Keyword}$ = {“xml”, “web”, “mail” }

$S(X_{Keyword}, X_{Keyword})$ =

[(S_{span} ({“xml”, “web”, “mail” }, {“xml”, “web”, “mail” })) +

($S_{content}$ ({“xml”, “web”, “mail” }, {“xml”, “web”, “mail” }))]

= [((3+3) / (2*3)) + (3 / 3)]

= [(6 / 6) + (3 / 3)]

= [1 + 1]

= 2

$S(X,X)$ = 2 + 2

= 4

$V = S(X,t)/S(X,X)$ = 2.125 / 4

= 0.531

Step 10 ทดสอบ Mismatch Reset

- $V \geq \rho$, 0.531 $\geq$ 0.4
- ไม่มีการยับยั้งโหนดเอาท์พุท Y_j ทำ Step 11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 11 ตรวจสอบว่า Class ที่ตอบ ตรงกับ Category Input หรือไม่

- เนื่องจาก Y_1 เป็นตัวแทนของ Class ที่ 2 และ Category Input ที่รับเข้ามาของชุดข้อมูลเข้าชุดนี้คือ category 2
- เพราะฉะนั้น Category ตรงกัน
- ทำ Step 13

Step 13 ปรับปรุงค่า Top-down Weight ของโหนดเอาท์พุทลำดับที่ 1 คือ Y_1

Title และ Keyword ของชุดข้อมูลเข้า ชุดที่ 3 คือ

X_{Title} = {“internet”, “java”}

$X_{Keyword}$ = {“xml”, “web”, “mail” }

ค่าสมาชิกใน Top-down Weight ชุดเดิมเป็นดังนี้

t_{11} = {(protocol, 0.50), (java, 0.50), (network, 0.50)}

t_{12} = {(mail, 0.50), (web, 0.50), (ftp, 0.50)}

หลังจากปรับปรุง Top-down Weight แล้ว ได้ค่าสมาชิกใน Top-down Weight ชุดใหม่ดังนี้

t_{11} = {(protocol, 0.48), (java, 0.52), (network, 0.48), (internet, 0.10)}

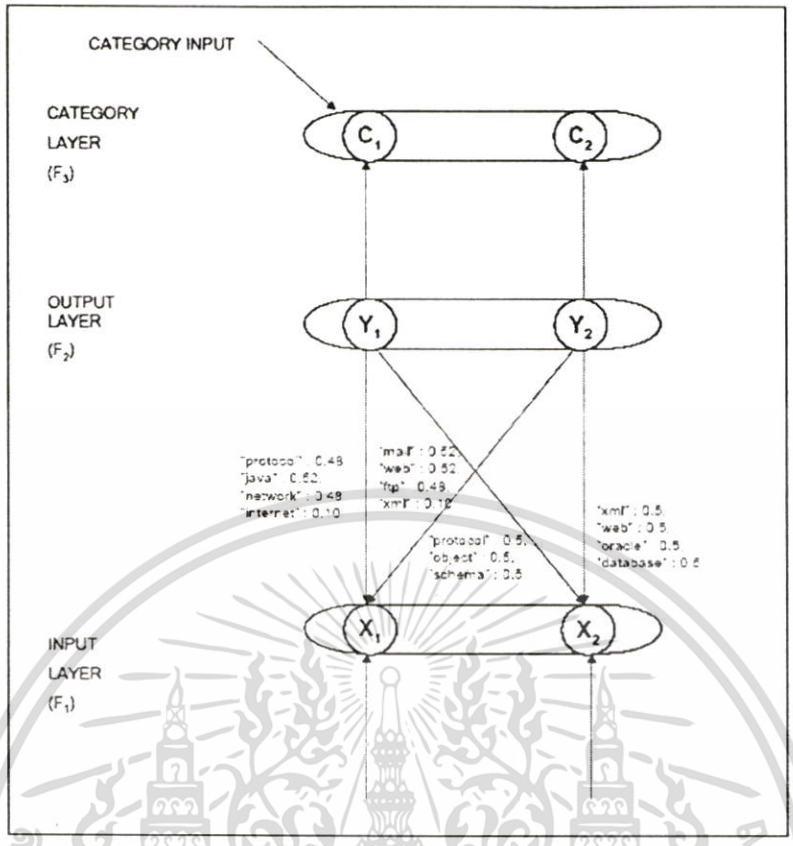
t_{12} = {(mail, 0.52), (web, 0.52), (ftp, 0.48), (xml, 0.10)}

Step 14 ตรวจสอบเงื่อนไขในการหยุด ไม่ตรงเงื่อนไขในการหยุด ให้รับชุดข้อมูลเข้าชุดใหม่ต่อไปจนครบทุกชุด และครบจำนวนรอบของการเรียนรู้ จึงหยุดการทำงาน

เมื่อโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ เรียนรู้จากข้อมูลสำหรับเรียนรู้ 3 ชุด จำนวน 1 รอบ เรียบร้อยแล้ว จะได้ค่า Top-down Weight แสดงดังรูปที่ 3.5

หลังจากที่สร้างและทดสอบการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ เรียบร้อยแล้ว ขั้นตอนต่อไปคือนำโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พไปทดลองกับข้อมูลที่ใช้งานจริง ซึ่งในบทที่ 4 จะกล่าวถึง การทดลองและผลการทดลอง พร้อมทั้งการวัดประสิทธิภาพการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ โดยข้อมูลที่ใช้ในการทดลองมี 4 ชุดข้อมูล ได้แก่ 1.ชุดข้อมูลที่สร้างจากกลุ่มตัวอักษรภาษาอังกฤษ จำนวน 3 กลุ่ม 2.ชุดข้อมูล CSTR [13] จำนวน 4 กลุ่มเอกสาร 3.ชุดข้อมูล Reuters-Top10 [14] จำนวน 10 กลุ่มเอกสาร และ 4.ชุดข้อมูล K-dataset [15] จำนวน 20 กลุ่มเอกสาร ซึ่งข้อมูลแต่ละชุดประกอบด้วยคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ รายละเอียดของการทดลองและผลการทดลองดังจะได้กล่าวต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.5 แสดงค่า Top-down Weight ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พหลังจากการเรียนรู้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

การทดลองและผลการทดลอง

4.1 การวัดประสิทธิภาพการทำงานของโมเดล

การวัดประสิทธิภาพการทำงานของโมเดล ได้ใช้ตัววัดประสิทธิภาพที่ใช้ในงานวิจัยทั่วไป คือตัววัดประสิทธิภาพอัตราความถูกต้อง (Accuracy Rate) และตัววัดประสิทธิภาพในงานวิจัยทางด้านการค้นคืนสารสนเทศ (Information Retrieval) คือตัววัดประสิทธิภาพ F-Measure และตัววัดประสิทธิภาพ Entropy โดยมีรายละเอียดของแต่ละตัวชี้วัดดังต่อไปนี้ [1] และ [8]

4.1.1 ตัววัดประสิทธิภาพอัตราความถูกต้อง (Accuracy Rate)

ตัววัดประสิทธิภาพอัตราความถูกต้อง มีรายละเอียดดังสมการที่ (4.1)

$$r = [(\sum_{i=1}^{c} doc_i) / n] * 100 \qquad (4.1)$$

โดยที่ c คือ จำนวนของข้อมูลที่ตอบถูกต้อง

n คือ จำนวนของข้อมูลทั้งหมด

4.1.2 ตัววัดประสิทธิภาพ F-Measure

ตัววัดประสิทธิภาพ F-Measure ใช้ในการวัดความถูกต้องของการจำแนกกลุ่มเอกสาร โดยจะมีการคำนวณหาค่า Recall และ Precision ค่า F-Measure จะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยที่ค่านี้จะบอกว่าผลการจำแนกกลุ่มเอกสารมีความถูกต้องมากน้อยเพียงใด ถ้าค่าที่คำนวณได้มีค่าน้อยแสดงว่ามีความถูกต้องต่ำ ถ้าค่าที่คำนวณได้มีค่าเป็น 1 แสดงว่ามีความถูกต้องสูงสุด การคำนวณหาค่า F-Measure $F(i)$ ของ Class ลำดับที่ i สามารถคำนวณได้จากสมการที่จะกล่าวต่อไป แต่ก่อนจะคำนวณค่า $F(i)$ จะต้องคำนวณหาค่า Recall และ Precision ซึ่งส่วนที่นำมาหาค่า Recall และ Precision พิจารณาจากรูปที่ 4.1

สมการที่ใช้สำหรับหาค่า Recall และ Precision คือสมการที่ (4.2) และ (4.3) ตามลำดับ

$$Recall(i) = \frac{|Ra_i|}{|R_i|} \qquad (4.2)$$

$$Precision(i) = \frac{|Ra_i|}{|A_i|} \qquad (4.3)$$

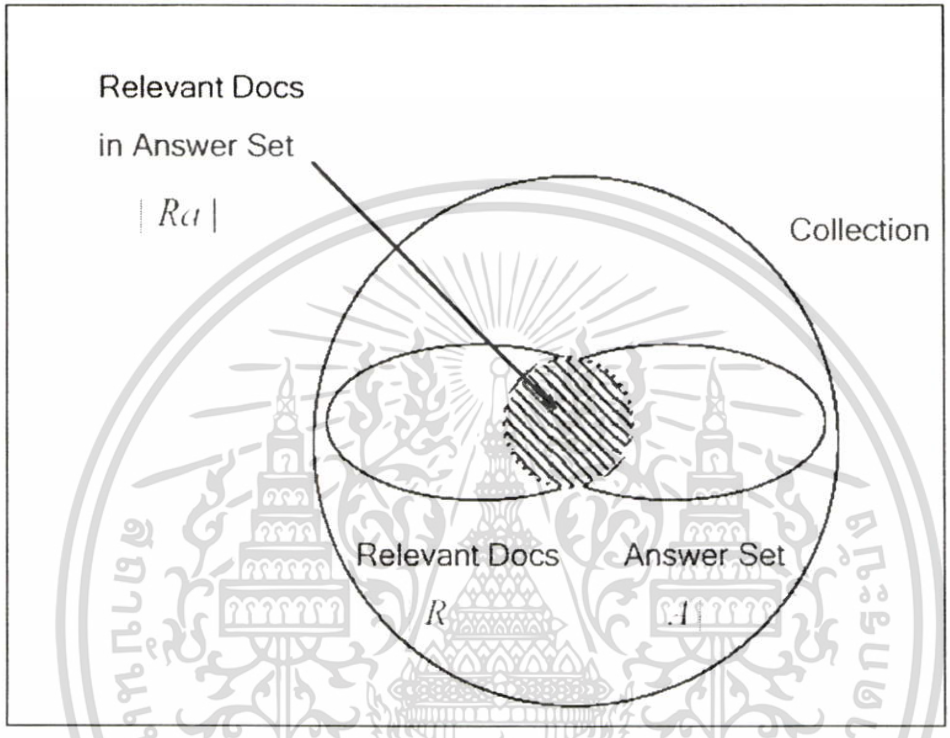
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่

$|Ra_i|$ คือ จำนวนของเอกสารจาก Category ลำดับที่ i ที่เป็นสมาชิกใน Class ลำดับที่ i

$|R_i|$ คือ จำนวนของเอกสารที่นำมาใช้ในการทดลอง ที่เป็นสมาชิกใน Class ลำดับที่ i

$|A_i|$ คือ จำนวนของเอกสารจากผลการทดลองที่อยู่ใน Category ลำดับที่ i



รูปที่ 4.1 แสดงส่วนที่ใช้วัดค่า Precision และ Recall ในการร้องขอสารสนเทศ [8]

การคำนวณหาค่า $F(i)$ ของ Class ลำดับที่ i สามารถคำนวณได้จากสมการที่ (4.4)

$$F(i) = \frac{2 * Recall(i) * Precision(i)}{Recall(i) + Precision(i)} \tag{4.4}$$

การคำนวณค่า F-Measure ของทุกๆ Class สามารถคำนวณได้จากสมการที่ (4.5)

$$F = \sum_{i=1}^{n} \frac{|R_i|}{N} F(i) \tag{4.5}$$

โดยที่ n คือ จำนวนของ Class

N คือ จำนวนเอกสารทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 ตัววัดประสิทธิภาพ Entropy

ตัววัดประสิทธิภาพ Entropy ใช้ในการวัดการซ้อนทับกันของกลุ่มเอกสาร จะมีค่าอยู่ระหว่าง 0 ถึง $\log_2$ (C) เมื่อ C คือจำนวนกลุ่มของเอกสาร ถ้ามีจำนวนกลุ่มเอกสาร 4 กลุ่ม ค่านี้จะอยู่ระหว่าง 0 ถึง 2 จำนวนกลุ่มเอกสาร 10 กลุ่ม ค่าจะอยู่ระหว่าง 0 ถึง 3.3219 และจำนวนกลุ่มเอกสาร 20 กลุ่ม ค่าจะอยู่ระหว่าง 0 ถึง 4.3219 โดยที่ค่านี้จะบอกว่าผลการจำแนกกลุ่มเอกสารมีการซ้อนทับกันมากน้อยเพียงใด ถ้าค่าที่คำนวณได้มีค่าเป็น 0 แสดงว่าไม่มีการซ้อนทับกันของเอกสารเลย การคำนวณหาค่า Entropy โดยการคำนวณหาค่าความน่าจะเป็นที่สมาชิกจากผลการทดลองของ Category ลำดับที่ i ที่อยู่ในแต่ละ Class j ซึ่งแทนความน่าจะเป็นนี้ด้วย p_{ij} และต้องคิดทุกๆ Class ที่นำมาใช้ในการทดสอบ ซึ่งค่า Entropy ของแต่ละ Class หาได้จากสมการที่ (4.6)

$$E_i = -\sum_{j=1}^{n} p_{ij} \log_2 (p_{ij}) \tag{4.6}$$

โดยที่ n คือ จำนวนของ Class

ค่า Entropy ของทุกๆ Class หาได้จากสมการที่ (4.7)

$$E = \sum_{i=1}^{n} \frac{|A_i| * E_i}{N} \tag{4.7}$$

โดยที่

$|A_i|$ คือ จำนวนเอกสารจากผลการทดลองที่อยู่ใน Category ลำดับที่ i

n คือ จำนวนของ Class

N คือ จำนวนเอกสารทั้งหมด

4.2 ชุดข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลองมี 2 แบบ คือ ข้อมูลที่สร้างจากกลุ่มตัวอักษรภาษาอังกฤษ โดยการสุ่มจากตัวอักษรตามข้อมูลแสดงลักษณะเฉพาะ (Profile) ที่กำหนดขึ้นเอง เรียกว่าข้อมูลสังเคราะห์ (Synthesized Dataset) เพื่อนำมาใช้ทดสอบการทำงานของโมเดล ก่อนนำไปใช้กับข้อมูลจริง และข้อมูลที่ใช้งานจริง ซึ่งเป็นข้อมูลที่ใช้ในการจำแนกประเภทเอกสาร ข้อมูลทุกชุดใช้คุณสมบัติชื่อเรื่อง และคุณสมบัติคำสำคัญ ข้อมูลแบบที่ 2 ที่ใช้ในการทดลองมีจำนวน 3 ชุดข้อมูลคือ 1.ชุดข้อมูล CSTR [13] 2.ชุดข้อมูล Reuters-Top10 [14] และ 3.ชุดข้อมูล K-dataset [15] ชุดข้อมูลทั้ง 3 ชุด เก็บ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อยู่ในรูปแบบแฟ้มข้อมูล HTML และ XML การที่จะนำข้อมูลเหล่านี้ไปใช้งานจะต้องมีการเตรียมข้อมูลก่อนนำเข้าสู่โมเดล ซึ่งรายละเอียดของขั้นตอนการเตรียมข้อมูลเป็นดังนี้

ขั้นตอนที่ 1 ตัดแท็ก แล้วแยกเอาเฉพาะข้อความที่เป็นข้อมูลของชื่อเรื่อง (Title) และข้อความในส่วนของข้อมูลที่เป็นคำสำคัญ (Keyword) ข องทุกๆ เอกสาร ซึ่งจะได้ข้อมูลคุณสมบัติชื่อเรื่อง และคุณสมบัติคำสำคัญ ของเอกสาร

ขั้นตอนที่ 2 นำข้อมูลในส่วนของคุณสมบัติคำสำคัญ ไปแยกเอาคำสำคัญของแต่ละเอกสาร ด้วยโปรแกรม Copernic Summarizer Version 2.1 โดยการหาคำสำคัญของเอกสารแต่ละชุด ได้กำหนดจำนวนคำไว้ที่ 30 คำ และตัดเป็นคำเดี่ยวไม่เกิน 50 คำ เพื่อนำเข้าสู่โมเดล

ขั้นตอนที่ 3 นำข้อมูลในส่วนคุณสมบัติชื่อเรื่อง ไปตัด Stopwords โดยใช้มาตรฐาน [16]

ขั้นตอนที่ 4 นำข้อมูล ทั้งคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ ไปหา Word Stemming โดยใช้ Porter's Algorithm [17] ซึ่งเป็นอัลกอริทึมสำหรับทำ Word Stemming ที่นิยมใช้กันอย่างแพร่หลาย โดยจะตัดส่วนที่เป็น Suffixes ของคำออก

ขั้นตอนที่ 5 หลังจากนั้นนำข้อมูลที่เตรียมเสร็จเรียบร้อยแล้ว ไปทำการแบ่งข้อมูลสำหรับให้โครงข่ายประสาทเทียมเรียนรู้ (Training) และทดสอบ (Testing) และทำให้อยู่ในรูปแบบที่สามารถนำเข้าสู่โครงข่ายประสาทเทียมได้ ด้วยโปรแกรม MATLAB Version 6.5 ในการเรียนรู้ของโครงข่ายประสาทเทียมนี้ ได้แบ่งข้อมูลสำหรับเรียนรู้ 70% และข้อมูลสำหรับทดสอบ 30% โดยในช่วงการเรียนรู้ได้ใช้ข้อมูลสำหรับทดสอบมาเป็นข้อมูลสำหรับตรวจสอบ เมื่อได้ผลที่ดีที่สุดของการเรียนรู้โดยตรวจสอบจากค่าอัตราความถูกต้องที่วัดได้แล้วจะหยุดการเรียนรู้ หลังจากนั้นนำผลที่เก็บอยู่ในรูปของ Weight มาทดสอบด้วยชุดข้อมูลสำหรับทดสอบซึ่งเป็นชุดเดียวกันกับชุดที่ใช้ตรวจสอบ ในการทดสอบจะใช้ตัววัดประสิทธิภาพ 3 แบบ คือตัววัดประสิทธิภาพอัตราความถูกต้อง ตัววัดประสิทธิภาพ F-Measure และตัววัดประสิทธิภาพ Entropy

4.3 ชุดข้อมูลตัวอักษรที่ใช้ในการทดลองและผลการทดลอง

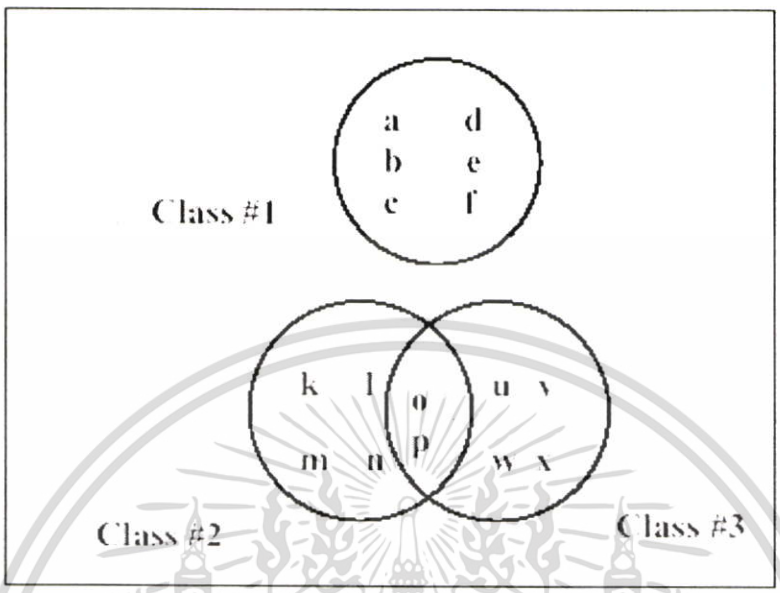
4.3.1 ชุดข้อมูลที่สร้างจากกลุ่มตัวอักษร (Synthesized Alphabet Document)

ข้อมูลชุดนี้สร้างจากกลุ่มตัวอักษรภาษาอังกฤษ 3 กลุ่ม ที่นำมาใช้ทดสอบการทำงานของโมเดลก่อนนำไปใช้กับข้อมูลจริง โดยข้อมูลแต่ละตัวอักษรเป็นตัวแทนของข้อมูลที่เป็นข้อความซึ่งมีคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ ดังแสดงตามรูปที่ 4.2 และ รูปที่ 4.3 ตามลำดับ

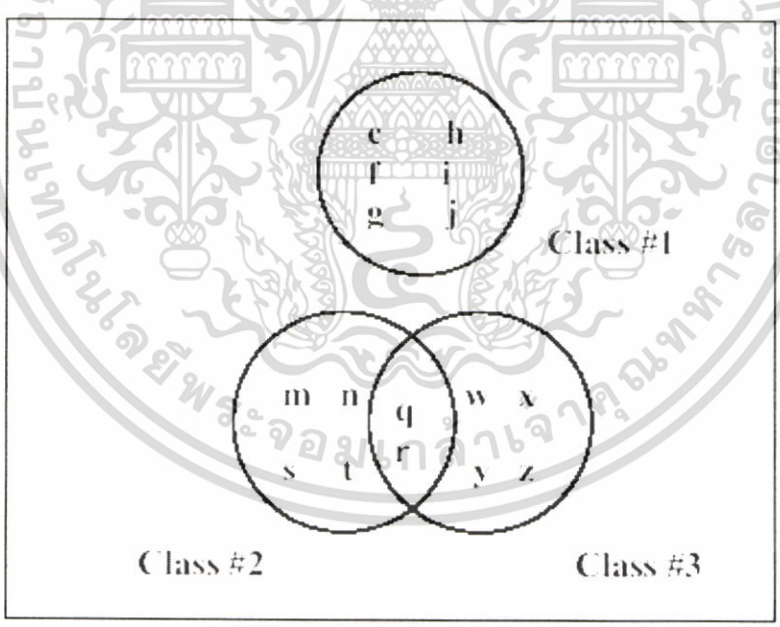
ข้อมูลที่สร้างจากกลุ่มตัวอักษร ประกอบด้วยคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ ซึ่งนำมาใช้ในการเรียนรู้ โดยใช้ข้อมูลจำนวน 200 ข้อมูล ซึ่งรายละเอียดของข้อมูลแต่ละ Class แสดงดังตารางที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างของข้อมูลที่สร้างจากข้อมูลตัวอักษร ที่นำมาใช้ในการเรียนรู้ 10 รายการแรก แสดงดังตารางที่ 4.2



รูปที่ 4.2 แสดงกลุ่มของตัวอักษรที่นำไปสร้างข้อมูลในคุณสมบัติชื่อเรื่อง



รูปที่ 4.3 แสดงกลุ่มของตัวอักษรที่นำไปสร้างข้อมูลในคุณสมบัติคำสำคัญ

ตารางที่ 4.1 แสดงจำนวนสมาชิกแต่ละ Class ของข้อมูลตัวอักษร ที่นำมาใช้ในการเรียนรู้

จำนวนสมาชิก		
Class ที่ 1	Class ที่ 2	Class ที่ 3
67	66	67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงตัวอย่างของข้อมูลตัวอักษรที่สร้างจากข้อมูลแสดงลักษณะเฉพาะ รูปที่ 4.2 และ รูปที่ 4.3 ที่นำมาใช้ในการเรียนรู้

ลำดับ ข้อมูล	ค่าของคุณสมบัติ		Category Input
	ชื่อเรื่อง	คำสำคัญ	
1	l,n,o,p	m,n,q,r,t	2
2	a,d,e,f	c,f,g,h	1
3	k,m,n	m,r,s,t	2
4	o,p,u,w	q,y,z	3
5	v,w,x	r,w,x,y,z	3
6	b,c,d	g,h,i,j	1
7	k,l,n,o	m,n,q,s	2
8	a,c,e,f	c,f,i	1
9	o,u	q,r,w,x,y,z	3
10	l,m,p	q,s,t	2

ข้อมูลที่นำมาใช้ในการทดสอบความถูกต้องของโมเดลที่เรียนรู้มาแล้ว โดยสร้างข้อมูลชุดนี้จากข้อมูลแสดงลักษณะเฉพาะชุดเดียวกันกับข้อมูลที่ใช้ในการเรียนรู้ ซึ่งมีจำนวนชุดข้อมูลทั้งหมด 5 ชุดๆ ละ 1,000 ข้อมูล รายละเอียดแสดงตามตารางที่ 4.3

ตารางที่ 4.3 แสดงจำนวนชุดข้อมูลตัวอักษรที่นำมาใช้ทดสอบความถูกต้องของโมเดล

ลำดับ ชุดข้อมูล	จำนวนสมาชิก		
	Class ที่ 1	Class ที่ 2	Class ที่ 3
1	330	336	334
2	328	338	334
3	332	333	335
4	332	336	332
5	330	332	338

4.3.2 ผลการทดลองของชุดข้อมูลตัวอักษร

ผลการทดลอง โดยใช้ข้อมูลตัวอักษรในการเรียนรู้จำนวน 200 ข้อมูล รายละเอียดตามตารางที่ 4.1 ซึ่งพารามิเตอร์ที่ใช้ในการทดลองเป็นดังนี้ Vigilance = 0.4 Learning Rate = 0.02

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Small Value = 0.02 Initial Degree = 0.5 และโหนดเอาท์พุทที่ได้จากการทดลองในชั้นเอาท์พุท (Output Layer) จำนวน 3 โหนด ที่เป็นตัวแทนของข้อมูลแต่ละกลุ่ม ชี้ไปที่ Class 1 2 และ 3 ซึ่งผลจากการทดสอบความถูกต้องปรากฏว่า โหนดเอาท์พุทชี้ไปที่ Class ทั้งสามได้อย่างถูกต้อง ทุกชุดข้อมูล

ผลจากการทดสอบความถูกต้องของโมเดลและการวัดประสิทธิภาพของโมเดล แสดงรายละเอียดตามตารางที่ 4.4

ตารางที่ 4.4 แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูลตัวอักษร

ลำดับชุดข้อมูล	จำนวนสมาชิก			ผลการทดลอง			อัตราความถูกต้อง	F-Measure	Entropy
	Class ที่ 1	Class ที่ 2	Class ที่ 3	Class ที่ 1	Class ที่ 2	Class ที่ 3			
1	330	336	334	330	336	334	100%	1.00	0.00
2	328	338	334	328	338	334	100%	1.00	0.00
3	332	333	335	332	333	335	100%	1.00	0.00
4	332	336	332	332	336	332	100%	1.00	0.00
5	330	332	338	330	332	338	100%	1.00	0.00
ค่าเฉลี่ยอัตราความถูกต้อง = 100.00%									
ค่าเฉลี่ย F-Measure = 1.00									
ค่าเฉลี่ย Entropy = 0.00									

ผลการทดลอง ของชุดข้อมูลตัวอักษรภาษาอังกฤษที่สร้างขึ้นเอง โดยให้ข้อมูล 1 ตัวอักษรแทนข้อความ 1 ข้อความ จากผลการทดลองแสดงให้เห็นว่าการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ สามารถจำแนกกลุ่มเอกสารได้อย่างถูกต้อง โดยค่าของความถูกต้องในการจำแนกกลุ่มเอกสาร วัดจากตัววัดประสิทธิภาพอัตราความถูกต้อง มีค่าเท่ากับ 100% ตัววัดประสิทธิภาพ F-Measure มีค่าเท่ากับ 1 และตัววัดประสิทธิภาพ Entropy มีค่าเท่ากับ 0 ทุกชุดข้อมูล ซึ่งแสดงให้เห็นว่าไม่มีการซ้อนทับกันของเอกสารกลุ่มอื่น

4.4 ชุดข้อมูล CSTR (CSTR Dataset) ที่ใช้ในการทดลองและผลการทดลอง

4.4.1 ชุดข้อมูล CSTR ที่ใช้ในการทดลอง

ชุดข้อมูล CSTR มาจาก Computer Science Technical Report [13] ซึ่งเป็นชุดข้อมูลที่ใช้ในการทดสอบการจำแนกประเภทเอกสาร (Text Categorization) และการจัดกลุ่มเอกสาร (Text Clustering) โดยชุดข้อมูล CSTR ที่ใช้ในการทดลอง เลือกข้อมูลที่อยู่ระหว่างปี ค.ศ. 1991 ถึง ค.ศ.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2002 ซึ่งมีเอกสารทั้งหมด 476 เอกสาร และมี 4 กลุ่มเอกสาร [18] โดยเก็บอยู่ในรูปแบบแฟ้มข้อมูล HTML ซึ่งมีรายละเอียดอยู่ในภาคผนวก ก. การทดลองนี้เลือกใช้ 2 คุณสมบัติ คือคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ รายละเอียดของแต่ละกลุ่มเอกสารแสดงในตารางที่ 4.5

ตารางที่ 4.5 แสดงจำนวนข้อมูล CSTR แต่ละกลุ่มเอกสาร

ลำดับที่	ชื่อกลุ่ม	จำนวนเอกสาร
1	AI	103
2	Robotics and Vision	74
3	Systems	172
4	Theory	127
รวมทั้งหมด		**476**

ในการทดลองได้แบ่งข้อมูลเอกสาร CSTR เพื่อใช้ในการเรียนรู้ 70% และใช้ในการทดสอบ 30% โดยใช้หลักการในการแบ่งข้อมูลสำหรับการเรียนรู้โดยการสุ่ม (Sampling Data) ซึ่งเป็นหลักการที่ใช้โดยทั่วไป ดังสมการที่ (4.8)

$$n = \frac{N}{1 + (N.e^2)} \tag{4.8}$$

โดยที่

n คือ จำนวนของเอกสารที่ใช้เรียนรู้ (Number of Sampled Data)

N คือ จำนวนเอกสารทั้งหมด (Number of Population)

e คือ ค่า Error ที่ยอมรับได้ (Sampling Error)

หลังจากแบ่งข้อมูลสำหรับใช้ในการเรียนรู้ตามสมการที่ (4.8) แล้ว จะได้ข้อมูลดังตารางที่ 4.6 ดังนี้

ตารางที่ 4.6 แสดงจำนวนข้อมูล CSTR ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร

ชื่อกลุ่ม	จำนวนเอกสารที่ใช้	
	เรียนรู้	ทดสอบ
AI	73	30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 (ต่อ)

ชื่อกลุ่ม	จำนวนเอกสารที่ใช้	
	เรียนรู้	ทดสอบ
Robotics and Vision	52	22
Systems	121	51
Theory	89	38
รวม	**335**	**141**

4.4.2 ผลการทดลองของชุดข้อมูล CSTR

การทดลองของชุดข้อมูล CSTR ซึ่งมี 4 กลุ่มเอกสาร จำนวนข้อมูลที่ใช้ในการทดสอบ 141 เอกสาร ซึ่งรายละเอียดของแต่ละกลุ่มตามตารางที่ 4.6 พารามิเตอร์ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุทที่ได้จากการทดลองชุดที่ดีที่สุดของแต่ละชุด แสดงดังตารางที่ 4.7 จำนวนโหนดเอาท์พุทนี้มีได้ไม่จำกัดขึ้นอยู่กับการกำหนดค่าวิจิแลนซ์ (Vigilance) ซึ่งโหนดเอาท์พุทเหล่านี้จะเก็บค่า Weight ที่ประกอบด้วยคำสำคัญของคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ กำหนดไว้ไม่เกิน 30 คำต่อโหนด

ตารางที่ 4.7 แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล CSTR ทุกชุด

ลำดับชุดข้อมูล	Vigilance	Learning Rate	Small Value	Initial Degree	Output Node
1	0.29	0.02	0.02	0.5	133
2	0.26	0.02	0.02	0.5	93
3	0.29	0.02	0.02	0.5	111
4	0.25	0.02	0.02	0.5	115
5	0.25	0.02	0.02	0.5	140
6	0.27	0.02	0.02	0.5	133
7	0.29	0.02	0.02	0.5	111
8	0.25	0.02	0.02	0.5	112
9	0.25	0.02	0.02	0.5	106
10	0.25	0.02	0.02	0.5	109
11	0.25	0.02	0.02	0.5	117

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลจากการทดลอง ของชุดข้อมูล CSTR แต่ละชุด แสดงจำนวนเอกสารทั้งหมดที่ได้จากการทดลองของแต่ละ Category รวมทั้งค่าอัตราความถูกต้อง ค่า F-Measure และค่า Entropy ดังตารางที่ 4.8

ตารางที่ 4.8 แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ทุกชุด

ลำดับชุดข้อมูล	จำนวนเอกสารใน Category ต่างๆ				อัตราความถูกต้อง	F-Measure	Entropy
	AI	Robotics	Systems	Theory			
1	29	21	54	37	86.52%	0.8639	0.6850
2	26	24	54	37	89.36%	0.8925	0.5445
3	26	26	54	35	87.94%	0.8806	0.5602
4	28	29	49	35	92.20%	0.9249	0.3924
5	24	30	48	39	92.20%	0.9240	0.3362
6	33	22	50	36	88.65%	0.8880	0.5548
7	18	32	52	39	89.36%	0.9811	0.4599
8	26	26	52	37	92.91%	0.9295	0.3586
9	27	27	49	38	92.20%	0.9235	0.3938
10	**25**	**26**	**53**	**37**	**92.91%**	**0.9294**	**0.3660**
11	26	26	52	37	92.91%	0.9295	0.3586
ค่าเฉลี่ยอัตราความถูกต้อง = 90.65%							
ค่าเฉลี่ย F-Measure = 0.9152							
ค่าเฉลี่ย Entropy = 0.4555							

รายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 1 ถึงชุดที่ 11 โดยใช้พารามิเตอร์แต่ละชุดตามตารางที่ 4.7 ซึ่งผลจากการทดสอบของข้อมูลแต่ละชุด แสดงรายละเอียดดังตารางที่ 4.9 ถึงตารางที่ 4.19 ตามลำดับ

ตารางที่ 4.9 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 1

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**23**	4	1	2	**30**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 (ต่อ)

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
Robotics and Vision	4	**15**	3	0	**22**
Systems	1	1	**49**	0	**51**
Theory	1	1	1	**35**	**38**
รวม	**29**	**21**	**54**	**37**	**141**
ค่าอัตราความถูกต้อง = 86.52%					
ค่า F-Measure = 0.8639					
ค่า Entropy = 0.6850					

ตารางที่ 4.10 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 2

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**22**	4	3	1	**30**
Robotics and Vision	3	**18**	1	0	**22**
Systems	0	1	**50**	0	**51**
Theory	1	1	0	**36**	**38**
รวม	**26**	**24**	**54**	**37**	**141**
ค่าอัตราความถูกต้อง = 89.36%					
ค่า F-Measure = 0.8925					
ค่า Entropy = 0.5445					

ตารางที่ 4.11 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 3

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**22**	7	1	0	**30**
Robotics and Vision	3	**17**	2	0	**22**
Systems	0	1	**50**	0	**51**
Theory	1	1	1	**35**	**38**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 (ต่อ)

รวม	**26**	**26**	**54**	**35**	**141**
ค่าอัตราความถูกต้อง = 87.94%					
ค่า F-Measure = 0.8806					
ค่า Entropy = 0.5602					

ตารางที่ 4.12 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 4

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**26**	4	0	0	**30**
Robotics and Vision	1	**21**	0	0	**22**
Systems	0	3	**48**	0	**51**
Theory	1	1	1	**35**	**38**
รวม	**28**	**29**	**49**	**35**	**141**
ค่าอัตราความถูกต้อง = 92.20%					
ค่า F-Measure = 0.9249					
ค่า Entropy = 0.3924					

ตารางที่ 4.13 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 5

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**23**	6	0	1	**30**
Robotics and Vision	1	**21**	0	0	**22**
Systems	0	3	**48**	0	**51**
Theory	0	0	0	**38**	**38**
รวม	**24**	**30**	**48**	**39**	**141**
ค่าอัตราความถูกต้อง = 92.20%					
ค่า F-Measure = 0.9240					
ค่า Entropy = 0.3362					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.14 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 6

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**25**	3	0	2	**30**
Robotics and Vision	4	**17**	1	0	**22**
Systems	1	1	**49**	0	**51**
Theory	3	1	0	**34**	**38**
รวม	**33**	**22**	**50**	**36**	**141**
ค่าอัตราความถูกต้อง = 88.65%					
ค่า F-Measure = 0.8880					
ค่า Entropy = 0.5548					

ตารางที่ 4.15 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 7

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**18**	9	1	2	**30**
Robotics and Vision	0	**21**	1	0	**22**
Systems	0	1	**50**	0	**51**
Theory	0	1	0	**37**	**38**
รวม	**18**	**32**	**52**	**39**	**141**
ค่าอัตราความถูกต้อง = 89.36%					
ค่า F-Measure = 0.8911					
ค่า Entropy = 0.4599					

ตารางที่ 4.16 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 8

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**24**	5	1	0	**30**
Robotics and Vision	1	**20**	1	0	**22**
Systems	0	1	**50**	0	**51**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 (ต่อ)

Theory	1	0	0	**37**	**38**
รวม	**26**	**26**	**52**	**37**	**141**
ค่าอัตราความถูกต้อง = 92.91%					
ค่า F-Measure = 0.9295					
ค่า Entropy = 0.3586					

ตารางที่ 4.17 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 9

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**25**	4	1	0	**30**
Robotics and Vision	1	**20**	1	0	**22**
Systems	1	3	**47**	0	**51**
Theory	0	0	0	**38**	**38**
รวม	**27**	**27**	**49**	**38**	**141**
ค่า อัตราความถูกต้อง = 92.20%					
ค่า F-Measure = 0.9235					
ค่า Entropy = 0.3938					

ตารางที่ 4.18 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 10

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**24**	5	1	0	**30**
Robotics and Vision	1	**20**	1	0	**22**
Systems	0	1	**50**	0	**51**
Theory	0	0	1	**37**	**38**
รวม	**25**	**26**	**53**	**37**	**141**
ค่าอัตราความถูกต้อง = 92.91%					
ค่า F-Measure = 0.9294					
ค่า Entropy = 0.3660					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.19 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR ชุดที่ 11

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category				
	AI	Robotics	Systems	Theory	**รวม**
AI	**24**	5	1	0	**30**
Robotics and Vision	1	**20**	1	0	**22**
Systems	0	1	**50**	0	**51**
Theory	1	0	0	**37**	**38**
รวม	**26**	**26**	**52**	**37**	**141**
ค่าอัตราความถูกต้อง = 92.91%					
ค่า F-Measure = 0.9295					
ค่า Entropy = 0.3586					

ผลจากการทดลองของข้อมูลแต่ละชุด แสดงให้เห็นว่าการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ สามารถจำแนกกลุ่มเอกสารได้อย่างถูกต้อง ซึ่งค่าของความถูกต้องในการจำแนกกลุ่มเอกสาร โดยวัดจากตัววัดประสิทธิภาพอัตราความถูกต้อง ตัววัดประสิทธิภาพ F-Measure และตัววัดประสิทธิภาพ Entropy ได้ค่าจากการคำนวณอยู่ในระดับที่ดี

ผลจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล CSTR แสดงให้เห็นว่ากลุ่มเอกสาร AI และ Robotics and Vision มีคำสำคัญที่ซ้อนทับกันมากกว่ากลุ่มอื่นๆ เนื่องจาก Robotics and Vision เป็นส่วนหนึ่งของ AI ซึ่งผลที่ได้ตามตารางที่ 4.9 ถึงตารางที่ 4.19 แสดงให้เห็นว่าทุกชุดข้อมูลของกลุ่มเอกสาร AI จะให้คำตอบเป็นกลุ่มเอกสาร Robotics and Vision จำนวนมากกว่ากลุ่มเอกสารอื่นทุกชุด ส่วนกลุ่มเอกสาร Theory ซ้อนทับกับกลุ่มเอกสารอื่นน้อย

4.4.3 ผลการเปรียบเทียบกับโมเดลต่างๆ ของชุดข้อมูล CSTR

โมเดลที่ใช้ในการเปรียบเทียบความถูกต้องของการทำงาน จากงานวิจัย [18] ซึ่งมีโมเดลต่างๆ ดังต่อไปนี้

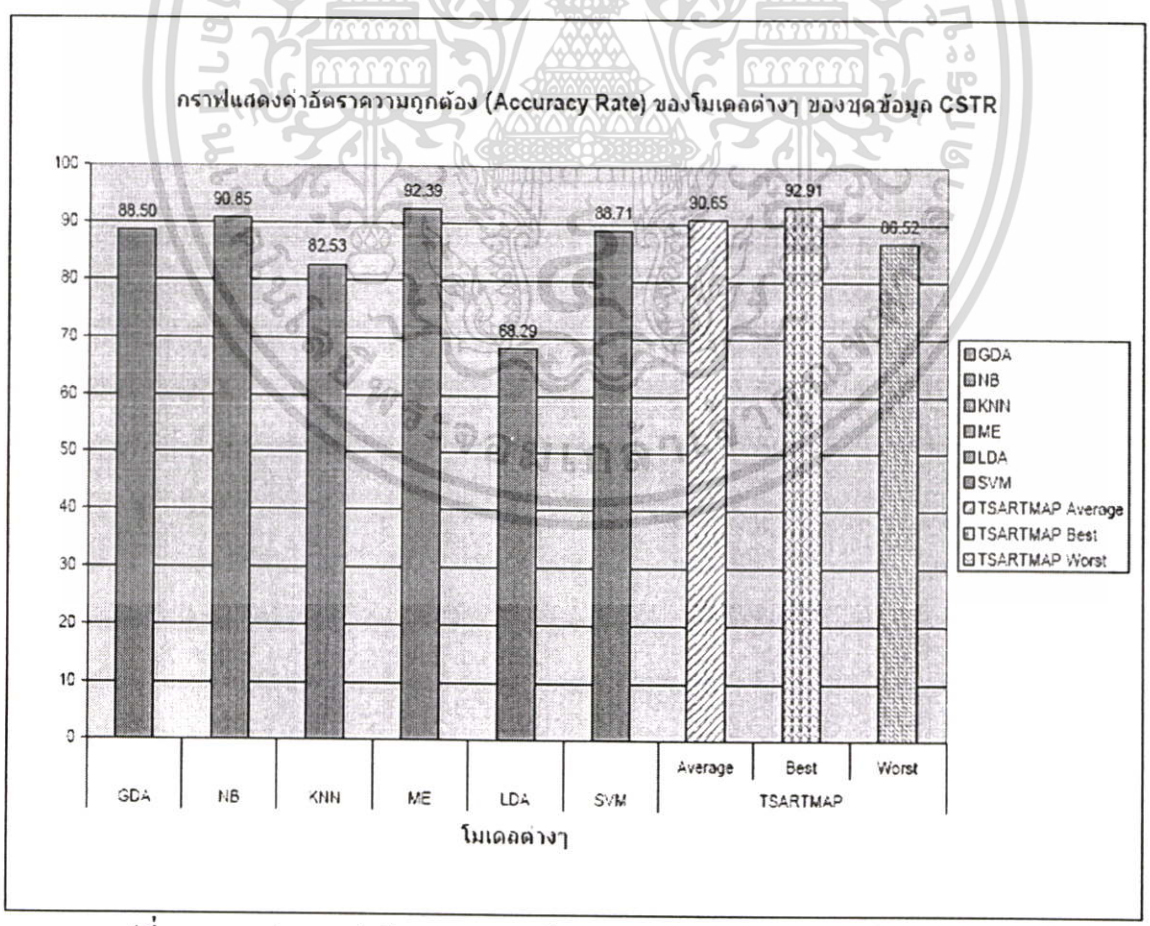
1. GDA คือ Generalized Discriminant Analysis
2. NB คือ Naive Bayes
3. KNN คือ K-Nearest Neighbor
4. ME คือ Maximum Entropy
5. LDA คือ Linear Discriminant Analysis
6. SVM คือ Support Vector Machine
7. TSARTMAP คือ Text Simplified ARTMAP Neural Network

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการทำงานของโมเดลต่างๆ จากงานวิจัย [18] ได้ใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง คิดเป็นเปอร์เซ็นต์ (%) ของแต่ละโมเดล โดยที่โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พใช้ค่าเฉลี่ยในการเปรียบเทียบกับโมเดลอื่นๆ ซึ่งมีค่าอัตราความถูกต้องเฉลี่ยใกล้เคียงกับโมเดลอื่นๆ รวมทั้งแสดงค่าที่ดีที่สุดซึ่งมีค่ามากกว่าโมเดลอื่นๆ เป็นอันดับที่ 1 และแสดงค่าที่เลวร้ายที่สุดของการทดลองแต่ละชุด ดังแสดงในตารางที่ 4.20 และรูปที่ 4.4 จำนวนคำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมนี้ จากการทดลองชุดที่ดีที่สุดของชุดข้อมูล CSTR มีจำนวนคำสำคัญ 1,555 คำ ซึ่งมีรายละเอียดในภาคผนวก ข.

ตารางที่ 4.20 แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดล ของชุดข้อมูล CSTR

จำนวนข้อมูลที่ใช้ในการทดลอง	โมเดลเปรียบเทียบ โดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง (%)								
	GDA	NB	KNN	ME	LDA	SVM	TSARTMAP		
							Average	Best	Worst
476	88.50	90.85	82.53	**92.39**	68.29	88.71	90.65	**92.91**	86.52



รูปที่ 4.4 กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของชุดข้อมูล CSTR

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลจากการเปรียบเทียบโดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง ที่ทดลองกับชุดข้อมูล CSTR ของโมเดลต่างๆ แสดงให้เห็นว่า โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ สามารถจำแนกประเภทเอกสารได้ถูกต้องในระดับที่ใกล้เคียงกับโมเดลอื่นๆ โดยดูผลที่ได้จากค่าอัตราความถูกต้องเฉลี่ย

4.5 ชุดข้อมูล Reuters-Top10 (Reuters-Top10 Dataset) ที่ใช้ในการทดลองและผลการทดลอง

4.5.1 ชุดข้อมูล Reuters-Top10 ที่ใช้ในการทดลอง

ชุดข้อมูล Reuters-Top10 คัดเลือกมาจาก ชุดข้อมูลข่าว Reuters-21578 [12] ซึ่ง David D. Lewis และคณะได้รวบรวม ปรับปรุงแก้ไขและจัดทำรูปแบบโดยจัดเก็บอยู่ในรูปแบบแฟ้มข้อมูล SGML มี 135 กลุ่มเอกสารข่าว จำนวน 21,578 เอกสารข่าว โดย Reuters-Top10 คัดเลือกเอาเฉพาะเอกสารข่าว 10 กลุ่มแรกที่มีจำนวนข่าวมากที่สุดจาก 135 กลุ่มเอกสารข่าว ซึ่งชุดข้อมูล Reuters-Top10 [14] เป็นชุดข้อมูลที่ใช้ในการทดสอบการจำแนกประเภทเอกสารและการจัดกลุ่มเอกสาร ที่ใช้กันอย่างแพร่หลาย ลักษณะการเก็บอยู่ในรูปแบบแฟ้มข้อมูล XML มี 10 กลุ่มเอกสาร จำนวน 2,775 เอกสาร ซึ่งมีรายละเอียดอยู่ในภาคผนวก ก. การทดลองนี้เลือกใช้ 2 คุณสมบัติ คือคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ รายละเอียดของแต่ละกลุ่มเอกสารแสดงในตารางที่ 4.21

ตารางที่ 4.21 แสดงจำนวนข้อมูล Reuters-Top10 แต่ละกลุ่มเอกสาร

ลำดับที่	ชื่อกลุ่ม	จำนวนเอกสาร
1	acq	719
2	corn	56
3	crude	189
4	earn	1,085
5	grain	148
6	interest	129
7	money-fx	176
8	ship	89
9	trade	113
10	wheat	71
รวมทั้งหมด		**2,775**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการทดลองได้แบ่งข้อมูลเอกสาร Reuters-Top10 เพื่อใช้ในการเรียนรู้ 70% และใช้ในการทดสอบ 30% โดยใช้หลักการในการแบ่งข้อมูลสำหรับการเรียนรู้โดยการสุ่ม ตามสมการที่ (4.8) ได้ข้อมูลดังตารางที่ 4.22 ดังนี้

ตารางที่ 4.22 แสดงจำนวนข้อมูล Reuters-Top10 ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร

ลำดับที่	ชื่อกลุ่ม	จำนวนเอกสาร	จำนวนที่ใช้เรียนรู้	จำนวนที่ใช้ทดสอบ
1	acq	719	504	215
2	corn	56	40	16
3	crude	189	133	56
4	earn	1,085	761	324
5	grain	148	104	44
6	interest	129	91	38
7	money-fx	176	124	52
8	ship	89	63	26
9	trade	113	80	33
10	wheat	71	50	21
	รวม	**2,775**	**1,950**	**825**

4.5.2 ผลการทดลองของชุดข้อมูล Reuters-Top10

การทดลองของชุดข้อมูล Reuters-Top10 ซึ่งมี 10 กลุ่มเอกสาร จำนวนข้อมูลที่ใช้ในการทดสอบ 825 เอกสาร รายละเอียดของแต่ละกลุ่มเอกสารตามตารางที่ 4.22 พารามิเตอร์ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุทที่ได้จากการทดลองชุดที่ดีที่สุดของแต่ละชุด แสดงดังตารางที่ 4.23 ซึ่งโหนดเอาท์พุทเหล่านี้จะเก็บค่า Weight ที่ประกอบด้วยคำสำคัญของคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ กำหนดไว้ไม่เกิน 40 คำต่อโหนด

ตารางที่ 4.23 แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล Reuters-Top10 ทุกชุด

ลำดับชุดข้อมูล	Vigilance	Learning Rate	Small Value	Initial Degree	Output Node
1	0.30	0.01	0.02	0.4	773

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.23 (ต่อ)

ลำดับชุดข้อมูล	Vigilance	Learning Rate	Small Value	Initial Degree	Output Node
2	0.30	0.01	0.02	0.4	783
3	0.30	0.009	0.02	0.4	796
4	0.30	0.01	0.02	0.4	914
5	0.30	0.009	0.02	0.4	916

ตารางที่ 4.24 แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ทุกชุด

ลำดับชุดข้อมูล	จำนวนเอกสารใน Category ต่างๆ										อัตราความถูกต้อง	F-Measure	Entropy
	1	2	3	4	5	6	7	8	9	10			
1	180	9	57	338	45	30	68	34	36	28	80.00%	0.8011	0.8957
2	183	9	54	337	46	30	67	32	38	29	80.85%	0.8096	0.8698
3	184	10	74	334	45	36	54	22	36	30	82.67%	0.8175	0.8221
4	219	8	57	325	43	31	58	22	34	28	80.12%	0.7992	0.8981
5	**200**	**11**	**64**	**334**	**48**	**35**	**52**	**27**	**35**	**19**	**84.24%**	**0.8503**	**0.6594**
ค่าเฉลี่ยอัตราความถูกต้อง = 81.58%													
ค่าเฉลี่ย F-Measure = 0.8155													
ค่าเฉลี่ย Entropy = 0.8290													

รายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 1 ถึงชุดที่ 5 โดยใช้พารามิเตอร์แต่ละชุดตามตารางที่ 4.23 ซึ่งผลจากการทดสอบของข้อมูลแต่ละชุดแสดงรายละเอียดดังตารางที่ 4.25 ถึงตารางที่ 4.29 ตามลำดับ

ตารางที่ 4.25 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 1

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**รวม**
1. acq	**170**	1	9	20	2	1	9	0	3	0	**215**
2. corn	0	**3**	0	0	7	0	0	0	1	5	**16**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.25 (ต่อ)

3. crude	3	0	**40**	2	1	0	1	8	1	0	**56**
4. earn	5	0	3	**313**	2	0	1	0	0	0	**324**
5. grain	0	4	2	2	**18**	0	0	3	1	14	**44**
6. interest	0	0	0	0	1	**21**	14	1	1	0	**38**
7. money-fx	0	0	0	0	0	8	**41**	1	2	0	**52**
8. ship	0	0	3	1	3	0	1	**18**	0	0	**26**
9. trade	0	0	0	0	4	0	1	1	**27**	0	**33**
10. wheat	2	1	0	0	7	0	0	2	0	**9**	**21**
รวม	**180**	**9**	**57**	**338**	**45**	**30**	**68**	**34**	**36**	**28**	**825**
ค่าอัตราความถูกต้อง = 80.00%											
ค่า F-Measure = 0.8011											
ค่า Entropy = 0.8957											

ตารางที่ 4.26 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 2

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**รวม**
1. acq	**175**	1	8	18	2	1	8	0	2	0	**215**
2. corn	0	**3**	0	0	7	0	0	0	1	5	**16**
3. crude	2	0	**39**	1	2	0	1	8	2	1	**56**
4. earn	6	0	2	**313**	1	0	1	0	1	0	**324**
5. grain	0	4	2	2	**19**	0	0	2	1	14	**44**
6. interest	0	0	0	0	1	**21**	13	1	2	0	**38**
7. money-fx	0	0	0	0	0	8	**42**	1	1	0	**52**
8. ship	0	0	3	1	3	0	1	**18**	0	0	**26**
9. trade	0	0	0	0	3	0	1	1	**28**	0	**33**
10. wheat	0	1	0	2	8	0	0	1	0	**9**	**21**
รวม	**183**	**9**	**54**	**337**	**46**	**30**	**67**	**32**	**38**	**29**	**825**
ค่าอัตราความถูกต้อง = 80.85%											
ค่า F-Measure = 0.8096											
ค่า Entropy = 0.8698											

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.27 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 3

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**รวม**
1. acq	**177**	1	13	14	2	1	3	0	3	1	**215**
2. corn	0	**3**	0	0	7	0	0	0	1	5	**16**
3. crude	1	1	**45**	2	1	0	0	6	0	0	**56**
4. earn	4	0	2	**315**	0	1	1	0	0	1	**324**
5. grain	0	4	3	1	**19**	0	0	2	1	14	**44**
6. interest	0	0	1	0	0	**25**	11	0	1	0	**38**
7. money-fx	1	0	1	0	0	9	**39**	0	2	0	**52**
8. ship	0	0	8	1	4	0	0	**13**	0	0	**26**
9. trade	1	0	0	0	4	0	0	0	**28**	0	**33**
10. wheat	0	1	1	1	8	0	0	1	0	**9**	**21**
รวม	**184**	**10**	**74**	**334**	**45**	**36**	**54**	**22**	**36**	**30**	**825**
ค่าอัตราความถูกต้อง = 82.67%											
ค่า F-Measure = 0.8175											
ค่า Entropy = 0.8221											

ตารางที่ 4.28 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 4

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**รวม**
1. acq	**184**	0	6	20	2	1	1	0	0	1	**215**
2. corn	0	**3**	0	0	7	0	0	0	1	5	**16**
3. crude	4	0	**41**	4	1	0	1	5	0	0	**56**
4. earn	27	0	0	**297**	0	0	0	0	0	0	**324**
5. grain	2	4	1	2	**18**	0	0	2	1	14	**44**
6. interest	0	0	1	0	0	**24**	12	0	1	0	**38**
7. money-fx	0	0	1	0	0	6	**43**	0	2	0	**52**
8. ship	1	0	7	0	3	0	1	**14**	0	0	**26**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.28 (ต่อ)

9. trade	0	0	0	0	4	0	0	0	**29**	0	**33**
10. wheat	1	1	0	2	8	0	0	1	0	**8**	**21**
รวม	**219**	**8**	**57**	**325**	**43**	**31**	**58**	**22**	**34**	**28**	**825**
ค่าอัตราความถูกต้อง = 80.12%											
ค่า F-Measure = 0.7992											
ค่า Entropy = 0.8981											

ตารางที่ 4.29 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ชุดที่ 5

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**รวม**
1. acq	**192**	0	7	11	3	0	1	0	0	1	**215**
2. corn	0	**3**	0	0	8	0	0	0	1	4	**16**
3. crude	2	0	**48**	0	0	1	0	5	0	0	**56**
4. earn	5	0	0	**319**	0	0	0	0	0	0	**324**
5. grain	0	4	2	2	**23**	0	0	2	1	10	**44**
6. interest	0	0	0	0	0	**26**	11	1	0	0	**38**
7. money-fx	1	0	1	0	0	7	**40**	1	2	0	**52**
8. ship	0	0	5	0	3	1	0	**17**	0	0	**26**
9. trade	0	2	0	0	0	0	0	0	**31**	0	**33**
10. wheat	0	2	1	2	11	0	0	1	0	**4**	**21**
รวม	**200**	**11**	**64**	**334**	**48**	**35**	**52**	**27**	**35**	**19**	**825**
ค่าอัตราความถูกต้อง = 84.24%											
ค่า F-Measure = 0.8503											
ค่า Entropy = 0.6594											

ผลจากการทดลองของข้อมูลแต่ละชุด แสดงให้เห็นว่าการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ สามารถจำแนกกลุ่มเอกสารได้อย่างถูกต้อง ซึ่งค่าของความถูกต้องในการจำแนกกลุ่มเอกสาร โดยวัดจากตัววัดประสิทธิภาพอัตราความถูกต้อง ตัววัดประสิทธิภาพ F-Measure และตัววัดประสิทธิภาพ Entropy ได้ค่าจากการคำนวณอยู่ในระดับที่ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

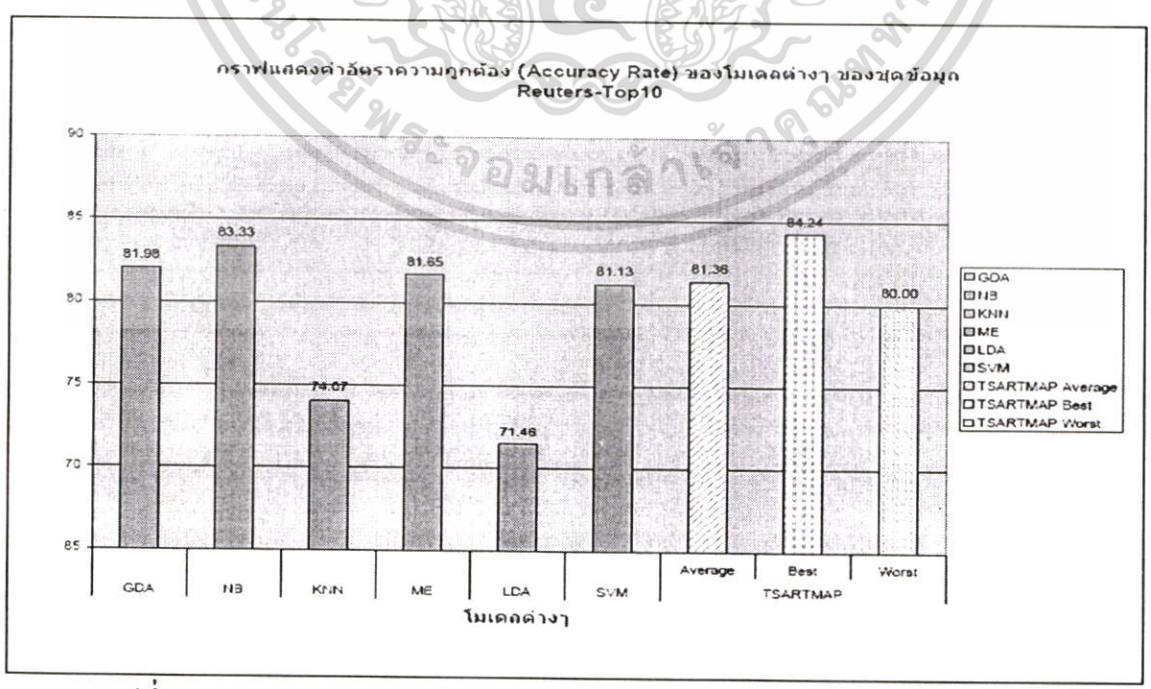
ผลจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล Reuters-Top10 ตามตารางที่ 4.25 ถึงตารางที่ 4.29 แสดงให้เห็นว่ากลุ่มเอกสาร corn, grain และ wheat มีคำสำคัญที่ซ้อนทับกัน เนื่องจาก corn และ wheat เป็นส่วนหนึ่งของ grain ส่วนกลุ่มเอกสาร interest ซ้อนทับกับกลุ่มเอกสาร money-fx และกลุ่มเอกสาร crude ซ้อนทับกับกลุ่มเอกสาร ship

4.5.3 ผลการเปรียบเทียบกับโมเดลต่างๆ ของชุดข้อมูล Reuters-Top10

จากผลการทำงานของโมเดลต่างๆ จากงานวิจัย [18] ได้ใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง คิดเป็นเปอร์เซ็นต์ (%) ของแต่ละโมเดล โดยที่โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พใช้ค่าเฉลี่ยในการเปรียบเทียบกับโมเดลอื่นๆ ซึ่งมีค่าอัตราความถูกต้องเฉลี่ยใกล้เคียงกับโมเดลอื่นๆ รวมทั้งแสดงค่าที่ดีที่สุดซึ่งมีค่ามากกว่าโมเดลอื่นๆ เป็นอันดับที่ 1 และแสดงค่าที่เลวร้ายที่สุดของการทดลองแต่ละชุด ดังแสดงในตารางที่ 4.30 และรูปที่ 4.5 จำนวนคำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมนี้จากการทดลองชุดที่ดีที่สุดของชุดข้อมูล Reuters-Top10 มีจำนวนคำสำคัญ 5,085 คำ ซึ่งมีรายละเอียดในภาคผนวก ข.

ตารางที่ 4.30 แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดล ของชุดข้อมูล Reuters-Top10

จำนวนข้อมูลที่ใช้ในการทดลอง	โมเดลเปรียบเทียบ โดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง (%)								
	GDA	NB	KNN	ME	LDA	SVM	TSARTMAP		
							Average	Best	Worst
2,775	81.98	**83.33**	74.07	81.65	71.46	81.13	81.36	**84.24**	80.00



รูปที่ 4.5 กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของชุดข้อมูล Reuters-Top10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ดัดแปลง เผยแพร่ หรือแสวงหาประโยชน์ทางการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลจากการเปรียบเทียบโดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง ที่ทดลองกับชุดข้อมูล Reuters-Top10 ของโมเดลต่างๆ แสดงให้เห็นว่า โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ สามารถจำแนกประเภทเอกสารได้ถูกต้องในระดับที่ใกล้เคียงกับโมเดลอื่นๆ โดยดูผลที่ได้จากค่าอัตราความถูกต้องเฉลี่ย

4.6 ชุดข้อมูล K-dataset ที่ใช้ในการทดลองและผลการทดลอง

4.6.1 ชุดข้อมูล K-dataset ที่ใช้ในการทดลอง

ชุดข้อมูล K-dataset [15] เป็นชุดข้อมูลจริงที่ใช้ในการทดลองชุดที่ 3 ซึ่งเป็นชุดข้อมูลที่ใช้ในการทดสอบการจำแนกประเภทเอกสารและการจัดกลุ่มเอกสาร เช่นเดียวกับ 2 ชุดที่กล่าวมา มี 20 กลุ่มเอกสาร จำนวน 2,340 เอกสาร โดยเก็บอยู่ในรูปแบบแฟ้มข้อมูล HTML ซึ่งมีรายละเอียดในภาคผนวก ก. การทดลองนี้เลือกใช้ 2 คุณสมบัติ คือคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ

ในการทดลองได้แบ่งข้อมูลเอกสาร K-dataset เพื่อใช้ในการเรียนรู้ 70% และใช้ในการทดสอบ 30% โดยใช้หลักการในการแบ่งข้อมูลสำหรับการเรียนรู้โดยการสุ่ม ตามสมการที่ (4.8) ซึ่งรายละเอียดของข้อมูลแต่ละกลุ่มเอกสารแสดงในตารางที่ 4.31 ดังนี้

ตารางที่ 4.31 แสดงจำนวนข้อมูล K-dataset ที่ใช้ในการเรียนรู้และทดสอบของแต่ละกลุ่มเอกสาร

ลำดับที่	ชื่อกลุ่ม	จำนวนเอกสาร	จำนวนที่ใช้เรียนรู้	จำนวนที่ใช้ทดสอบ
1	B-Business	142	100	42
2	E-Entertainment	9	7	2
3	Ea-Art	24	17	7
4	Ec-Cable	44	31	13
5	Ecu-Culture	74	52	22
6	Ef-Film	278	195	83
7	Ei-Industry	70	50	20
8	Em-Media	21	15	6
9	Emm-Multimedia	14	10	4
10	Emu-Music	125	88	37
11	Eo-Online	65	46	19
12	Ep-People	248	174	74
13	Er-Review	158	111	47
14	Es-Stage	18	13	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.31 (ต่อ)

ลำดับที่	ชื่อกลุ่ม	จำนวนเอกสาร	จำนวนที่ใช้เรียนรู้	จำนวนที่ใช้ทดสอบ
15	Et-Television	187	131	56
16	Ev-Variety	54	38	16
17	H-Health	494	347	147
18	P-Politics	114	80	34
19	S-Sports	141	99	42
20	T-Technology	60	43	17
รวม		**2,340**	**1,647**	**693**

4.6.2 ผลการทดลองของชุดข้อมูล K-dataset

การทดลองของชุดข้อมูล K-dataset ซึ่งมี 20 กลุ่มเอกสาร จำนวนข้อมูลที่ใช้ในการทดสอบ 693 เอกสาร ซึ่งรายละเอียดของแต่ละกลุ่มตามตารางที่ 4.31 พารามิเตอร์ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุทที่ได้จากการทดลองชุดที่ดีที่สุดของแต่ละชุด แสดงดังตารางที่ 4.32 ซึ่งโหนดเอาท์พุทเหล่านี้จะเก็บค่า Weight ที่ประกอบด้วยคำสำคัญของคุณสมบัติชื่อเรื่องและคุณสมบัติคำสำคัญ กำหนดไว้ไม่เกิน 40 คำต่อโหนด

ตารางที่ 4.32 แสดงค่าพารามิเตอร์ต่างๆ ที่ใช้ในการทดลองและจำนวนโหนดเอาท์พุท ของชุดข้อมูล K-dataset ทุกชุด

ลำดับชุดข้อมูล	Vigilance	Learning Rate	Small Value	Initial Degree	Output Node
1	0.32	0.01	0.02	0.4	934
2	0.30	0.01	0.02	0.4	808
3	0.30	0.01	0.02	0.4	852
4	0.29	0.01	0.02	0.4	815

ผลจากการทดลอง ของชุดข้อมูล K-dataset แต่ละชุด แสดงจำนวนเอกสารแต่ละ Category รวมทั้งค่าอัตราความถูกต้อง ค่า F-Measure และค่า Entropy ดังตารางที่ 4.33 ซึ่งแสดงในหน้าถัดไป

รายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 1 ถึงชุดที่ 4 โดยใช้พารามิเตอร์แต่ละชุดตามตารางที่ 4.32 ซึ่งผลจากการทดสอบของข้อมูลแต่ละชุด แสดงรายละเอียดดังตารางที่ 4.34 ถึงตารางที่ 4.37 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.33 แสดงผลลัพธ์ที่ได้จากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ทุกชุด

ลำดับชุดข้อมูล	จำนวนเอกสารใน Category ต่างๆ																				อัตราความถูกต้อง	F-Measure	Entropy
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20			
1	34	4	13	9	23	66	27	4	5	33	21	68	56	8	59	16	152	40	41	14	83.12%	0.8297	0.8656
2	43	3	10	11	24	69	27	3	6	36	22	71	49	8	54	16	143	40	41	17	81.53%	0.8169	0.9718
3	45	1	9	16	21	60	20	3	6	34	21	84	44	8	59	16	145	40	45	16	84.27%	0.8417	0.8197
4	**48**	**1**	**12**	**10**	**20**	**75**	**21**	**6**	**2**	**39**	**19**	**79**	**43**	**6**	**60**	**16**	**143**	**35**	**42**	**16**	**87.45%**	**0.8741**	**0.6921**
ค่าเฉลี่ยอัตราความถูกต้อง = 84.09%																							
ค่าเฉลี่ย F-Measure = 0.8406																							
ค่าเฉลี่ย Entropy = 0.8373																							

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ตารางที่ 4.34 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 1

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
1. B	27	2	0	0	0	1	0	0	0	0	1	2	0	0	1	0	0	5	0	3	42
2. E	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	2
3. Ea	0	0	4	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	7
4. Ec	0	0	0	6	0	3	1	0	0	0	0	0	0	0	3	0	0	0	0	0	13
5. Ecu	0	0	0	0	17	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	22
6. Ef	0	0	4	1	1	56	5	0	1	2	0	5	3	1	2	0	1	0	1	0	83
7. Ei	0	0	0	0	1	0	17	0	0	0	0	0	0	0	0	0	0	0	0	2	20
8. Em	0	0	0	1	0	0	0	3	0	0	0	1	0	0	1	0	0	0	0	0	6
9.Emm	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4
10.Emu	0	0	3	1	1	4	0	0	0	26	0	0	2	0	0	0	0	0	0	0	37
11.Eo	1	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	1	19
12.Ep	0	2	2	0	3	1	0	1	0	1	0	59	1	1	1	0	1	0	1	0	74
13.Er	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	47
14.Es	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5
15.Et	0	0	0	0	0	0	3	0	0	1	0	0	2	0	48	0	0	1	1	0	56
16.Ev	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	16
17.H	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	147	0	0	0	147
18.P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	33	0	0	34
19.S	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	0	37	0	42
20.T	6	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	1	0	7	17
รวม	34	4	13	9	23	66	27	4	5	33	21	68	56	8	59	16	152	40	41	14	693
ค่าอัตราความถูกต้อง = 83.12%																					
ค่า F-Measure = 0.8297																					
ค่า Entropy = 0.8656																					

ตารางที่ 4.35 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 2

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
1. B	29	2	0	0	1	0	0	0	0	0	1	2	0	0	0	0	1	3	0	3	42
2. E	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
3. Ea	0	0	2	0	0	0	1	0	1	0	0	2	0	0	0	0	0	0	1	0	7
4. Ec	0	0	1	6	0	3	2	0	0	0	0	0	0	0	1	0	0	0	0	0	13

โอนเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.35 (ต่อ)

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
5. Ecu	0	0	0	0	18	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	22
6. Ef	0	0	1	3	1	57	5	0	1	0	0	7	2	1	3	0	0	2	0	0	83
7. Ei	1	0	0	1	1	0	15	0	0	0	0	0	0	0	0	0	0	0	1	1	20
8. Em	0	0	0	0	0	0	2	3	0	0	0	0	0	0	0	0	0	1	0	0	6
9.Emm	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4
10.Emu	2	0	1	0	1	2	0	0	0	28	0	1	1	0	0	0	1	0	0	0	37
11.Eo	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	1	19
12.Ep	2	1	4	0	2	4	0	0	0	2	0	57	0	0	0	0	1	0	1	0	74
13.Er	0	0	1	0	0	0	0	0	0	0	0	1	44	0	1	0	0	0	0	0	47
14.Es	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5
15.Et	0	0	0	0	0	0	2	0	0	3	0	0	1	0	48	0	1	0	1	0	56
16.Ev	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	16
17.H	2	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	137	2	0	2	147
18.P	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	31	0	0	34
19.S	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	1	0	37	0	42
20.T	5	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	10	17
รวม	43	3	10	11	24	69	27	3	6	36	22	71	49	8	54	16	143	40	41	17	693
ค่าอัตราความถูกต้อง = 81.53%																					
ค่า F-Measure = 0.8169																					
ค่า Entropy = 0.9718																					

ตารางที่ 4.36 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 3

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
1. B	28	1	0	0	1	1	0	0	0	0	1	2	0	0	0	0	0	7	0	1	42
2. E	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
3. Ea	0	0	3	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	7
4. Ec	0	0	1	7	0	3	1	0	0	0	0	0	0	0	1	0	0	0	0	0	13
5. Ecu	0	0	0	0	18	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	22
6. Ef	0	0	1	2	1	53	1	0	2	1	0	10	2	2	3	0	1	2	2	0	83
7. Ei	1	0	0	1	1	0	15	0	0	0	0	0	0	0	0	0	0	0	0	2	20
8. Em	0	0	0	1	0	0	0	3	0	0	1	0	0	0	1	0	0	0	0	0	6
9.Emm	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.36 (ต่อ)

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
10.Emu	1	0	2	0	0	3	0	0	0	29	0	0	1	0	0	0	1	0	0	0	37
11.Eo	1	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	1	19
12.Ep	1	0	0	2	0	0	0	0	0	2	0	69	0	0	0	0	0	0	0	0	74
13.Er	1	0	1	1	0	0	0	0	0	0	0	2	41	0	1	0	0	0	0	0	47
14.Es	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5
15.Et	0	0	1	1	0	0	2	0	0	0	0	0	0	0	52	0	0	0	0	0	56
16.Ev	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	16
17.H	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	142	0	0	2	147
18.P	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	34
19.S	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	0	42
20.T	5	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	10	17
รวม	45	1	9	16	21	60	20	3	6	34	21	84	44	8	59	16	145	40	45	16	693
ค่าอัตราความถูกต้อง = 84.27%																					
ค่า F-Measure = 0.8417																					
ค่า Entropy = 0.8197																					

ตารางที่ 4.37 แสดงรายละเอียดจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ชุดที่ 4

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
1. B	34	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	3	42
2. E	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2
3. Ea	0	0	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	7
4. Ec	0	0	1	8	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	13
5. Ecu	0	0	0	0	19	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
6. Ef	0	0	2	0	0	63	3	1	0	0	0	4	1	1	6	0	1	0	1	0	83
7. Ei	1	0	0	0	1	1	15	0	0	0	0	0	0	0	1	0	0	0	1	0	20
8. Em	0	0	0	1	0	0	0	3	0	0	0	1	0	0	0	0	0	1	0	0	6
9.Emm	0	0	0	0	0	1	0	0	2	1	0	0	0	0	0	0	0	0	0	0	4
10.Emu	0	0	1	0	0	0	0	1	0	34	0	1	0	0	0	0	0	0	0	0	37
11.Eo	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	2	19
12.Ep	1	0	0	0	0	0	0	1	0	2	0	70	0	0	0	0	0	0	0	0	74
13.Er	1	0	1	0	0	2	0	0	0	1	0	1	41	0	0	0	0	0	0	0	47
14.Es	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.37 (ต่อ)

กลุ่มเอกสาร (Class)	จำนวนเอกสารที่ได้จากการทดลองในแต่ละ Category																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	รวม
15.Et	0	0	1	0	0	1	1	0	0	0	0	0	1	0	52	0	0	0	0	0	56
16.Ev	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	16
17.H	1	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	141	0	0	2	147
18.P	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	31	0	0	34
19.S	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	40	0	42
20.T	6	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	9	17
รวม	48	1	12	10	20	75	21	6	2	39	19	79	43	6	60	16	143	35	42	16	693
ค่าอัตราความถูกต้อง = 87.45%																					
ค่า F-Measure = 0.8741																					
ค่า Entropy = 0.6921																					

ผลจากการทดลองของข้อมูลแต่ละชุด แสดงให้เห็นว่าการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ สามารถจำแนกกลุ่มเอกสารได้อย่างถูกต้อง ซึ่งค่าของความถูกต้องในการจำแนกกลุ่มเอกสาร โดยวัดจากตัววัดประสิทธิภาพอัตราความถูกต้อง ตัววัดประสิทธิภาพ F-Measure และตัววัดประสิทธิภาพ Entropy มีค่าจากการคำนวณอยู่ในระดับที่ดี

ผลจากการทดสอบการทำงานของโมเดล ของชุดข้อมูล K-dataset ตามตารางที่ 4.34 ถึงตารางที่ 4.37 แสดงให้เห็นว่ากลุ่มเอกสาร film มีคำสำคัญที่ซ้อนทับกับกลุ่มเอกสารอื่นหลายกลุ่ม โดยเฉพาะกลุ่มเอกสาร people และ television ส่วนกลุ่มเอกสาร stage และ variety มีการซ้อนทับกับกลุ่มเอกสารอื่นน้อยมาก

ผลจากการวัดค่าอัตราความถูกต้องของชุดการทดลองที่ดีที่สุด มีค่าน้อยกว่าโมเดลที่ดีที่สุด เนื่องจาก ชุดข้อมูล K-dataset นี้มีกลุ่มเอกสารบางกลุ่มจำนวนน้อยเกินไป ซึ่งไม่เพียงพอสำหรับการเรียนรู้ อาทิเช่น กลุ่มเอกสาร entertainment และ art เป็นต้น

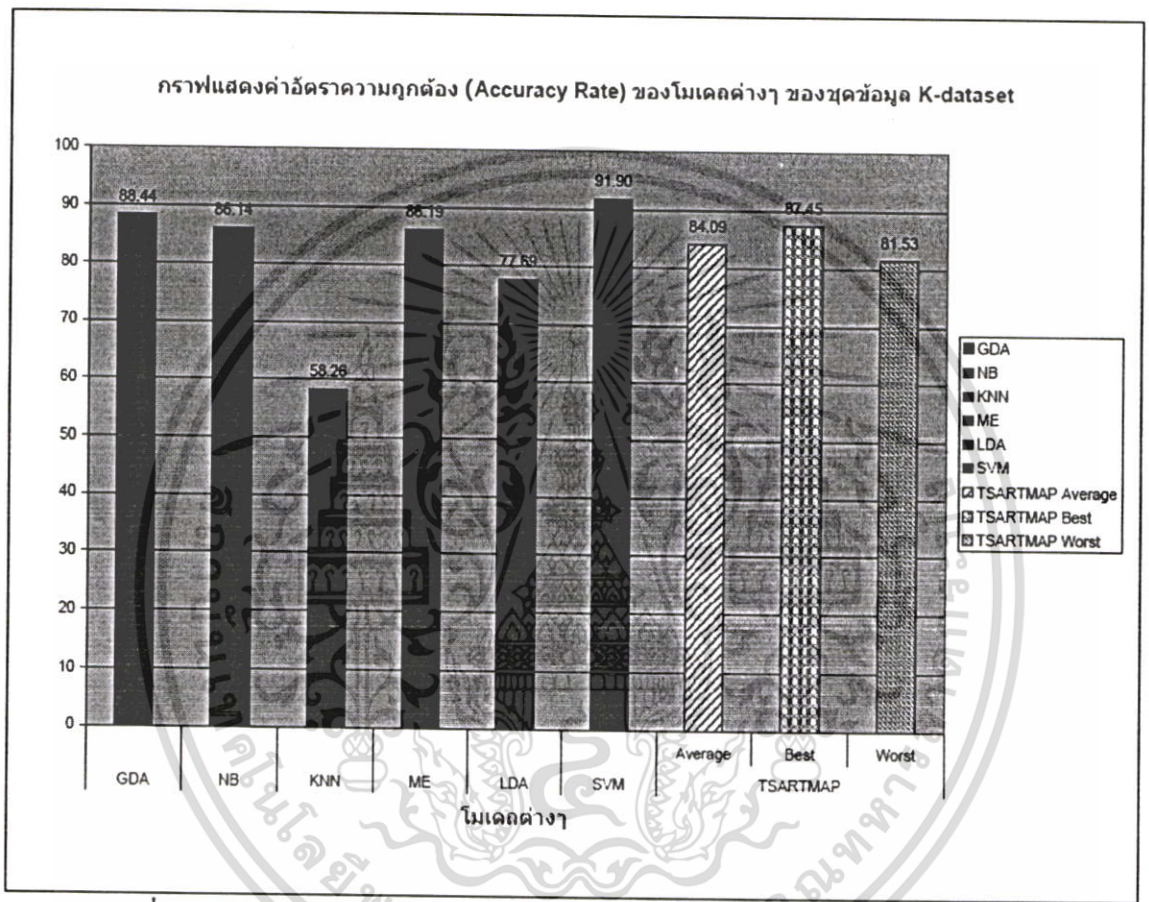
4.6.3 ผลการเปรียบเทียบกับโมเดลต่างๆ ของชุดข้อมูล K-dataset

จากผลการทำงานของโมเดลต่างๆ จากงานวิจัย [18] ได้ใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง คิดเป็นเปอร์เซ็นต์ (%) ของแต่ละโมเดล โดยที่โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พใช้ค่าเฉลี่ยในการเปรียบเทียบกับโมเดลอื่นๆ ซึ่งมีค่าอัตราความถูกต้องเฉลี่ยใกล้เคียงกับโมเดลอื่นๆ รวมทั้งแสดงค่าที่ดีที่สุดและค่าที่เลวร้ายที่สุดของการทดลองแต่ละชุด ดังแสดงในตารางที่ 4.38 และรูปที่ 4.6 จำนวนคำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมนี้จากการทดลองชุดที่ดีที่สุดของชุดข้อมูล K-dataset มีจำนวนคำสำคัญ 9,614 คำ ซึ่งมีรายละเอียดในภาคผนวก ข.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.38 แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดล ของชุดข้อมูล K-dataset

จำนวนข้อมูลที่ใช้ในการทดลอง	โมเดลเปรียบเทียบ โดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง (%)								
	GDA	NB	KNN	ME	LDA	SVM	TSARTMAP		
							Average	Best	Worst
2,340	88.44	86.14	58.26	86.19	77.69	**91.90**	84.09	87.45	81.53



รูปที่ 4.6 กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของชุดข้อมูล K-dataset

ผลจากการเปรียบเทียบโดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง ที่ทดลองกับชุดข้อมูล K-dataset ของโมเดลต่างๆ แสดงให้เห็นว่าโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พสามารถจำแนกประเภทเอกสารได้ถูกต้องในระดับที่ใกล้เคียงกับโมเดลอื่นๆ โดยดูผลที่ได้จากค่าอัตราความถูกต้องเฉลี่ย

4.7 ผลการเปรียบเทียบกับโมเดลต่างๆ ของทุกชุดข้อมูล

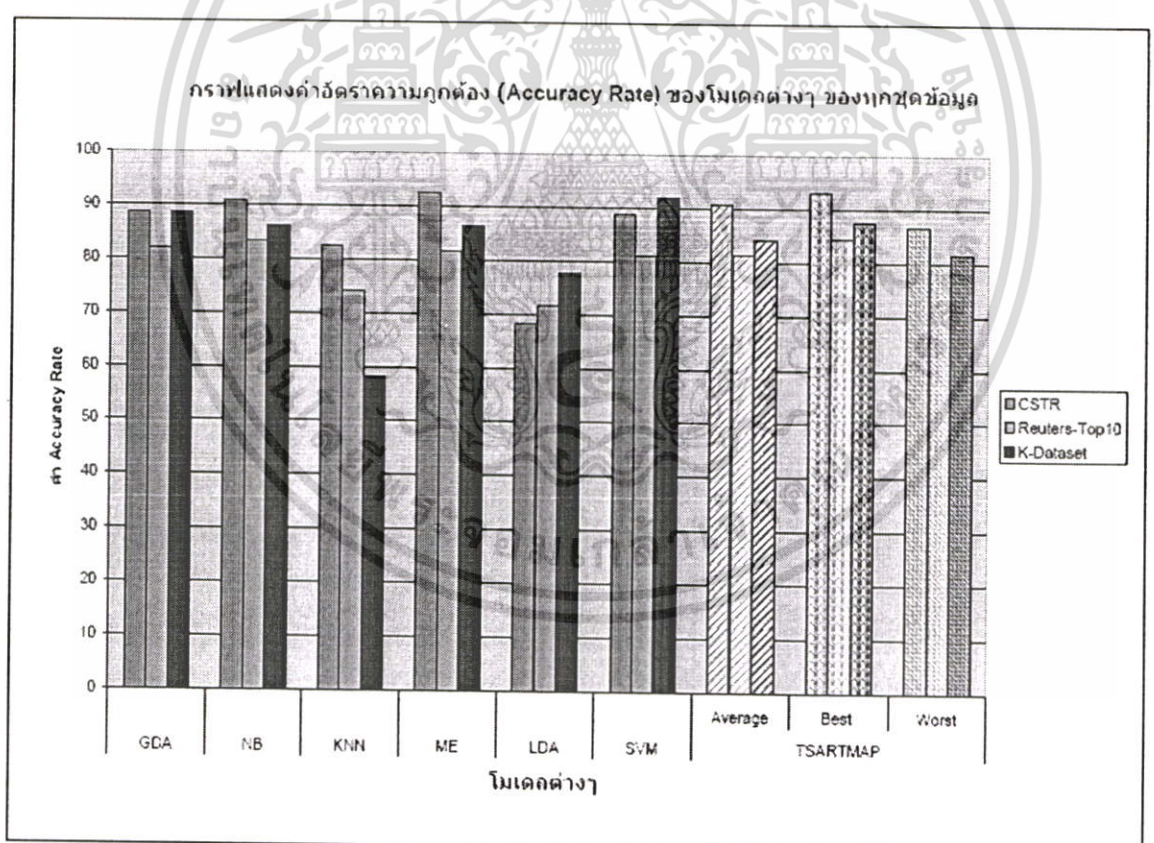
ในการทดลองได้เปรียบกับงานวิจัย [18] โดยใช้ตัววัดประสิทธิภาพ อัตราความถูกต้อง ซึ่งผลของค่าอัตราความถูกต้องเฉลี่ยของชุดข้อมูลแต่ละชุดจะมีค่าใกล้เคียงกับโมเดลอื่นๆ และผลของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความถูกต้องของชุดการทดลองที่ดีที่สุด มีค่ามากกว่าทุกๆ โมเดล จำนวน 2 ชุดข้อมูล คือชุดข้อมูล CSTR และชุดข้อมูล Reuters-Top10 ส่วนชุดข้อมูล K-dataset มีความถูกต้องเป็นลำดับที่ 3 จาก 7 โมเดล ดังแสดงในตารางที่ 4.39 และรูปที่ 4.7 ซึ่งแสดงให้เห็นว่าโครงข่ายประสาทเทียมนี้สามารถจำแนกประเภทเอกสารได้ถูกต้อง

ตารางที่ 4.39 แสดงผลการเปรียบเทียบการทำงานแต่ละโมเดลของทุกชุดข้อมูล

ชุดข้อมูล	จำนวน	โมเดลเปรียบเทียบ โดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง (%)								
		GDA	NB	KNN	ME	LDA	SVM	TSARTMAP		
								Average	Best	Worst
CSTR	476	88.50	90.85	82.53	92.39	68.29	88.71	90.65	**92.91**	86.52
Reuters-Top10	2,775	81.98	83.33	74.07	81.65	71.46	81.13	81.36	**84.24**	80.00
K-dataset	2,340	88.44	86.14	58.26	86.19	77.69	**91.90**	84.09	87.45	81.53



รูปที่ 4.7 กราฟแสดงค่าอัตราความถูกต้องของโมเดลต่างๆ ของทุกชุดข้อมูล

การนำชุดข้อมูลทั้ง 3 ชุด มาใช้ในการทดลอง จะต้องมีการเตรียมข้อมูลเหล่านี้ให้อยู่ในรูปแบบที่จะสามารถนำเข้าสู่โมเดลได้ ซึ่งข้อมูลที่นำเข้าสู่โครงข่ายประสาทเทียมแบบเท็กซิมพลิ-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไฟด์อาร์ทแม็พ สามารถรับข้อมูลที่เป็นข้อความ (Text) ได้โดยตรง โดยไม่ผ่านขั้นตอนการแปลงข้อมูลแบบข้อความไปเป็นตัวเลข ซึ่งการรับข้อความเข้าสู่โมเดลโดยตรงเป็นจุดเด่นของโครงข่ายประสาทเทียมนี้ ลักษณะการทำงานโครงข่ายประสาทเทียมจะเป็นตัวคัดเลือกคำสำคัญที่เหมาะสมหรือคำที่สามารถจำแนกประเภทเอกสารนั้นๆ ให้เอง ส่วนโมเดลอื่นๆ จะรับข้อมูลเข้าเป็นแบบตัวเลข ซึ่งจะต้องมีการแปลงข้อมูลจากข้อความไปเป็นตัวเลขเสียก่อน วิธีที่เป็นที่นิยม คือวิธี Vector Space Model แต่วิธีการนี้จะทำให้สูญเสียเวลาในขั้นตอนการแปลงข้อมูล เนื่องจากจะต้องเลือกคำสำคัญที่จะใช้เป็นตัวแทนของเอกสารให้เหมาะสมเสียก่อน และวิธีนี้ยังเกิดปัญหา High Dimensions เนื่องจากคำสำคัญที่ใช้เป็นตัวแทนของเอกสารแต่ละเอกสารมีจำนวนคำมาก และจำนวนของ Dimension จะมากตามจำนวนของคำสำคัญที่ใช้

ในบทที่ 5 จะกล่าวถึงผลสรุปของงานวิจัยและข้อเสนอแนะ ซึ่งจะเป็นประโยชน์ในการพัฒนาการทำวิจัยต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้ ได้นำเสนอการจำแนกประเภทเอกสาร ด้วยโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ (Text Simplified ARTMAP Neural Network) โดยได้ประยุกต์ใช้โครงข่ายประสาทเทียมแบบซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ (Simplified Fuzzy ARTMAP Neural Network) การวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็ค (Symbolic Objects) และหลักการของโครงข่ายประสาทเทียมแบบเท็กอะแดปทีพเรโซแน้นเทียรี (Text Adaptive Resonance Theory Neural Network) ซึ่งทำให้โครงข่ายประสาทเทียมที่นำเสนอนี้สามารถรับชุดข้อมูลเข้า (Input Pattern) ที่เป็นข้อความ (Text) ได้โดยตรงโดยไม่ผ่านขั้นตอนการแปลงค่าข้อมูล จากค่าเชิงคุณภาพ (Qualitative Value) ไปเป็นค่าเชิงตัวเลข (Numerical Value) ซึ่งจะทำให้ประหยัดเวลาในจัดการกับข้อมูลก่อนนำเข้าสู่โมเดล วิธีการนำข้อมูลที่เป็นข้อความเข้าสู่โมเดลโดยตรงนี้ สามารถช่วยลดปัญหาการเกิด High Dimensions เมื่อมีการแทนเอกสารด้วยวิธี Vector Space Model

คุณสมบัติ (Feature) ของเอกสารที่เลือกมาใช้ในโครงข่ายประสาทเทียมคือ คุณสมบัติชื่อเรื่อง (Title Feature) และคุณสมบัติคำสำคัญ (Keyword Feature) ทำให้โหนดอินพุท (Input Node) ของโครงข่ายประสาทเทียมมี 2 โหนด ส่วนโหนดเอาท์พุท (Output Node) มีได้ไม่จำกัด ขึ้นอยู่กับการกำหนดค่าวิจิแลนซ์ (Vigilance) แทนด้วย ρ ซึ่งเป็นค่าที่ควบคุมการจำแนกกลุ่มในชั้นเอาท์พุทจะมีค่าอยู่ระหว่าง 0.25 ถึง 1 ค่านี้จะกำหนดเริ่มแรกเป็น Base-line Vigilance (แทนด้วย $\overline{\rho}$) เพราะจะมีการเพิ่มขึ้นเรื่อยๆ ในระหว่างการเรียนรู้ การที่กำหนดค่าวิจิแลนซ์ไว้สูง แสดงว่าชุดข้อมูลเข้าต้องคล้ายคลึงกันมากถึงจะจำแนกไว้ในกลุ่มเดียวกันได้ ซึ่งจะทำให้โหนดเอาท์พุทมีขนาดเล็กและมีจำนวนโหนดมาก แต่ถ้ากำหนดค่าวิจิแลนซ์ไว้ต่ำ ชุดข้อมูลเข้าไม่ต้องคล้ายคลึงกันมากก็สามารถจำแนกไว้ในกลุ่มเดียวกันได้ ทำให้โหนดเอาท์พุทมีขนาดใหญ่และมีจำนวนโหนดน้อย

นอกจากนี้ ยังมีส่วนของ Match Tracking ทำหน้าที่ควบคุมความถูกต้องในการเรียนรู้ว่าชุดข้อมูลเข้าถูกจำแนกไว้ในโหนดเอาท์พุท ที่ชี้ไปยังคำตอบ (Category) ในชั้นแคททิกอรี่ (Category Layer) ตรงกับชุดข้อมูลเข้า คือ Category Input ที่กำลังเรียนรู้อยู่หรือไม่ ถ้าคำตอบไม่ตรงกันแสดงว่าชุดข้อมูลเข้าชุดนี้ไม่เหมาะสมที่จะจำแนกอยู่ในโหนดเอาท์พุทนี้ ถ้าเหตุการณ์นี้เกิดขึ้น ระบบจะกำหนดให้โหนดเอาท์พุทที่ชนะในขณะนั้นมีค่าเป็น 0 หรือค่าน้อยๆ เพื่อไม่ให้ระบบกลับมาเลือกโหนดเอาท์พุทนั้นอีก และเพิ่มค่าวิจิแลนซ์ขึ้นเล็กน้อย แล้วหาโหนดเอาท์พุทที่ชนะโหนดถัดไป และตรวจสอบตามขั้นตอนเดิมจนกว่าโหนดเอาท์พุทที่เลือกจะมีคำตอบตรงกับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดข้อมูลเข้า ถ้าคำตอบตรงกัน ถือว่าการจำแนกกลุ่มสำเร็จ เรียกว่าเกิด Resonance ทำการปรับปรุงค่า Top-down Weight เมื่อปรับปรุงเสร็จ แสดงว่าระบบได้เรียนรู้ชุดข้อมูลเข้าชุดนั้นเรียบร้อยแล้ว

ในการทดลองได้เปรียบเทียบกับงานวิจัย [18] โดยใช้ตัววัดประสิทธิภาพอัตราความถูกต้อง (Accuracy Rate) ซึ่งผลของค่าอัตราความถูกต้องเฉลี่ยของชุดข้อมูลแต่ละชุดจะมีค่าใกล้เคียงกับโมเดลอื่นๆ และผลของค่าอัตราความถูกต้องของชุดการทดลองที่ดีที่สุด มีค่ามากกว่าทุกๆ โมเดล จำนวน 2 ชุดข้อมูล คือชุดข้อมูล CSTR และชุดข้อมูล Reuters-Top10 ส่วนชุดข้อมูล K-dataset มีค่าอัตราความถูกต้องเป็นลำดับที่ 3 จาก 7 โมเดล แสดงให้เห็นว่าโครงข่ายประสาทเทียมนี้สามารถจำแนกประเภทเอกสารได้อย่างถูกต้อง

แต่อย่างไรก็ตาม การวัดประสิทธิภาพด้วยตัววัดประสิทธิภาพอัตราความถูกต้อง ไม่ได้เป็นจุดเด่นหลักของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ จุดเด่นของโมเดลนี้คือสามารถรับข้อมูลที่เป็นข้อความ (Text) ได้โดยตรง โดยไม่ผ่านขั้นตอนการแปลงข้อมูลแบบข้อความไปเป็นตัวเลข ซึ่งการรับข้อความโดยตรงนี้ โครงข่ายประสาทเทียมจะเป็นตัวคัดเลือกคำสำคัญที่เหมาะสมหรือคำที่สามารถจำแนกประเภทเอกสารนั้นๆ ให้เอง ส่วนโมเดลอื่นๆ จะรับข้อมูลเข้าเป็นแบบตัวเลข ซึ่งจะต้องมีการแปลงข้อมูลจากข้อความไปเป็นตัวเลขเสียก่อน วิธีที่เป็นที่นิยม คือวิธี Vector Space Model แต่วิธีการนี้จะทำให้สูญเสียเวลาในขั้นตอนการแปลงข้อมูล เนื่องจากจะต้องเลือกคำสำคัญที่จะใช้เป็นตัวแทนของเอกสารให้เหมาะสมเสียก่อน และยังเกิดปัญหา High Dimensions เนื่องจากคำสำคัญที่ใช้เป็นตัวแทนของเอกสารแต่ละเอกสารมีจำนวนคำมาก และจำนวนของ Dimension จะมากตามจำนวนของคำสำคัญที่ใช้

การวัดประสิทธิภาพการทำงานของโมเดลนี้ ได้ใช้ตัววัดประสิทธิภาพ 3 แบบ คือ ตัววัดประสิทธิภาพอัตราความถูกต้อง ตัววัดประสิทธิภาพในงานวิจัยทางด้านการค้นคืนสารสนเทศ (Information Retrieval) 2 แบบคือตัววัดประสิทธิภาพ F-Measure และตัววัดประสิทธิภาพ Entropy ซึ่งตัววัดประสิทธิภาพทั้ง 3 แบบ ได้ค่าจากการคำนวณของชุดการทดลองที่ดีที่สุดอยู่ในระดับที่ดี

5.2 ข้อเสนอแนะ

โครงข่ายประสาทเทียมที่นำเสนอนี้ ได้ทดลองกับชุดข้อมูลเอกสาร (Text Document) จำนวน 3 ชุดข้อมูล ซึ่งได้ผลการทดลองอยู่ในระดับที่ดี นอกจากนี้ยังสามารถนำไปทดลองกับชุดข้อมูลอื่นๆ ได้ และมีความเป็นไปได้ที่จะนำไปประยุกต์ใช้กับ Web Mining โดยไม่ต้องเตรียมข้อมูลก่อนนำเข้าสู่โมเดล (Data Preprocessing) และประยุกต์ใช้กับงานด้านการค้นคืนสารสนเทศเพื่อช่วยการค้นคืนสารสนเทศจากเอกสารที่มีอยู่จำนวนมาก โดยการจำแนกให้เป็นกลุ่มเสียก่อน จากนั้นผู้ใช้สามารถค้นหาเอกสารตามหมวดหมู่ที่ต้องการได้

ปัญหาในการจำแนกเอกสารคือ การจำแนกเอกสารด้วยโครงข่ายประสาทเทียมนี้ จะให้คำตอบเพียงกลุ่มหรือ Class เดียวเท่านั้น เอกสารที่มีการซ้อนทับกัน (Overlap) คือมีคำสำคัญใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณสมบัติต่างๆ อยู่ในหลายเอกสาร อาจทำให้การจำแนกเอกสารผิดกลุ่มได้ มีบางเอกสารเป็นสมาชิกมากกว่า 1 กลุ่ม แต่จำแนกได้เพียง 1 กลุ่ม ทำให้การจำแนกเอกสารในลักษณะนี้ผิดกลุ่มที่ควรจะเป็น

ในการพัฒนาต่อไป อาจเป็นไปได้ที่จะสามารถจำแนกเอกสารได้มากกว่า 1 กลุ่ม เพื่อให้ได้สารสนเทศที่เป็นประโยชน์กับผู้ใช้มากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] Salton G. **Automatic Text Processing: the transformation, analysis and retrieval of information by computer**. New York: Addison Wesley Publishing Company. 1989.

[2] ทรงพล ชุติพงศ์พัฒนกุล. "เท็กโปรเซสซิ่งโคโอเนนนิวรอลเน็ตเวิร์คโดยใช้กระบวนการเรียนรู้แนวใหม่." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2546.

[3] สมคิด แสนเสนาะ. "การแบ่งกลุ่มเอกสารโดยใช้เทคนิคการประมวลผลข้อความด้วยโครงข่ายประสาทเทียมที่เรียนรู้แบบหาผู้ชนะ." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2547.

[4] นรเศรษฐ์ จันทสูตร. "เท็กอะแคปทีพเรโซแน้นเทียรีนิวรอลเน็ตเวิร์ค." วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2547.

[5] Ravi, T.V., and Gowda, K.C. "Clustering of Symbolic Objects Using Gravitational Approach." **IEEE Trans. on Systems, Man, and Cybernetics**, vol. 29, no. 6, 1999. pp. 888-894.

[6] Gowda, K.C. and Diday, E. "Symbolic Clustering Using a New Similarity Measure." **IEEE Trans. on Systems, Man, and Cybernatics**, vol. 22, no. 2, 1992. pp. 368-378.

[7] El-Sonbaty, Y. and Ismail, M.A. "Fuzzy Clustering for Symbolic Data." **IEEE Trans. on Fuzzy Systems**, vol. 6, no. 2, 1998. pp. 195-204.

[8] Baeza-Yates, R. and Ribeiro-Neto, B. **Modern Information Retrieval.** New York : Addison Wesley Publishing Company. 1999.

[9] Aas, K. and Eikvil, L. **Text Categorisation : A Survey.** Oslo : Norwegian Computing Center. 1999.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[10] Fausett, L. **Fundamentals of Neural Networks Architecture, Algorithms and Application**. New Jersey : Prentice Hall International. 1994.

[11] Kasuba, T. "Simplified Fuzzy ARTMAP.", **AI Expert**, vol. 8, Nov, 1993. pp.18-25.

[12] Lewis, D.D. "**Reuters-21578 text categorization test collection distribution 1.0.**", 2004. [Online]. Available : http://www.daviddlewis.com/resources/testcollections/reuters21578/

[13] [Online]. Available : http://www.cs.rochester.edu/trs/

[14] [Online]. Available : http://www.reeltwo.com/datasets.html

[15] [Online]. Available : ftp://ftp.cs.umn.edu/dept/users/boley/PDDPdata/

[16] [Online]. Available : http://thomas.loc.gov/home/stopwords.html

[17] [Online]. Available : http://www.tartarus.org/~martin/PorterStemmer/

[18] Li, T., Zhu, S., and Ogihara, M. "Efficient Multi-way Text Categorization via Generalized Discriminant Analysis." **Proceedings of the Twelfth International Conference on Information and Knowledge Management**, 2003. pp. 317-324.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.
แสดงตัวอย่างชุดข้อมูลต่างๆ ที่ใช้ในการทดลอง

ก.1 ชุดข้อมูล CSTR

ชุดข้อมูล CSTR มีรายละเอียดในขั้นตอนการเตรียมเอกสารต่างๆ ดังนี้

ก.1.1 ไฟล์เอกสาร CSTR ในรูปแบบ HTML

รูปแบบของเอกสารก่อนการตัดแท็ก เป็นดังนี้

<LI>Scott, M.L., "Non-Blocking Timeout in Scalable Queue-Based Spin Locks", TR773, Computer Science Dept., U. Rochester, February 2002.
<A href="ftp://ftp.cs.rochester.edu/pub/papers/systems/02.tr773.Non-blocking_timeout_in_scalable_queue-based_spin_locks.pdf.gz">02.tr773.Non-blocking_timeout_in_scalable_queue-based_spin_locks.pdf.gz</A>

<P>Keywords: non-blocking synchronization; timeout; user-level spin locks; preemption; scalability.

<BLOCKQUOTE>Queue-based spin locks allow programs with busy-wait synchronization to scale to very large multiprocessors, without fear of starvation or performance-destroying contention. Timeout-capable spin locks allow a thread to abandon its attempt to acquire a lock; they are used widely in real-time systems to avoid overshooting a deadline, and in database systems to recover from transaction deadlock and to tolerate preemption of the thread that holds a lock. In previous work we showed how to incorporate timeout in scalable queue-based locks. Technological trends suggest that this combination will be of increasing commercial importance. Our previous solutions, however, require a thread that is timing out to handshake with its neighbors in the queue, a requirement that may lead to indefinite delay in a preemptively multiprogrammed system. In the current paper we present new queue-based locks in which the timeout code is

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

non-blocking. These locks sacrifice the constant worst-case space per thread of our previous algorithms, but allow us to bound the time that a thread may be delayed by preemption of its peers. We present empirical results indicating that space needs are modest in practice, and that performance scales well to large machines. We also argue that constant per-thread space cannot be guaranteed together with non-blocking timeout in a queue-based lock. </BLOCKQUOTE>

ก.1.2 ไฟล์เอกสาร CSTR ที่ตัดเอาเฉพาะข้อมูล Title และ Keyword

ในส่วนของ Title ตัดเอาเฉพาะข้อความที่อยู่ในเครื่องหมาย " " และส่วนของ Keyword ตัดเอาข้อมูลในแท็ก <BLOCKQUOTE>...</BLOCKQUOTE> ทั้งหมด ซึ่งมีรายละเอียดดังนี้

ส่วนของ Title เป็นดังนี้

Non-Blocking Timeout in Scalable Queue-Based Spin Locks

ส่วนของ Keyword เป็นดังนี้

Queue-based spin locks allow programs with busy-wait synchronization to scale to very large multiprocessors, without fear of starvation or performance-destroying contention. Timeout-capable spin locks allow a thread to abandon its attempt to acquire a lock; they are used widely in real-time systems to avoid overshooting a deadline, and in database systems to recover from transaction deadlock and to tolerate preemption of the thread that holds a lock. In previous work we showed how to incorporate timeout in scalable queue-based locks. Technological trends suggest that this combination will be of increasing commercial importance. Our previous solutions, however, require a thread that is timing out to

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

handshake with its neighbors in the queue, a requirement that may lead to indefinite delay in a preemptively multiprogrammed system. In the current paper we present new queue-based locks in which the timeout code is non-blocking. These locks sacrifice the constant worst-case space per thread of our previous algorithms, but allow us to bound the time that a thread may be delayed by preemption of its peers. We present empirical results indicating that space needs are modest in practice, and that performance scales well to large machines. We also argue that constant per-thread space cannot be guaranteed together with non-blocking timeout in a queue-based lock.

ก.1.3 ข้อมูล Title นำไปตัด Stopwords และ ข้อมูล Keyword นำไปหาคำสำคัญ ด้วยโปรแกรม Copernic Summarizer

ข้อมูล Title เป็นดังนี้

non-blocking, timeout, scalable, queue-based, spin, locks

ข้อมูล Keyword เป็นดังนี้

locks, timeout, queue-based, preemption, non-blocking, performance-destroying, multiprocessors, scale, spin, constant, per-thread, space, argue, machines, practice, needs, indicating, peers, algorithms, code, indefinite, delay, queue

ก.1.4 นำข้อมูลทั้ง Title และ Keyword ไปหา Word Stemming

ข้อมูล Title เป็นดังนี้

non-block, timeout, scalabl, queue-base, spin, lock

ข้อมูล Keyword เป็นดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

lock, timeout, queue-base, preemption, non-block, perform-destroi
multiprocessor, scale, spin, constant, per-thread, space, argu
machin, practice, need, indic, peer, algorithm, code, indefinit
delai, queue

ก.2 ชุดข้อมูล Reuters-Top10

ชุดข้อมูล Reuters-Top10 มีรายละเอียด ดังนี้

ก.2.1 ไฟล์เอกสาร Reuters-Top10 ในรูปแบบ XML

รูปแบบของเอกสารก่อนการตัดแท็ก เป็นดังนี้

<?xml version="1.0"?>
<MESSAGE>
DANISH CROPS TWO WEEKS BEHIND, OFFICIAL REPORT

COPENHAGEN, June 29 - Danish crops are up to two weeks behind normal growth levels due to the cold weather in May and June, when mean temperatures were up to four centigrade below average, the State Plant Cultivation Bureau said in a report.

At the season's first crop test on June 26, the bureau rated crops at an overall 94, the same as on June 20 last year. The figure of 100 represents normal in a year of average growth conditions with no crop damage.

The test gave ratings expressed as a factor of 100, as follows - winter wheat 96 (last year 91), winter rye 97 (98), winter barley 92 (90), spring barley 94 (98), oats 96 (98), fodder and sugar beets 87 (96), winter rapeseed 98 (94), spring rapeseed 94 (98).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Final harvest figures for 1986 were not yet available.

Reuter

</MESSAGE>

ก.2.2 ไฟล์เอกสาร Reuters-Top10 ที่ตัดเอาเฉพาะข้อมูล Title และ Keyword

ส่วนของ Title เป็นดังนี้

DANISH CROPS TWO WEEKS BEHIND, OFFICIAL REPORT

ส่วนของ Keyword เป็นดังนี้

Danish crops are up to two weeks behind normal growth levels due to the cold weather in May and June, when mean temperatures were up to four centigrade below average, the State Plant Cultivation Bureau said in a report.

At the season's first crop test on June 26, the bureau rated crops at an overall 94, the same as on June 20 last year. The figure of 100 represents normal in a year of average growth conditions with no crop damage.

The test gave ratings expressed as a factor of 100, as follows - winter wheat 96 (last year 91), winter rye 97 (98), winter barley 92 (90), spring barley 94 (98), oats 96 (98), fodder and sugar beets 87 (96), winter rapeseed 98 (94), spring rapeseed 94 (98).

Final harvest figures for 1986 were not yet available.

หลังจากนั้น นำข้อมูล Title ไปตัด Stopwords และ นำข้อมูล Keyword ไปหาคำสำคัญ ด้วยโปรแกรม Copernic Summarizer แล้วนำข้อมูลที่ได้ทั้งข้อมูล Title และข้อมูล Keyword ไปหา Word Stemming

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.3 ชุดข้อมูล K-dataset

ชุดข้อมูล K-dataset มีรายละเอียด ดังนี้

ก.3.1 ไฟล์เอกสาร K-dataset ในรูปแบบ HTML

รูปแบบของเอกสารก่อนการตัดแท็ก เป็นดังนี้

```
http://www.yahoo.com/text/headlines/970908/entertainment/stories/art_holocaust_1.html
<html>
<head>
<title>Yahoo! - Jewish Group To Form Unit On Looted Art</title>
</head>
<body>
:
:
<!-- TextStart -->
<p>
    NEW YORK (Reuter) - The Jewish group that negotiated with
Switzerland for the return of unclaimed Holocaust-era accounts
now plans to investigate what happened to the billions of
dollars worth of art looted by the Nazis.
<p>
    The World Jewish Restitution Organization (WJRO) plans at
an executive meeting Tuesday to set up a special commission
headed by Museum of Modern Art chairman Ronald Lauder, a former
U.S. ambassador to Austria, to look into the question of looted
art and what should be done, a group spokesman said.
<p>
    The WJRO, a coalition of several Jewish groups plus Israel,
is headed by World Jewish Congress President Edgar Bronfman, who
led it in negotiations with Switzerland on the question of
unclaimed Holocaust-era accounts in Swiss banks.
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<p>

A WJRO spokesman said that a search of Second World War archives showed that some 55,000 works of art stolen by the Nazis in France alone were not returned to their rightful owners after the war.

<p>

Most of the works stolen in France, including Impressionist and Old Master paintings, belonged to Jewish owners, the group spokesman said.

<p>

He added that of the 55,000 works, 2,058 hang in French museums and 14,000 were publicly auctioned off by the French government after the war.

<p>

The WJRO is also expected to make a decision at its Tuesday meeting on the first allocation of money from a special fund set up by Swiss banks and businesses to aid Holocaust survivors.

<p>

The group was expected to approve the allocation of $12 million in October from the $200 million fund to help East European Holocaust survivors.

<p>

The fund was set up by Swiss banks and businesses in response to the controversy, which followed revelations that Switzerland did extensive business with the Nazis, including selling and laundering of looted gold.

<p>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    Reuters/Variety
<!-- TextEnd -->
:
:
</body>
</html>
```

ก.3.2 ไฟล์เอกสาร K-dataset ที่ตัดเอาเฉพาะข้อมูล Title และ Keyword

ส่วนของ Title เป็นดังนี้

```
Jewish Group To Form Unit On Looted Art
```

ส่วนของ Keyword เป็นดังนี้

```
ตัดข้อความตั้งแต่ <!-- TextStart -->... <!-- TextEnd -->
```

หลังจากนั้น นำข้อมูล Title ไปตัด Stopwords และ นำข้อมูล Keyword ไปหาคำสำคัญ ด้วยโปรแกรม Copernic Summarizer แล้วนำข้อมูลที่ได้ทั้งข้อมูล Title และข้อมูล Keyword ไปหา Word Stemming

เมื่อตัด Word Stemming เรียบร้อยแล้ว ถือว่าสิ้นสุดขั้นตอนการเตรียมข้อมูล หลังจากนั้นนำข้อมูลที่ได้ ไปทำให้อยู่ในรูปแบบของแฟ้มข้อมูลที่จะนำเข้าสู่โมเดล ด้วยโปรแกรม MATLAB ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข.
แสดงตัวอย่างของคำสำคัญที่ได้จากการทดลองของชุดข้อมูลต่างๆ

ข.1 คำสำคัญที่ได้จากชุดข้อมูล CSTR

คำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ของชุดข้อมูล CSTR ชุดที่ดีที่สุด มีจำนวน 1,555 คำ ดังต่อไปนี้

3d	affin	applet/bytec	assur	belong	burkhardt
aba	affix	appli	asynchron	benchmark	burstein
abandon	agent	applic	attent	beta	bursti
abil	agent-base	applic-level	augment	better	busi-wait
absolut	aggreg	apportion	autoepistem	bh97	bytecod
abstract	aggress	approach	autom	bia	cach
acceler	agreement	approxim	automat	biochem	cach-miss-in
accept	agument	arbitr	automata	bit	cai
access	airlift	arbitrari	auxiliari	black-and-wh	cai-furst
accommod	algebra	arbter	auxiliari-pu	block	calibr
accomplish	algorithm	architectur	averag	boltzmann	camera
accord	allend	architectur/	axiomat	bonu	capabl
account	allevi	area	backward	boolean	capac
achiev	alloc	argu	bair	born	captur
acquisit	almost-every	argument	balanc	boston	car
act	alphabet	arithmet	bandwidth	bottleneck	care
action	altern	arma	barrier	bottom-up	carrol
activ	amort	arrai	base	bought	carv
acycl	amount	arriv	baselin	bound	cascad
adapt	amplif	art	basi	boundari	cashmer
adjac	analysi	articul	basic	bpp	categor
adjust	analyz	artifact	basket	brain-comput	categori
adopt	anim	aspect	bayesian	branch	causal
advantag	annot	assess	beach	brassard	caus-effect
adventur	anteced	assign	becom	brown	cc-numa
adverbi	apart	assist	beek	bu-base	cell
advic	api	associ	behavior	buffer	censu
advisor	appear-base	assum	belief	build	center
affect	appendic	assumpt-base	believ	built	central

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

certif	collabor	connect	cortex	deictic	displai
chang	collaps	conp	cortic	delet	dissert
channel	collect	consciou	co-specifi	deleteri	dissip
character	collector	consequ	cost	deliber	distanc
characterist	combin	consider	count	demonstr	distant
check	combinatori	consist	counterpart	denot	distinct
checker	commit	constrain	countryu	dens	distinguish
chemistri	committ	constraint	coupl	densiti	distract
chief	common	construct	creat	densiti-base	distractor
chines	commonplac	consum	creation	depend	distribut
chip	commun	consumpt	criterion	depth-first	distribut-me
choic	commut	content	crossov	dequeu	divers
choos	compact	content-free	cryptograph	deriv	dividend
circuit	compar	context	cue	descriptor	dna
circular	comparison	contextu	cull	design	dna-chemic
circumnavig	compat	continu	curv	detail	dnp
citi	compil	contour	curve	detect	dodgson
claim	compil-time	contrast	custom	determin	domain
clariti	complet	contribut	cutpoint	determinist	domin
class	complex	control	cycl	develop	downward
classic	complex-theo	convent	danc	devic	draw
classif	complianc	converg	data	dextrou	drift
classifi	complic	convers	databas	dialog	drive
claus	compon	convex	datacub	dialogu	drug
client-serve	compon-base	convoi	dataset	dichotomi	durat
close	composit	cooper	deadlin	differ	dvi2ps
closest	comprehens	coordin	deadlock	dimens	dynam
closur	compress	core	deal	direct	e-admiss
closur-base	compris	corefer	debug	disadvantag	eager
cluster	comput	corel	decis	disambigu	easi
clutter	concept	corpora	decis-make	disciplin	easi-hard
coars	concurr	corpu	decomposit	discours	easili
code	condit	corpu-base	decreas	discov	easi-to-us
cofewp	confer	correct	deduct	discover	econom
cognit	congress	correctli	default	discoveri	eeg
coher	congression	correl	definit	discret	effect
coign	conjectur	correspond	degre	disjoint	efficaci
coincid	conjunct	corrupt	degre-of-fre	disk	effici

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

egg
elect
element
elimin
embed
embodi
emerg
emphasi
empti
enabl
encod
encount
endstop
energi
engin
eniron
entropi
enumer
envirobn
environ
ephor
equal
equival
error
estim
euclidean
evalu
event
event-base
evid
evok
evolut
evolutionari
exact
exact-comput
exceed
exclus
execut
exhibit
exig
exist
existenti
exp
expect-maxim
experi
experiment
explan
explicit
explicitli
exploit
explor
exponenti
exposur
express
extend
extens
extra
extract
extractor
extrem
ey
facil
facilit
faculti
fail
fair
fairer
fall
fals
fast
faster
fastest
fault
faulti
fault-toler
feasibl
featur
featur-base
featureless
feedback
feedforward
feigenbaum
ferguson
fewnp
fewp
field
file
filter
filter-base
filter-like
final
find
finit
finit-state
fire
first
fix
fixat
flag
flaw
flexibl
forai
forc
forc-control
form
formal
formul
formula
foundat
fourier
fraction
frame
frame-base
frameview
framework
free
freedom
freeli
free-view
frequenc
frequent
front
fulli
function
futur
gain
galaxi
gallei
game
gap
garbag
gaze
gener
generaliz
genet
geometr
gestur
given
goal
goal-orient
good
grammar
granular
graph
grasp
gripper
grollmann
ground
groundwork
group
grow
guarante
guard
gui
guid
hand
haptic
hard
hardwar
harm
hartmani
hash
head
head-to-head
hea
heavili
helmet
help
hemaspaandra
heterogen
hgp
hidden
hide
hierarch
hierarchi
high
higher
highest
high-frequen
high-level
highlight
hinkelman
histori
hmd
hmm
hole
horn
hull
human
hybrid
hypothet
i/o
ibm
ica
icon
identif
idl
illustr
ilp
imag
imageflow
imageri
immun
impact
implement
impli
import
improv
inact
incept
inclus
incomplet
inconsist
incorpor
incorrect
increas
increment
incur
inde
independ
indic
individu
indoor/outdo
ineffici
infer
inferenc
influenc

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

inform	intuit	lattic-theor	low	mean	modular
infrastructu	invalid	lda	lower	measur	modulatori
inher	invari	learn	low-level	mechan	modulo
inherit	invers	least-squar	lwpp	media	molecul
inhibit	invoc	legaci	lymphocyt	mediat	moment
initi	irix	lemma	machin	medic	monitor
input	iscan	level	machineri	member	monoton
input/output	isol	leverag	magnitud	membership	monro
input-driven	i-speech	lewi	mai	memori	moreov
insecur	item	lexic	mainstream	memori-map	morph
insight	itemset	li	maintain	mercuri	morpholog
instanc	java	light	major	mere	motion
instanc-base	jit	light-weight	mammalian	merg	motiv
instruct	join	likelihood	manag	mesh-connect	move
insuffici	joint	limit	maneuv	messag	movement
integ	justifi	lineal	mani-on	messi	msu
integr	just-in-time	linear	manipul	method	multi-agent
intellectu	jvm	linear-nonde	manner	methodolog	multihead
intellig	kalman	link	manual	metric	multi-model
intend	karp-lipton	link-base	map	microarchite	multi-parti
interact	kart	lisp	mark	middlewar	multipl
interconnect	kei	list	marker	milan	multipl-goal
interfac	kernel	literatur	markov	mine	multiprocess
interfc	kernel-appli	live	martin-l	minim	multiprogram
interfer	kernel-kerne	load	masquerad	minimum	multi-select
interleav	know	local	mat	minor	multi-speake
intern	knowledg	local-access	match	mint	mundi
internet	kolmogorov	locat	matric	mirror	mutual
interpret	kolmogorov-s	lock	matrix	mispredict	mv200
inter-queri	label	lock-free	matroshka	miss	myth
intersect	lack	logarithm-sp	matter	mix-initi	name
inter-transa	lambda	logic	matur	mobil	nativ
interv	languag	logspac	maxim	modal	natur
intervent	larg	longer	maxim-inform	model	navig
inton	larg-scale	longstand	maximum	model-base	nearbi
intra-queri	latenc	loop	mdl	moder	nearer
introduc	later	loos	mdp	modifi	nearest
introduct	latter	lost		modul	necessarili

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

neg
neighbor
netprof
network
network-base
neural
neuron
neurophysiol
nevertheless
new
node
nois
nomin
nonadapt
non-convex
non-dens
nondetermin
nondetermini
nonhomogen
nonlinear
non-modular
non-monoton
nonparametr
nonrigid
non-speech
non-stationa
normal-form
notat
note
notion
notori
noun
novel
novic
np
np-complet
npmv
np-select
numa
number
numer
object
object-affin
object-base
observ
obtain
occlud
occlus
occup
odometri
offload
off-the-shel
ogiwara
on-dimension
on-line
ontolog
on-wai
open
oper
opportun
optim
optimist
oracl
order
ordinari
organ
orient
origin
orthogon
orthograph
outgrowth
outperform
output
overal
overcom
overhead
overli-const
overproduct
overs
over-the-cou
p300
pair
pan
paper
paradigm
parallel
paramet
pars
parti
partial
partit
pass
path
patient
pattern
pebbl
pedict
peer
percept
perceptu
perceptuo-mo
perfect
perform
permut
persist
phase
phenomena
phenomenon
phone
phonet
photo
photo-consis
photograph
phrase
pictur
p-immun
pl
plai
plan
plane
planner
platform
player
ploxoma
plu
p-m-complet
p-m-degre
point
point-base
pointer
polhemu
polhemu-puma
polici
polici-insen
polynomi
polynomi-tim
poor
popular
port
portabl
pose
posit
posteriori
post-parser
post-process
postscript
potenti
power
practic
precis
predic
predict
predictor
preemptibl
preemption-s
prefetch
prefix
preliminari
prepar
prescript
presenc
preserv
presum
presupposit
previous
primari
primarili
prime
primit
principl
prior
prioriti
privat
probabilist
probabl
problem
problem-solv
proceed
process
processor
produc
profil
program
programm
project
promis
pronomin
pronoun
proof
properti
proportion
propos
prosodi-wise
prospect
protocol
prototyp
provabl
prove
provid
provoc
p-select
pseudo
pseudo-code
pseudorandom
pspace
psych
psychophys
purpos
pushdown
q-learn
qsl
qualiti
quantif
quantifi
quantum
quasi-static
queri
question
queue
quorum
rabi
race
radio
rais
random
rang
rapid

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

rapidli
rate
ratio
reader
reader-write
real
realist
realiti
realiz
real-life
real-time
real-valu
real-world
reason
recaptur
receiv
recenc-base
recept
reciproc
recogn
recognit
recommend
reconcil
reconstruct
record
recov
recoveri
rectif
recurr
recurs
reduc
reduct
re-establish
reevalu
refer
refresh
refut
regimen
region
regist
regularli-st
reinforc
rel
relat
related
relationship
relativ
relearn
relev
reli
reliabl
reliev
remaind
remot
render
repair
repeat
repetit
report
represent
request
requir
research
residu
resolut
resolv
resourc
respond
respons
restaur
restor
restrict
restrictt
restructur
result
resutl
reus
reusabl
reveal
review
re-weight
rhet
rhetor
rice-style
rich
rivest
robot
robust
role
rotat
roth
round
rout
routin
row
rule
run
run-time
saccad
sacrif
salienc
sampl
sandewal
sarsa
satan
satisfact
satisfi
scalabl
scale
scale-up
scenario
scene
schedul
schemata
scheme
scope
sdsm
s-dsm
search
second
second-level
secret-kei
secur
seemingli
segment
select
self-check
self-guid
self-reduc
self-specifi
self-wit
semant
semi-decis
semi-rankabl
sens
sensori
sensori-moto
separ
sequenc
sequent
sequenti
serial
serializ
set
set-base
settl
shape
share
share-memori
shepherd-com
sherman
shift
show
signal
signatur
signific
sikorski
similar
simon
simplic
simplifi
simul
simul-anneal
simultan
singl
singl-head
singl-level
singular
situat
situat-theor
size
skill
sky
slight
slog
slot-fifo
slow
smarter
smooth
smp
smp-base
soft
softwar
solut
solut-phase
solv
sort
sourc
space
space-bound
space-effici
space-time
spade
span
spars
sparta
spatial
spatiotempor
spatula
speaker
special
special-purp
specif
specifi
spectrum
speech
speed
speedup
speed-up
spehlmann
spent
spike
spin
spin-down
spin-up
split
spoken
spontan
spot
spp
spread
spun
squar
stack
standard
state
state-action

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

statement	superset	tempor	train-95	uncertainti	vari
static	supervis	tendon	train-96	unclear	variabl
statist	support	tension	trajectori	uncommun	varianc
stem	surfac	term	transact	unconstrain	variant
stent	surpris	terminolog	transcript	underli	variat
stereo	survei	test	transform	underproduct	varieti
stick	svm	testb	transform-in	understand	vector
stimuli	switch	theorem	transit	unexpect	vehicl
stochast	symbol	theoret	translat	unifi	veloc
store	symposium	theori	transmiss	uniform	verb
stori	synchron	third	transpar	uniqu	verifi
strang	syntact	thread	transput	unit	version
strategi	syntax	threshold	trap	univers	versu
stream	synthesi	tight	traum	unix	vertic
strict	synthet	tightest	travelstar	unknown	viciou
string	system	time	travers	unlik	video
strong	system-appli	tod	tree	unlimit	view
structur	systemat	toda	tremend	unp	viewer
studi	tactic	toler	trial	unrel	viewpoint
sub-goal	taho	tool	trivial	unreli	virtual
subgraph	talli	toolkit	true	unseen	visibl
sub-lattic	tangent	topolog	truth-tabl	unsupervis	vision
subpopul	tape	tori	tube	updat	vision-base
subset	target	total	tune	upper	visual
substructur	task	tpd	ture	upward	vot
subtract	task-base	trace	ture-complet	us	vpl
success	task-orient	trace-driven	tutori	usabl	vr
suffer	taxonomi	tracer	twice	usag	vr4
suffic	tea-1	track	two-dimensio	user	wait
suffici	techinqu	track-and-co	two-head	useru	wang
suggest	techniqu	tracker	two-level	usual	warrant
suit	technolog	trade	two-wai	utah/mit	wave
suitabl	teleassist	tradeoff	type	util	weak
summari	teleoper	trade-off	typic	utter	weaken
summat	telephoni	tradit	udm	valid	wear
supercomput	tell	trailer	unambigu	valoi	weight
superexponen	teller	train	unauthent	valu	well-founded
superior	templat	train93	uncertain	van	wherev

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

wider	window	withdraw	wordnet	workshop	zero
wide-rang	wireless	within-class	work	world	zissermann
wide-us	wit	word	workload	yield	

ข.2 คำสำคัญที่ได้จากชุดข้อมูล Reuters-Top10

คำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ของชุดข้อมูล Reuters-Top10 ชุดที่ดีที่สุด มีจำนวน 5,085 คำ ดังต่อไปนี้

abandon	account	adob	ag	akiman	alon
abat	accru	adopt	aga	akio	al-otaibi
abbett	accumul	adt	ageement	akron	al-sabah
abbott	accus	adult	agenc	alabama	altern
abboud	acg	advanc	agenda	al-ahm	alto
abc	achiev	advantag	agent	alan	altogeth
abdul	acm	adventur	aggreg	alaska	altron
abdul-aziz	acquir	advers	aggress	alba	altu
abil	acquisit	advic	agip	al-bader	aluminium
abitibi-price	acquisiton	advis	ago	albani	aluminum
ablaz	acquisitt	advisor	agre	alberta	alver
abnorm	acqur	advisori	agreement	alcan	am
abolish	acr	aegean	agricola	alert	amadeu
abolit	acreag	aegon	agricultur	alex	amatil
abroad	act	ael	agrimont	alexand	amax
absenc	action	aep	agrio	alfa	ambassador
absorb	activ	aerospac	agrnom	algeria	ambrosiano
absorpt	ad	affair	ahead	ali	amca
abu	adam	affect	ahmanson	allan	amcast
abus	add	affili	aid	alleg	amend
academ	adddit	afford	aim	allegheni	amendend
academi	addison	africa	air	allegi	america
acceler	addit	african	airbu	allegiswestin	american
accept	address	after	aircraft	alli	american-flag
access	adequ	aftermath	airlin	allianc	american-isra
accessori	adjourn	afternoon	airport	alli-signal	amertek
accompani	adjust	aftertax	airsensor	alloc	amex
accomplish	adm	after-tax	ajust	allow	amfac
accord	administr	afterward	aker	allwast	amman

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

amoco	anti-takeov	argosystem	associ	automobil	ballot
amort	antitrust	arieff	assum	autotrol	ballpark
amoskeag	anygrain	aris	assumptrion	aveng	baltimor
amount	anzola	aristech	assur	averag	balzac
amp	apart	aritech	astra	avg	ban
ampl	apartheid	ariz	asylum	avgpric	banc
amr	aphid	arkansa	atcor	aviat	bancgroup
amsouth	api	arm	athei	avoid	banckshar
amsterdam	apiec	armel	athen	avon	banco
amt	apollo	armi	athlon	avx	bancorp
amus	appar	armor	atla	award	bancorpor
ana	apparel	aromat	atlant	awhil	bancshar
analyst	appeal	aros	atlanta	axilrod	banctexa
anchor	appl	arous	atlanti	axl	band
anchorag	appli	arrai	attack	azpur	bangladesh
ancillari	applic	arrang	attend	bach	bank
and	appoint	arriv	attent	back	bank/imf
andrea	appreci	arrog	attornei	backfir	bank/north
anew	approach	arrow	attract	background	bankeast
angel	appropri	artic	attribut	backlog	banker
anger	approv	articl	atwood	bad	bankruptci
angola	approxim	artilleri	auction	bag	bankshar
anheus-busch	april	arturo	audio/video	baggag	banner
anhui	aqazadeh	aruba	audiovisu	baghdad	banponc
anim	aquino	asa	audit	bahrain	bar
announc	aquir	asamera	auditor	baht	barbara
annual	arab	asbesto	aug	bai	barber
ansett	arabia	aschoff	augenthal	bake	barco
answer	arabian	ashraf	auger	baker	bare
anthoni	aramco	asia	august	baker/stolten	bargain
anti-american	arbitr	asian	austec	bakeri	barlei
anticanc	archer-daniel	ask	auster	bakeriesbid	barnei
anticip	architect	assault	austral	balanc	barnett
anti-dump	arco	assembl	australasia	baldrig	barrag
anti-inflamma	arden	assess	australia	bale	barrel
anti-inflat	area	asset	australian	balfour	barri
antil	argentin	assist	author	ball	barrier
	argentina	assoc	authoris	balladur	base

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

baseload
basi
basic
basin
basix
basket
bass-led
bastian
batch
batignol
batter
batteri
battl
bavadra
baxter
baybank
bazoli
bbc
bbl
bcv
bcw
beach
beacon
bean
bear
beard
bearish
beaufort
bec
beckman
becor
bed
beef
beer
beet
beforehand
begin
begun

beirut
belgian
belgium
belgrad
believ
bell
bellmon
bellsouth
belo
beloit
belt
ben
bench
benedetti
benefici
benefit
benel
bennett
bensenvil
benson
benzen
berg
bergen-richar
berkei
bernardino
bernic
best
better-than-e
beverag
bevi
bfi
bhd
bi-annual
bicol
bid
bidder
bidtek
bigger

biggest
bii
bilater
bildner
bilion
bill
billion
billlion
bilzerian
biolog
biotech
biotechnolog
biphenyl
blah
blame
blaze
blinch
block
blockad
blockag
blood
blount
blown
blue
bluegrass
blunt
board
boart
boat
bodi
boe
boiler
bolivar
bolster
bomb
bombai
bond
bonn

bonu
bonus
book
boom
boost
booth
borax
borden
border
borg-warner
borrow
boston
botchwei
both
bottl
bottom
bought
bound
boveri
bowat
bp
bpcc
bpd
brace
bracket
brake
bramal
bran
branch
brand
brass
brazil
brazilian
breach
breadmak
break
breakdown
bregman

brenda
brent
brenton
brew
brewer
brian
bridg
brief
brine
bristol
bristold
bristol-myer
britain
british
british-base
brito
bro
broad
broadcast
broadli
brockton
brockwai
brodi
broke
broken
broker
brokerag
brook
brooklyn
brophi
brother
broussard
brown
brown-ferri
bruce
brussel
bryan
bsd

bsn
btx
bu
budget
buffalo
buffet
bug
bui
build
builder
build-up
built
built-in
bui-out
buitoni
bulk
bulletin
bullish
bumper
bundesbank
bung
burden
burdensom
bureau
burgdorff
burlington
burmah
burn
burndi
burnham
buse
bush
bushel
busi
busin
butan
buttros
buyer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

buyout
buyt
c.o.m.b.
cabinet
cabl
cablevis
cade
cadillac
cadnetix
caesar
cafeteria
cai
cain
cairn
caiss
cake
cal
calcul
calendar
calf
calif
califora
california
call
calm
calmark
calmat
calstar
cam
cama
camp
campaign
campbel
campeau
can
canada
canada-dome
canadian
canadian/bbl
canbra
cancel
candid
cannon
canola
capac
capacitor
cape
capit
capitain
capitol
captain
captur
car
caraca
carbid
card
cardi
cardiovascula
care
carefulli
caremark
cargil
cargo
caribbean
carl
carlo
carlson
carmark
carnei
carolco
carolina
carri
carrier
carrol
carryforward
carson
cartel
carter
casablanca
case
cash
caspar
cassava
castl
castor
casualti
catalog
categori
caterpillar
cattl
caus
caution
cayuga
cb
cbc
cbo
ccc
ccf
ccr
ccx
cd
cdc
cdi
cdu
cea
ceas
ceil
celanes
celebr
celeron
cellular
cement
cent
cental
centel
center
centerbank
centerr
centimetr
centrabank
central
centr-west
centuri
cepe
ceram
cereal
certain
certif
cetu
cfbk.o
cfcf
cff
chain
chairman
chaitanya
challeng
champion
champlin
chan
chanc
chancellor
chanceri
chang
channel
chanthong
chaotic
charact
charg
chargeoff
charger
charg-off
chariot
charl
chart
charter
charter-crell
chase
chattanooga
cheat
check
cheif
chemetron
chemic
chemlawn
cherrin
chesebrough
chevron
cheyenn
chicago
chicken
chief
child
chimerin
chimi
china
chip
chittagong
chose
christian
christi-tyler
christoph
chrobok
chronar
chronowitz
chrysler
church
ci
ciba-geigi
cie
cif
cigar
cincinnati
circl
circuit
circul
cite
citi
citibank
citicorp
citizen
citru
cityf
civil
civilian
claim
clara
clarifi
clark
clash
class
classic
classifi
clayton
clayton-marcu
clean
clear
clercq
cleveland
cleveland-cli
clevit
client
cliff
climat
climb
clinton
close
closur

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

cloth
cloud
clout
clutch
cml
cnb
co
coal
coars
coast
coat
coca
cochran
coconut
cod
code
code-a-phone
coebergh
coeur
cofac
coffe
cola
cold
coleman
colland
collaps
collasp
collect
cologn
colombia
colombian
colombo
coloni
color
colorado
colson
colt
columbia
columbian
colvil
comalco
comanav
comapni
combat
combin
come
comerica
command
comment
commer
commerc
commerci
commiss
commission
commit
committe
commod
common
commonweal
th
commun
communiqu
compani
compaq
compar
compat
compens
compensatori
compet
competit
competitor
complain
complaint
complementa
ri
complet
complex
complianc
compon
composit
compound
compris
compromis
comptek
comp-u-card
comput
computerland
computrac
comput-relat
comsat
comsat-
contel
comserv
comshar
comstock
comtrex
concentr
concern
concert
concess
concord
condens
condit
condition
condominium
confer
confid
confidenti
confirm
conflict
confront
congress
congression
congressman
coniston
conjunct
conn
connect
connecticut
conneticut
connor
conoco
conrac
consensu
consent
conserv
consid
consider
consist
consol
consolid
consortium
constanza
constel
constitu
constitut
constraint
construct
consult
consum
consumm
consumpt
cont
contac
contain
contel
contempl
contend
contest
continent
continental
conting
continu
contl
contract
contractor
contran
contribut
control
conveni
converg
convers
convert
convey
conveyor
convinc
convoi
cook
cooper
co-oper
coopervis
coordin
cope
copi
copor
copper
copperbelt
corbi
core
cori
corn
cornerston
corona
corp
corpor
correct
correct-1st
cosmo
cost
costco
cost-control
cost-cut
cost-reduct
cotton
cotton-y
council
counsel
counselor
count
countermeasur
counti
countri
coup
courier
cours
court
court-approv
courtauld
cover
coverag
cow
cpc
cr
cra
craft
crandal
crane
crash
crazi
creation
credibl
credit
creditor
creek
crew
crise
crisi
critic

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

criticis
crop
crow
crowlei
crown
crucial
crucibl
crude
crush
crystal
csi
csr
ct
ct/bbl
ct/shr
ctn
ctsd
cubic
cueva
cullinet
cumul
curb
cure
curent
currei
currenc
current
current-accou
curtail
curtic
cush
custom
cut
cutoff
cutter
cw.to
cwt
cxr

cyanamid
cycar
cycl
d
dai
daili
dairi
dakota
dalen
dalla
damag
damascu
damm
danforth
danger
danida
daniel
danish
danworth
dar
darbi
dart
data
datacard
dataflex
date
datron
david
dayton
dayton-hudson
db
de
dead
deadlin
deadlock
deal
dealer

dean
deapart
dearden
debentur
debt
debt-burden
debthold
dec
dechamp
decid
decis
decis/capit
declar
declin
decreas
deduct
deem
deep
defenc
defend
defens
defer
defici
deficit
definit
dei
dekalb
del
delai
delawar
delawar-chart
delchamp
deleg
deliber
delic
deliv
deliveri
delm

deloitt
delor
delta
delux
demand
democraci
democrat
deni
denim
denounc
dens-pac
denver
depart
departur
depend
deploi
deposit
depot
depreci
depress
dept
deputi
deriv
desert
desig
design
desk
despit
desptp
destin
destock
destrehan
destroi
destroy
destruct
detail
detect
deter

deterior
determin
deterr
detroit
deutsch
devalu
develcon
develop
devic
dexter
dfm
dhabi
di
diagnost
dialogu
diamant
diamond
diana
diaz
diceon
differ
differenti
digest
digiorgio
diht
dilemma
dilig
dillard
dillon
dilut
diminish
dinar
dinkov
dinner
diplomaci
diplomat
direct
director

dirham
dirti
disallow
disappoint
disast
disc
discharg
disclos
disclosur
disc-o-mat
discontinu
discontinuedo
discount
discov
discoveri
discrep
discuss
diseas
disillusion
disloc
dismantl
dismiss
disnei
disobedi
displai
dispos
disposit
disput
disrupt
dissens
distanc
distil
distilleri
distort
distribut
distributor
district
distrit

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

disturb
disut
div
divers
diversif
diversifi
divest
divestitur
divid
dividend
divis
dlr
dlr/bbl
dlr/share
dlr/shr
dlr/tonn
dlre
dlrsl
dmc
docker
dockwork
dodg
doe
dollar
dom
dome
domest
domin
dominick
dominion
domm
domtar
donald
donaldson
donegan
door
doorstep
dosher
dot
dotronix
doubl
doubl-digit
doubt
dougla
dow
downward
drain
dramat
drastic
draw
dresser
drew
dreyfu
drill
drink
drive
drop
drought
drug
drummond
dry
dsm
dtuch/shell
du
dual
dublin
due
duffour
duggan
duisburg
duli-form
dumez
duplic
durakon
durapip
durham
duri
durum
dutch
dutch/shell
duti
duti-free
dwt
dynam
e.f.
e12c
e15b
earli-develop
earn
earthquak
eas
easi
east
eastek
eastern
eastman
eastov
eaton
ec
ec-12
ecgd
eckenfeld
ecolab
econom
econometr
economi
economist
ecopetrol
ecu
ecu/tonn
ecuador
ecuadorean
ecusta
eddi
edelman
edmonton/swan
edouard
edson
eduard
edward
eec
eep
effect
effici
effort
egypt
egyptian
ei
eia
eishiro
elabor
elder
elect
electr
electr/ga
electrohom
electrolysi
electrolyt
electromagnet
electron
electro-optic
electrospac
element
elev
elimin
elizabethtown
elmira
elmo
eln
embargo
emcor
emerg
emhart
emi
emir
emma
employ
employe
emulex
enabl
encompass
encourag
end
endang
end-april
end-april/ear
end-decemb
end-mai
end-march
endors
endow
end-year
energa
energi
energi/u.s.
enforc
engag
engin
england
enhanc
eni
enjoin
enorm
enrol
enter
enterpris
entertain
entex
enthusiasm
entiti
entitl
entourag
entr
entranc
entregrowth
entrench
entri
environ
environment
environmment
epa
equibank
equimark
equip
equiti
equiti-accoun
equiticorp
equival
era
erc
eric
erickson
ericsson
erupt
escal
escap
escort
esop
espinosa
esso
est
estat
estim
etl
eugen
eurobond

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

eurodollar
euromarket
europ
european
european/medi
evacu
evalu
evas
eventu
evergreen
evid
exacerb
exagger
exce
exceed
excess
exchang
exchequ
exclud
exclus
ex-dividend
execut
exercis
exist
expand
expans
expd-e
exp-e
expect
expedi
expedit
expel
expenditur
expens
experi
expert
expir
explain
explan
explicit
explor
export
exposur
express
extend
extens
extern
extinguish
extract
extraordianri
extraordinari
exxon
f.a.
f.s.b.
f.w.
faber
fabk.o
fabric
face
face-off
facet
facil
facilit
facilti
fact
faction
factor
factori
fail
failur
fair
fairchild
fairer
fairfield
fairview
fakkan
fall
famili
fan
fao
far
fargo
farm
farmer
farmer-held
farmland
fasten
fastest-grow
fattah
favor
favour
faygo
fe
fear
feasibl
feb
februari
fed
feder
feder-mogul
fee
fee-base
feed
feedgrain
feet
felix
fend
ferguson
ferrara
ferri
ferruzzi
fertilis
fhlb
fhlbb
fiber-optic
fibr
fidata
fidel
fidelcor
fiduciari
field
fielder
fifth
fight
fighter
figur
fiji
fild
file
filipino
final
finalis
financ
financi
finanziaria
find
findlei
finl
finland
fintech
fio
fip
fire
firm
firmer
first
first-quarter
fiscal
fish
fit
fitwat
fitzwat
fix
fixman
fixtur
fla
flag
flagler
flat
flax
fleet
flexibl
float
flood
floor
florida
flotat
flotilla
flour
flow
fluctuat
fluid
fluidiz
fluorocarbon
fly
fob
follow
followin
fondiaria
food
foodgrain
foodservic
footwear
forai
forb
forc
ford
forecast
foreign
foreign-flag
foremost
foresaw
forest
forestal
form
forma
formal
format
former
formerli
formica
formid
formula
forrest
forstmann
forsyth
fort
forthcom
fortran
fortun
forward
foster
fouad
founder
foundri
four
fourth
fraction
framework
franc
francisco
francx
frank
frankfurt
franklin
fraud
fraudal

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

fred
free
freedom
freer
freez
freight
freight-on-bo
french
frenzi
fresh
fri
friction
frigat
fromartz
front-end
frontier
frost
frozen
fructos
fsb
fslic
ftc
ft-se
fuel
fujisaki
fujitsu
fulfil
full
full-court
fuller
full-fledg
fulli
full-scale
fumiko
fund
fundament
fungicid
funktion
furnitur
futur
fww
g-5
g-7
ga
gabelli
ga-burn
gaf
ga-fire
gain
galaxi
galileo
gallah
galveston
game
ga-n-all
gander
gang
ga-oil
gaon
gap
garag
garcin
garfinckel
garn
garnac
gase
gasolin
gatewai
gather
gatt
gaug
gaughan
gauntlet
gcc
ge
gear
gec
geelong
gelco
gem
gemina
gencorp
gener
general
genet
genoa
genoves
geodyn
geoffrei
geolog
georg
geraldez
gerhard
german
germani
gestur
get
ghana
gholamreza
giant
gibraltar
gift
gillett
giordano
giorgio
giovanni
give
given
glass
glatfelt
glaxo
glen
glendal
glenf
glenn
glimps
glitch
global
globe
glori
glynw
gmbh
gna
gnp
go
gobert
go-it-alon
gold
gold-bear
golden
goldendal
goldman
good
goodman
goodrich
goodwil
goodyear
gordon
goria
gotco
gov
govern
government
governor
govett
gowland
graan
grab
grade
grain
grain/oilse
granari
grand
grant
grapevin
graphic
gravel
graviti
great
greater
greec
greek
green
greenman
greenspan
greenvil
greenwich
greet
greg
grenfel
greyhound
greyhound-tra
grime
grip
grisanti
gro
gross
grossli
ground
groundnutse
group
grove
grow
growmark
growth
growth-orient
gsp
gte
guarante
guaranti
guard
guardian
guid
guidanc
guilder
guin
gulf
gullfak
gustaaf
gwinnett
gyrat
h
h.f.
h.k.
hackensack
hadson
hail
hainan
hal
halcyon
hale
half
half-stake
hall
halmi
halt
hamlei
hamper
hand
handl
handler
hanna
hanov
hanson
happi
haq
harcourt

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hard	heinz	histor	hovnanian	ihlwan	index
harder	helen	hit	how	ill	india
hardi	helicopt	hk	howel	illeg	indian
hardwar	helmut	hmo	hr	illinoi	indiana
harm	help	hoar	hrw	imag	indic
harman	hemispher	hoax	hss	imbal	indirect
harmon	henan	hoechst	hudson	imf	individu
harnischfeg	henderson	hoffman	huge	immedi	indonesia
hartford	henlei	hog	hugh	immediatlei	indonesian
hartmarx	henri	hohenwald	humphrei	immin	industri
harvei	hepworth	hold	huntington	impact	industri-prod
harvest	herb	holder	huron	implic	inelectra
haskin	herbicidesand	holidai	hurt	implicit	inflat
hatfield	heritag	holleb	husk	import	inflationari
haut	hermiston	holli	hutmann	impos	inflow
havoc	hernan	holling	hutton	imposit	influenc
hawaiian	hernandez	home	hybrid	imposs	influenti
hawkey	heron	homebuy	hydro	improperli	inform
hawkin	heublein	homemad	hydrocarbon	improv	infrastructur
hawlei	heusen	homestead	hydrodealkyl	impun	infring
hbo	heyman	hondura	hyper-inflat	imput	infus
head	high	honeywel	hytek	inaccur	ing
headquart	higher	hong	i.u.	inadequ	ingersol-rand
headwai	highest	honor	iae	inc	inher
health	highland	hope	ian	incent	iniezion
healthi	high-level	horizon	ibc	inch	initi
healthsouth	highli	hormon	iberica	inchon/india	initit
hear	highlight	hormuz	ibm	includ	inject
heat	high-qualiti	horsham	ic	inclus	injunct
heatlh	high-technolo	hospit	icahn	incom	injuri
heavi	hike	hostil	icc	incorpor	ink
heavier	hill	hot	icn	incorrect	inmobiliaria
hebei	hillsdown	hotel	idc	increas	inoc
hectar	hing	hous	identifi	increment	inou
heed	hint	housewar	ideolog	incur	inquiri
hei	hiram	houston	ifr	ind	insecticid
heineken	hirschberg	hover	igl	indefinit	insert
heinold	hisanobu	hovi	igon	independ	insid

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

insist	invis	jaim	jose	kelli	koaq-fm
inspect	invit	jake	joseph	kellogg	kodak
inspiron	involv	jamail	journalist	kelsei-hay	koehler
instabl	iowa	jame	jovanovich	kennamet	koether
instal	ipco	jamesport	juan	kenner	kohl
institut	iran	jan	judg	kenneth	kondo
instrument	iranian	janeiro	judgment	kent	kong
insult	iran-iraq	jannock	judi	kentucki	koninklijk
insur	iraq	januari	judson	kernel	korea
int	iraqi	janunari/marc	juic	kerr	korean
intact	irna	japan	jukebox	keswick	korea--usda
integr	iron	japan/india	juli	keycorp	kotc
intellicorp	irrig	japanes	jump	keyston	kraft
intelstat	irv	jardin	june	khafji	krohn
intend	irwin	jenrett	jungl	khamenei	krutikhin
intent	isl	jeopardi	junior	kharrazi	kuna
interbank	island	jeopardis	justic	khor	kunio
intercar	isomer	jeremi	justif	kidder	kuroda
intercep	israel	jerri	jute	kiena	kuwait
interest	issu	jersei	jwt	kiichi	kuwaiti
interfirst	issuer	jet	kahan	kill	kwacha
interim	isuzu	jewelleri	kai	killington	kwik
interior	it	jewelri	kain	kilowatt/hour	l.p.
intermitt	itali	jiangsu	kaiser	kim	la
intern	italian	jig	kalbag	kimbark	lab
international	italiana	jim	kamal	kind	label
interst	item	job	karl	king	laborarori
inter-tel	itoh	jobless	karlsruh	kingdom	laboratori
interuterin	iv	johannesburg	kato	kinnear	labour
interven	ixl	john	kaufman	kio	lacklustr
intervent	j.p.	johnson	kaunda	kirk	lacled
interview	j.w.	johnston	kawakami	kleinwort	ladd
intiti	jackaman	johnstown	keep	klm	ladi
intl	jackson	join	keg	kluwer	lag
inventori	jackup	joint	kei	km	lago
invest	jacob	jonathan	keidanren	kmw	lai
investig	jacor	jone	keith	knudsen	laidlaw
investor	jacquelin	jordan	kellei	koa-am	lake

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

lakehead	lb	levi	live	low-incom	mahbubul
lambert	lbo	lewi	livestock	lp	mai
lamson	ldc	li	livingston	lpg	mail
lanc	ldmf.o	liabil	lloyd	lsi	mailer
land	lead	liber	lo	luctvaart	mainfram
landa	leader	liberalis	load	ludlum	mainichi
landown	leagu	liber-minist	loan	lufkin	mainten
landshir	learn	liberti	loan-loss	lukman	maiz
langlei	leas	licenc	local	lump-sum	major
languag	leasehold	licens	lock	luncheon	make
lanka	leather	lichter	loctit	lusinchi	maker
lap-top	leav	lieberman	logist	luxembourg	maki
larosier	lebanes	liemandt	loma	luxuri	maladi
larri	leblanc	life	lomak	lvi	malais
larsson	led	lifo	lombard	lynch	malawi
laser	lee	lift	london	lyng	malayan
lasmo	leftist	light	longer	lynn	malaysia
lastest	legal	lighter	longitud	lyon	malpractic
last-minut	legg	light-weight	long-stand	m.a.	man
late	legionario	like-date	long-term	m.d.c.	manag
latest	legisl	likelihood	lonrho	m-1	mandarin
latin	lehman	lilli	look	maanag	mandat
latta	leigh	limit	loos	maatschappij	mandatori
lauderdal	leighti	lincoln	loral	macandrew	maneuv
launch	leisur	linda	lord	macao	manhattan
laurentii	lend	lindberg	lorri	macgregor	manila
lautoka	lender	line	lose	machin	mankowski
lavan	lennar	link	loss	machineri	manner
law	leonard	linse	lost	mackenzi	manufactur
lawmak	leroux	lipstick	losx	maclain	mar
lawrenc	lessen	liquefi	loui	maclean	marai
lawrencevil	letter	liquid	louisiana	macluan	marathon
lawson	lettershop	liquifi	louisiana-pac	madeira	marcad
lawsuit	leung	lire	louisvil	madison	march
lawyer	leutwil	lisa	lousiana	madrid	margin
laydai	leuzzi	list	louvr	mae	marika
layoff	level	litig	low	magellan	marin
lazo	leverag	littl	lower	magic	marjori

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

mark
market
marketplac
marlei
marlin
marri
marseil
marsh
marshal
mart
martha
martim
martin
maru
maryland
mascot
mason
massiv
master
materi
materiel
matheson
matter
matur
maxicar
maxim
maximis
maximum
maxwel
mayaguez
mayban
mayfair
mca
mcc
mccarthi
mcchip
mccune
mcdougal
mcente
mcgin
mcginlei
mcgraw-hill
mcgroarti
mcguin
mckesson
mclain
mclennan
meadow
meanwhil
measur
meat
mechan
medchem
medco
media
medic
medicin
mediterranean
meditrust
medium
medtec
mee
meet
meetng
melbourn
melbourn-base
mellon
melvil
member
membership
memori
mepc
mercer
merchandis
merchant
merg
merger
meriden
merieux
merit
merril
mesa
met
metal
met-coil
metex
method
metr
metro
metropolitan
mexican
mexico
meyer
mich
michael
michigan
micro
microchip
microdyn
micron
micronet
microsoft
microsystem
microwav
mid-contin
middai
middl
mideast
midg
midland
midlant
midnight
midpoint
mid-siez
midwest
migrat
miguel
mij
mikhail
milan-base
mildew
mile
militari
milk
mill
mill-basi
miller
mina
mine
miner
minesweep
mini
minim
minimum
minist
ministri
minneapoli
minneopoli
minnesota
minnetonka
minntech
minor
mir-hossein
mirror
miscellan
mislead
missil
mission
mississippi
missouri
mistak
misunderstand
mite
miti
mitsui
miyazawa
mlln
mln
mlnv
mmc
mo
mobil
moder
modern
modestli
modifi
mohler
moistur
molturacion
mondai
monei
monetari
monier
moniterm
monitor
monolith
monopol
monsanto
monsoon
montagn
montari
montauk
montedison
montgomeri
month
monthli
moon
moor
morgan
morger
mori
morita
morn
morocco
morrison
mortgag
mosar
moscow
motech
motor
mould
mount
mountain
mousavi
mouth
move
movi-theater
mozambiqu
mpb
mpt
msr
mtech
mth
mthly
mull
multibank
multifood
multin
multi-union
municip
munim
murfreesboro
murphi
murrai
musicland
mutchi

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

mutual
myer
myrtl
n.c.
n.sea
n.v.
n.y.
n.z.
nabisco
nacion
nadi
naif
nakason
nalco
name
name-a
napa
napach
napco
napl
narrow
nasdaq
nashvil
nat
natali
nation
national
nationalist
nationwid
natl
natn
natsuo
natur
naval
navi
navig
nazmi
nbc
ncnb
ncr
nearbi
nearer
near-record
nebraska
necessari
necessarili
nederlanden
need
needl
negoti
negotiaiton
neiman-marcu
nekoosa
nerci
nerco
net
netback
netherland
neti
nettleton
network
neutral
nevertheless
new
newbuild
newel
newhal
newli
newli-form
newli-nomin
new-mexico-ba
newmont
neworld
newport
newslett
newspap
nfa
nicol
nicopoli
nicosia
nidera
nigel
nightmar
nikko
nikonov
nil
nin
nine
nine-recur
nippon
nl
nois
nomin
nomine
non-accru
non-accrual
non-bank
nonbelliger
nonbind
non-bind
non-complianc
non-copper
non-delhi
non-ec
nonetheless
non-hear
non-interest-
non-iraqi
non-negoti
non-opec
non-oper
nonpartisan
nonprofit
non-profit
nonrecur
non-recur
non-spanish
non-tradit
nonutil
non-wirelin
norcro
norg
norio
normal
norquest
nor-quest
norsk
nortek
north
northair
northeast
northern
northtown
northview
northwest
northwestern
norwai
norwegian
note
notion
nova
npak
npo
ns/dn
nsw
nuclear
nui
nuisanc
nuton
nvhome
nycor
nyhamm
nynex
o/e
oak
oakit
oakland
oaklei
oat
object
oblig
obligatori
obtain
obvious
occas
occidental
occur
ocean
oceania
oct
octob
odet
oecd
offend
offer
offic
offici
offier
offload
offset
offshor
ofth
ohio
ohs
oil
oil/ga
oilfi
oilfield
oilfil
oil-produc
oilsand
oilse
okada
oklahoma
okutsu
oldham
oman
omani
omc
omit
omni
omnicar
on
oncor
on-dai
ondustri
oneok
on-for-11
on-for-50
ongo
on-hundreth
onic
onset
onshor
ontario
on-third
on-time
on-wai
onward
onyx
op
opeat
opec
open
open-market

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

oper
opinion
opium
opportun
oppos
opposit
opr
optimist
option
or
orang
orang-co
orang-nassau
orbit
order
orderli
ordinari
ordnanc
oreffic
oregon
organ
organis
origin
orinoko
orion
orlando
orlean
oroleum
orthoxylen
osaka
osborn
oseberg
oslo
osp
oth
oth4r
otto
ouachita
oullivan
ounc
outer
outflow
outgo
outlet
outlin
outlook
out-of-state
outpac
output
outright
outstand
oven
overal
overcharg
overcom
overdu
overfund
overhaul
overland
overnight
overproduct
overreact
oversea
overshoot
oversubscrib
over-the-week
ow
owen
owen-illinoi
owl
owner
ownership
p.h.
pa
pabx
pacif
pacificorp
pack
packag
packer
pact
pai
paid
pain
painewebb
pak
pakistan
pakistani
palai
palito
pall
panamanian
pancho
panel
panic
papandr
paper
par
para
parachut
parcel
parent
paretti
pari
parish
pariti
park
parker
parliament
parliamentari
part
parti
particip
partisan
partli
partli-own
partner
partnership
pass
passeng
passiv
passport
pasta
patent
patienc
patient
patricia
patrick
patrol
pattamapong
pattern
patti
paue
paul
paxar
payment
payout
pc
pct
pdt
pdvsa
peabodi
peak
peat
peavei
pellet
pemberton
penalti
penc
pend
penetr
pengo
peninsula
pennsylvania
pennwalt
pennzoil
pensacola
pension
pentagon
peopl
perelman
perfidi
perform
period
periood
peripher
perman
permiss
permit
perpich
perri
per-share
persist
person
personel
personnel
persuas
pertamina
perugina
peseta
peso
pest
pet
petcord
peter
petit
petro-canada
petrochem
petrochemci
petrochem-rel
petrolan
petroleum
petroli
pfister
pharmaceut
pharmaci
pharmacontrol
phase
phelp
phh
philippin
philippin-pap
phillip
phillip-van
philmac
phoenix
phone
photon
photron
phylli
physician
piccadilli
pictur
piedmont
pierr
pigmeat
pill
pillar
pillsburi
pilot
pine
pineappl
pinella
pioneer
piosec

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

pipe
pipelin
pipework
piri
pisgah
pit
pittsburgh
pittston
pl
pl-480
place
placement
plai
plan
plane
plant
plantat
planter
plastic
plasti-line
plate
plateau
platform
plaza
plc
pldt
pleas
pleasanton
pledg
plenti
plough
plu
plumb
plume
plung
ply
pm
pneumat
poehl
point
point/burling
point/ibm
pois
poison
poland
polic
policeman
polici
polici-maker
polimeri
polish
polit
politician
pollard
pollin
pollut
polydex
polym
polyphenyl
ponc
pont
pool
pool-of-inter
poor
porex
pork
port
portfolio
portion
portland
posit
possibl
post
postpon
post-split
post-war
potenti
potlatch
poultri
pound
powder
powderi
power
powerpl
practic
pratt
prebbl
precari
precipit
precis
predict
prefer
preferenti
preliminari
premadasa
prematur
premier
premium
pre-notificat
prepar
prepay
presenc
preserv
presid
presidenti
press
pressur
preston
pre-summit
pretax
pre-tax
prev
prevail
prevent
preventedf
previosuli
previous
previous-anno
price
primari
primarili
prime
primerica
princ
princip
principl
print
printer
printronix
prior
prioriti
prior-period
privat
privat-held
privatis
privat-own
pro
probabl
probe
problem
proce
proced
procedur
proceed
process
procter
procur
produc
product
profession
professor
profit
program
programm
progress
project
promin
promis
promissori
promot
prompt
proofer
propan
propel
proper
properti
proport
propos
prorat
prospect
prospectu
prosper
protect
protein
protest
prove
proven
provid
provinci
provis
provision
provoc
proxi
proxmir
prudenc
prudenti
prupos
psycholog
psyllium
pty
public
publicli
publicli-held
publish
pueblo
puerto
pull
pullout
pump
punit
punta
purchas
purol
purpos
pursu
pursuant
pursuit
push
put
qatar
qet
qtly
qtr
qtrly
quadrangl
quak
qualifi
qualiti
quantiti
quaotiti
quarter
quarterli
quartz
quebec
queensland
queensland/ja
queri

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

queue
quick
quick-fix
quiet
quietli
quinci
quinn
quot
quota
ra
racin
racket
radar
radevormwald
radio
radnik
radwel
raffineri
rafsanjani
rai
raid
railroad
railwai
rain
rainfal
rainier
rais
rakhsh
raleigh
ralli
rand
randi
random
rang
rank
rape
rapese
rapid
rapitech
rare
rashadat
rate
ratifi
ratio
rationalis
ravenna
raw
raybesto
raycomm
raytech
rbi
rca
reach
react
reaction
reactiv
reader
readi
reaffirm
reagan
real
realiz
realtech
realti
realtor
reason
rebat
rebuild
recapit
reced
receiv
recept
recess
reciev
recind
reclassif
recogn
recommend
reconsid
reconstruct
reconven
record
record-set
recov
recoveri
red
redeem
redekop
redemt
redeploi
redland
reduc
reduct
reebok
reed
re-establish
re-export
refer
refin
refineri
reflag
reflat
reflect
reform
reform-hydrot
refrain
refund
reg
regain
regal-beloit
regan
regard
regardless
regenc
regim
region
regist
registr
regret
regul
regular
regularli
regulatori
rehabilit
reichhold
reid
reimpos
reinforc
reinstat
reinvest
reiter
reject
rel
relat
relaunch
relax
releas
relet
reliacar
relianc
relief
reloc
reluct
remain
remaind
remark
re-market
remedi
remitt
remot
remov
renam
renault
renegoti
renew
rent
renta
rent-a-car
rental
reoprt
reord
reorgan
repaid
repay
repeat
repeatedli
replac
repo
report
reportedli
reposit
repres
represent
repris
repubblica
republ
republican
republicbank
repurchas
request
requir
requri
resalat
rescind
resdel
research
research/deve
research-cott
reserv
reserv-ad
reshadat
resid
residenti
residu
resign
resili
resist
resolut
resolv
resort
resourc
respect
respond
respons
rest
restat
restaur
restock
restrain
restraint
restrict
restructur
result
retail
retain
retali
retaliatori
retir
retreat
retribut
retroact
reuni
reuter
rev
revalu
revamp
revco

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

reveal
reveng
revenu
revers
review
revis
reviv
revlon
revolutionari
revolv
rexham
rhine
rhm
rhone-poulenc
rican
rice
rice/maiz
richard
richfield
richmond
rico
ricoh
ridicul
rie
rig
rigg
right
rinehart
rio
riot
rise
risen
risk
rite
rival
river
riyal

rj
rjr
rko
roach
road
roast
rober
robert
robertson
robin
robl
robson
rochest
rocket
rocki
rockport
rodriguez
roebuck
roger
roi
role
roll
roll-over
rolm
roma
roma-brisban
romero
romo
ronald
room
rooter
rorer
rori
rose
rosenkranz
rostam
roth
roth-

american
rothman
roto-rooter
rotterdam
round
rourk
rout
routin
row
rowan
rowena
royal
royc
royex
rp
r-pa
rpc
rpt
rsl
rte
rtz
rubber
rubbermaid
ruder
rudi
rule
rumor
rumour
run
rupe
rupert
rush
russel
ryan
ryan-mcfarlan
ryder
rye

s.african
s.atlant
sabah
sach
sack
safco
safe
safeti
safewai
sagawa
sai
said
sail
sait
saito
salaam
salari
sale
sall
salli
salomon
salt
sama
samjen
samuel
san
san/bar
sanchez
sanction
sandalwood
sandvik
sandwich
santa
santi
santo
sara
saskatchewan
sassan

satellit
satisfactori
satisfi
satoshi
sauc
saudi
saunder
saunderson
save
save/rocki
saver
savoi
sayst
scale
scandinavia
scatter
scenario
schaik
schedul
scheme
schere-plough
schlesing
schlumberg
schmidt
schneider
schoenheim
schreyer
schroeder
schuler
schulman
schuster
schwartz
sci
scicom
scienc
scientif
sci-med

scope
scott
scottish
scrap
scrutini
sea
seagat
seal
sear
season
seat
seattl
sec
second
secondari
second-hand
secret
secretari
secretariat
sector
secur
securitiesd
sedco
sedgwick
see
seed
seek
seen
seger
segment
seizur
sejerstad
selbi
select
self
self-defens
self-suffici
self-tender

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

sell	sfr	shortag	singer	societ	southmark
selloff	shabwa	short-circuit	singl	societi	southold
sell-off	shad	short-cover	singl-famili	societi/save	southward
semi-annual	shaffer	shortfal	sir	soft	southwest
semicon	shaki	short-term	sissonvil	soften	south-west
semiconductor	shamrock	shot	site	softwar	sovereignti
semi-finish	shandong	show	sith	soi	soviet
seminar	shanghai	showboat	situat	sokorri	soya
semi-process	shanxi	shr	six	solang	soybean
sen	share	shultz	six-dai	solar	spa
senat	share-for-sha	shumwai	ski	sold	space
send	sharehold	shun	s-k-i	solici	spain
seneg	shareholkd	shut	skill	solicit	spanish
senegales	sharp	shutdown	sky-high	solidar	spark
senior	sharpli	sichuan	skyrocket	solitron	speak
sensit	shawmut	side	slash	solomon	speaker
sentiment	shearson	sidetrack	slate	solut	special
seoul	sheer	sieg	slide	solv	specialist
separ	sheet	sierra	slifer	somehow	specialti
sept	sheikh	sight	slight	somerset	spectr
septebm	sheldahl	sign	slope	son	spectra-physi
septemb	sheldon	signal	slow	soni	spectrum
sequest	shelf	signet	slowdown	sonntag	specul
seram	shell	signfic	sluggish	soon	speech
seri	shelter	signifi	slump	sooth	speed
serious	sherman	signific	small	sorg	speedboat
serv	shi	significantli	small-boat	sorghum	speedier
servic	shield	sign-up	smelt	sort	spell
seslowski	shift	silent	smelter	sosnoff	spencer
session	ship	silicon	smith	soufr	spend
set	shipbuild	silkworm	smithklin	sound	spent
settl	shipment	silver	smuggl	sour	sperri
settlement	shipmet	sime	snake	sourc	spie
seven	shipshewana	simmon	snap-on	south	spiegler
seven-dai	shipyard	simon	snow	southam	spin
sever	shop	simsa	soar	southeast	spirit
sevic	shore	sinclair	social	southern	spirt
sew	short	singapor	sociedad	southland	spite

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

split
spokesman
spokeswoman
sport
sportschannel
sportswear
spot
spotlight
spot-next
spread
spree
spring
spring/dn
sprinkel
sprint
spun
spur
spy
squar
squeez
squibb
squir
sri
ssmc
st
stabil
stabilis
stabl
staci
staff
stage
stagnat
stai
stake
stalei
stall
stanadyn
stand
standard
stapl
star
start
startup
state
state-flag
stategi
statement
state-own
state-run
statfjord
station
statist
statistician
statoil
statpip
statu
statut
stauffer
stave
stck
steadi
steadier
steak
stearn
steel
steer
ste-geneviev
stem
stephen
step-saver
sterivet
sterl
steven
stevenson
stewart
stg
stifel
stiff
stiffen
stig
stimul
stimulu
stipul
stock
stock/brown
stockbrok
stock-for-sto
stockhold
stoltenberg
stop
storag
store
stowag
strackbein
strain
strait
strata
stratabit
strateg
strategi
strathcona
stratu
street
street"
strength
strengthen
stress
stretch
strike
stringfellow
stronger
strongest
strongli
struck
structur
struggl
stubbl
stuck
student
studi
studio
stutz
subcommitte
subordin
subroto
subscrib
subsid
subsidari
subsidi
subsidiari
subsoil
substanti
substitut
suburb
suburban
succe
success
successor
sudden
suez
suffer
suffici
sugar
suharto
suit
suitor
sullivan
sullom
sum
sumita
sumitomo
summer
summerfallow
summit
sun
sundstrand
sunflow
sunflowerse
sunken
sunris
suntrust
super
superfund
superior
supermarket
superpow
superstor
supervisori
supplementari
suppli
suppli/demand
supplier
suppoort
support
suppport
suprem
surcharg
surfac
surg
surplu
surplus
surpris
survei
surveil
surviv
susbidiari
suspect
suspend
suspens
susquehanna
sustain
sustain"
swamp
swap
swedish
sweep
sweet
swell
swift
swiss
switzerland
sybron
sydnei
symbol
symtron
synar
syndic
syracus
syscom
syscon
system
t.e.a.m.
tacit
tactic
taft
tag
tailspin
taiwan
take
takeov
take-up
talk
talman
tamba
tamper
tamura

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

tan
tandem
tandi
tangibl
tanker
tanura
tanzania
tape
taper
target
tariff
tarriff
task
tasman
tass
tata
taunton
tax
tax-free
taxpay
t-bill
tbn
t-bond
tea
team
tech
technic
technoexport
technolog
tecsyn
tehran
tektronix
telco
telco
telecharg
telecom
telecommun
tele-commun
teleconcept
telecraft
telefon
telegraph
telemarket
telephon
teleproduct
telesi
televis
telex
telkom
tell
temper
temperatur
templ-inland
temporari
tender
tenera
tenn
tennaco
tennant
tennesse
tenni
tension
terenc
term
termiflex
termin
terr
territori
territorri
terror
test
tetsuo
texa
texa/okla
texa-base
texaco
text
textil
thai
thailand
thema
theodor
therapi
thermal
third
third-quarter
thoma
thomaston
thomson
thorn
thorough
thou
threat
threaten
three-dai
threshold
thrift
thrifti
thrust
thudershow
thundershow
thwart
ticker
tie/commun
tigera
tight
tighten
tighter
tiller
time
tin
tinto-zinc
tit-for-tat
tl-f
tlx
tn
tnt
tobacco
tobi
todai
todd
toher
toi
toiletri
tokyo
toledo
toluen
tomorrow
ton
toni
tonight
tonn
tonnag
took
tool
top
tornado
toronto
torotel
toshiba
total
touch
tough
tour
town
township
toyobo
trace
tract
tractor
trade
trader
tradevest
trade-weight
trafalgar
trailwai
train
tran
tranch
transact
transamerica
transcanada
transfer
transit
transport
transworld
tranzon
trautman
travaux
travel
travenol
tre
treadmil
treasuri
treati
treatment
trend
trial
trigger
trigon
trim
triniti
triton
troi
troop
tropic
trotter
troubl
truck
true
trump
truscott
trust
trustcorp
trw
try
ttl
ttl-f
tube
tucker
tug
tumaco
tunisia
turco
turgut
turk
turkei
turkish
turmoil
turn
turnaround
turnov
tuskaloosa
tussl
tvi
twa
twenti-foot
twice
twin
twister
two-dai
two-for-five
two-for-on
two-sentenc
two-tier
type
u
u.k.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

u.n.
u.s
u.s.
u.s./soviet
u.s.-iran
u.s.-japan
u.s.senat
u.s.-west
uae
ual
uccel
uchida
ula
ulcer
ullman
ulric
unabl
unambigu
unbil
uncertainti
unchang
uncollect
und
undercut
undergo
underground
underli
underscor
undersecretar
understand
undertak
underwai
underwrit
undevelop
undisclos
unemploi
unemploy
unfair
unfairli
unfavor
unfavour
unfinish
unforeseen
uni
unicorp
unidentifi
unifi
unilev
uni-mart
unimedia
union
unisi
unit
unithold
univers
unjust
unknown
unlik
unlist
unload
unmil
unnam
unnam-b
unnecessari
unoc
unpreced
unprepar
unprofit
unpublish
unqtd
unregist
unrest
unschedul
unsecur
unsolicit
unspecifi
unstabl
unsuccess
unusu
unveil
unwelcom
unwil
uopn
up
upbeat
uphold
upper
upsurg
uptrend
upward
uranium
urban
urg
urgenc
urgent
uruguai
usa
usag
usair
usda
usdaprj-g
user
usg/morocco
usg/taiwan
usp
uspci
ussr
ust
usual
usx
utah
util
utiliaz
utilicorp
vaalian
vagu
valdivieso
valero
valhal
vallei
valor
valu
valuat
valv
vamand
van
vanasek
vancouv
varieit
varieti
vaughan
vbi
vc
veget
vehicl
vein
vend
venezuela
venezuelan
venezuelan-ow
venic
ventur
vermillion
vernitron
vero
versi
veslefrikk
vessel
viacom
vic
vice
vice-chairman
vice-presid
vice-versa
victoria
video
vieill
vienna
vigil
vigor
vik
vilat
villespedu
vintner
violat
virginia
virul
vismara
visual
vital
voe
voic
volatil
volcker
volkska
volkswagen
volta
volum
vorman
vornado
vote
vow
vulcan
w.germani
wage
wait
walker
walker-
goodma
wall
wallac
walli
wal-mart
walter
waltham
wang
want
war
warburg
ward
warehous
warfront
warn
warner
warrant
warship
wash
washington
washington-ba
washita
wast
watch
watchdog
water
waterwai
watt
wave
weak
weaker
wean
weapon
wear
weather
webb
wedd

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

wedgeston	westwego	wilson	woodhead-faul	wtc	yoshiko
wednesdai	westworld	wimpei		wti	younger
week	weybridg	win	woodstock	wyse	youth
weekdai	whale	wind	woodward	xerox	yr
weekend	wharton	winder	woolowrth	xicor	yugoslav
weekli	whatev	wine	woolworth	xidex	yugoslavia
weigh	whatsoev	winlei	word	xylen	zair
weight	wheat	winn	worgroc	yaacobi	zambia
weil	wheatlei	winston	work	yaccobi	zambian
weinberg	whelan	winston-salem	worker	yale	zealand
welbac	whim	winter	workforc	yard	zenex
welbilt	whirlpool	wireless	world	yarn	zen-noh
welcom	white	wirelin	worldscal	yasuhiro	zero
weld	wholesal	wisconsin	worldwid	yate	zhang
well	wholli-own	wisconsin-bas	worri	year	zinc
well-head	wi	wisdom	worsen	year-ago	zloti
welsh	widen	wit	wort	year-end	zoet
wesbanco	wider	withdraw	worth	yearlei	zone
west	widespread	withdrawn	wpp	yellow	zoran
westamerica	wilhelm	withdrew	wright	yellowknif	zurich
westburn	wilk-barr	withheld	wrightson	yemen	
westburneshar	willa	withhold	write	yen	
western	willamett	witter	writedown	yesterdai	
westin	willard	wmc	write-down	yeutter	
westinghous	willi	wood	writeoff	yield	
westminst	william	woodhead	write-off	york	
westpac	willing		written	york-base	
			wrongdo		

ข.3 คำสำคัญที่ได้จากชุดข้อมูล K-dataset

คำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ของชุดข้อมูล K-dataset ชุดที่ดีที่สุด มีจำนวน 9,614 คำ ดังต่อไปนี้

aafpr	abba	aberlich	abort	absenc	abstin
aama	abbei	abil	abound	absent	absurd
aaron	abbott	abl	abroad	absinth	abus
aback	abc	abnorm	abronsiu	absolut	ac
abandon	abdelatif	aboard	absat	absorb	academi

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

acceler
accept
access
accessori
accid
acclaim
accompani
accord
accordion
account
accredit
accumben
accumul
accuraci
accus
acer
ach
achiev
achingli
acid
ackland
acknowled
acl
aclu
acoem
acquir
acquisit
acr
act
action
action-ty
activ
activis
actor
actor-dan
actor-dir
actor-tur
actress
actress/s
acuiti
ad
ada
adam
adapt
add
addam
addario
addict
adept
adequ
adher
adhes
adieu
adio
adjani
adjust
adminis
administr
admir
admiss
admit
admitt
adnyamath
ado
adolesc
adolf
adopt
adopte
ador
adparam
adrian
adriana
adriat
adspac
ad-spend
adult
adulter
advanc
advantag
adventur
advertis
advic
advis
advisori
advoc
advocaci
aerosmith
afar
affair
affect
affer
affidavit
affili
affirm
affleck
affluent
afford
afi
afraid
africa
african
african-a
aftermath
afternoon
aftershoc
after-tax
afterward
ag
agath
ag-base
agenc
agenda
agent
agerton
aggress
agn
agora
agre
agreement
ag-relat
agricultu
ahead
ahearn
ahi
ahmanson
ahn
aho
aid
aim
aime
air
air-bag
aircraft
airig
airlin
airport
airwai
aisl
al
ala
alabama
alabaman
alamo
alan
alani
alarm
alaska
albert
alberta
alberto
albi
alborta
albright
album
album-roc
alc
alcohol
alcohol-i
alec
alejandro
alendron
alert
alex
alexand
alexandra
alford
alfr
alhambra
ali
alic
alien
alison
a-list
aliv
all
allan
all-busi
alleg
allegedli
allegra
allei
allen
allend
allerg
allergen
allergi
allergi-i
all-fcc-f
alli
allianc
allison
all-java
allman
alloc
allow
all-sport
all-star
all-time
aloft
alomar
alon
alongsid
alpha
alreadi
alreadi-s
alright
altar
alter
altern
altern-un
altima
alto
altogeth
altri
aluminium
alzheim
ama
amanda
amantadin
ambassado
ambuehl
amc
amd
amelio
amend
america
american
american-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ameritrad	angela	antihyper	appleg	arlington	ashlei
amezcua	anger	anti-infl	applic	arm	ashoub
ami	angi	antioxid	appoint	armageddo	ashraf
amicu	angina	anti-pira	appointe	armand	asia
amid	angioplas	antiqu	appreci	armando	asian
amidst	angl	anti-raci	appropri	armani	asia-paci
amino	anglo	antiretro	approv	armbrust	asid
amlf	angola	anti-semi	apron	armi	ask
amman	angotti	anti-smok	apt	armistead	asleep
ammar	angri	anti-trus	aqua	armistic	aspidastr
amongst	anim	antitrust	arabia	armori	aspir
amort	animalist	anti-war	arabian	armrest	assassin
amount	anita	antoni	arab-ira	arni	assault
amp	anixt	antonin	arafat-ne	arnold	assembl
amphetami	ann	antonio	arata	aros	assenmach
ampl	anna	antwerp	arbitron	arquett	assess
amuck	annabella	anwar	arch	arrai	asset
amus	annal	anxieti	archer	arraign	assign
ana	annan	anymor	archerd	arrang	assist
anaheim	annaud	anytim	archeri	arrest	assn
analog	anni	aol	architect	arriv	associ
analysi	annihil	aorta	archiv	arsenio	assort
analyst	anniversa	aortic	ard	art	ast
analyz	announc	apach	ardent	arteri	asthma
anand	annual	apart	arduou	arteri-cl	astin
anatomi	anonym	aphrodisi	area	arteriosc	astrolog
anchor	anoth	api	arena	arthous	astronaut
ancient	answer	apiec	aretz	arthriti	asylum
anderl	anthem	apo-4	argentin	arthrotom	athen
anderson	anthoni	apolog	argentina	arthur	atheroscl
andi	antibiot	apostl	argu	articl	athlet
andr	antibodi	appar	ari	artist	atkin
andrea	anti-canc	apparel	arista	artistri	atlant
andrew	anticip	appeal	aristocra	arvid	atlanta
anem	antidepre	appear	aristotl	ash	atmospher
anesthesi	anti-disc	appel	arizona	ashburi	atop
aneurysm	anti-ging	appet	arkansa	ashford	atp
angel	anti-gun	appl	arledg	ashle	atrium

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

att	autobiogr	babysitt	ballet	barefoot	batteri
attach	automak	bac	balli	barenboim	battl
attack	automobil	bachelor	balloon-t	barf	battlegro
attain	av	back	ballot	bargain	bausch
attempt	ava	backboard	ballroom	barksdal	bavarian
attend	avail	backdrop	ballyhoo	barlow	baylor
attende	avalanch	backend	baltimor	barn	bayou
attent	avalon-go	backer	balzac	barnett	baywatch
attest	avant-gar	backlash	ban	baroqu	bbc
attila	avarici	backlog	banal	barr	bdzv
attitud	averag	backpack	banc	barrag	beach
attornei	avert	backstag	bancorp	barrel	beacon
attract	avian	backward	band	barrett	beam
aua	avian-hum	backyard	bandag	barri	bean
auction	aviat	bacon	bandera-a	barrier	bear
audienc	avid	bacqu	bandwagon	barron	beaslei
audio	avium	bacteri	bang	barroom	beasti
audionet	avoid	bacteria	bangemann	bart	beat
audit	avon	bad	banglades	barton	beatl
audrei	aw	badli	bank	base	beatric
aufschnai	awai	bafta-la	banker	basebal	beauti
aug	await	bag	bankgesel	base-load	beck
auggust	awak	baggag	bankruptc	baseman	becker
august	awar	bagman-ma	banner	bash	beckinsal
augustin	award	bai	banquet	basi	becom
augusto	award-win	bail	bantam	basic	bed
auletta	awash	bailei	baotian	basin	bedelia
aunt	awfulli	baio	bar	basketbal	bedsheet
auspic	awhil	bairstow	barasch	basqu	bedsid
aussi	awkward	baker	barbado	bass	bee
austin	awri	balanc	barbara	bassett	beeban
australia	axl	baldwin	barbersho	bat	beer
austria	ayn	balfour	barbi	batch	beer-love
austrian	aznavour	balkan	barbra	bate	beforehan
author	azoff	ball	barcelona	bateman	begh
authorita	babi	ballad	bardem	bath	begun
autist	babi-boom	ballard	bardot	batman	behav
auto	babyfac	ballerman	bare	batter	behavior

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

behemoth	bergman	bigger-th	biweekli	bloodi	bold
behind	berklei	biggest	bi-weekli	bloodiest	bolivia
beij	berkoff	biggest-s	biz	blood-lik	bolivian
bein	berlin	bigotri	bizarr	bloomer	bollenbac
beisbol	bernhard	big-scree	bjoux	bloomingd	bolling
bela	bernhardt	bike	black	blow	bolshevik
belaru	berni	bill	black-and	blue	bolster
belgian	bernstein	billboard	blackhawk	blue;	bolt
belgium	berrien	billi	blackston	blue-chip	bomb
belief	bert	billion	blackstre	bluesman	bombshel
believ	bertelsma	billionai	blackwel	bluff	bon
bell	berth	bin	blain	blunt	bond
bellami	bess	bind	blair	bluntli	bone
belli-but	best	bing	blake	blvd	bonfiglio
belliger	best-paid	binki	blame	bmg	bonham
bello	bestsel	bio-chemi	blanco	board	bonn
bellow	best-sell	bioethici	blasphem	boardwalk	bonni
belong	bet	biograph	blasphemi	boast	bono
belov	beta	biographi	blast	boat	bonu
below	betaresea	biolog	blatant	boateng	bonus
belt	bethesda	biologi	blaxploit	bob	boogi
beltwai	betrai	biondi	blaze	bobbi	book
belushi	better	biopic	blazer	bobcat	booker
bemoan	beverag	biopsi	bledso	bochco	bookmark
ben	beverli	bipartisa	bleed	bodi	booksel
benazzi	bevi	bird	blei	bodili	bookshop
bendetson	bevin	birdbath	bless	bodi-pier	bookstor
benefit	bewar	birdi	blethyn	bodmer	book-styl
benetton	bibliophi	birmingha	blige	bodnar	boom
bengali	bicycl	birt	blight	bodyguard	boomer
benitez	bid	birth	blind	boe	boost
benjamin	bidder	birth-cou	blitz	boer	boot
bennett	bifida	birthdai	bln	boffo	booth
bent	big	bishop	block	bogdanovi	bootleg
berardi	bigami	bit	blockag	bogota	bop/warne
bereng	big-budge	bite	blocker	boi	border
beretta	bigelow	bitter	blond	boil	bordick
berg	bigger	bitterswe	blood	boi-toi	bordin

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

bore
bori
born
born-agai
borowieck
borrow
bosnia
boss
boston
bother
bottl
bottl-gre
bottom
bottum
bought
boulevard
bounc
boundari
bout
bow
bowel
bowi
bowl
box
boyaca
boycott
boyfriend
boyz
bp
brace
brad
bradi
bradlei
bradshaw
brai
braill
brain
bran
brand
brandi
brando
brandon
bransford
branson
brass
brave
bravo
brawl
brazil
brca
brca1
breach
bread
break
breakdanc
breakfast
break-in
breakthro
breakup
breast
breastfe
breastfee
breastmil
breath
breathtak
breiter
brendan
brendon
brennan
brenneman
brent
brentwood
brett
brew
bre-x
brian
bride
bridget
brief
briefli
bright
brightest
brightman
brilliant
brim
bring
brinklei
brio
brisban
bristl
bristow
britain
britannia
britian
british
british-b
briton
bro
broad
broadcast
broaden
broadwai
brochur
brockman
broderick
broke
brokedown
broken
broker
brokerag
bromontan
bronchiti
bronchosc
bronco
bronfman
bronson
bronx
bronx-bor
bronz
brook
brooklyn
brother
brother-i
brown
browner
brows
browser
bruce
bruguera
brunner
brush
brutal
bryant
bskyb
bu
bubbl
bucchino
buck
buckingha
bud
budapest
buddhist
buddi
budget
buena
buff
buffalo
bug
bui
buick
build
buildup
built
built-in
bulgaria
bulimia
bulk
bull
bullock
bunch
bundi
bundl
bunni
bureau
bureaucra
burgeon
burger
burgess
burgundi
burial
burk
burn
burner
burnett
burnham
burnham-d
burrow
burst
burstyn
burt
burton
buse
bush
busi
busin
butt
butterfie
butt-head
buyback
buyer
buyout
buzz
bypass
by-pass
byrd
byrn
byron
bystand
c16
c45
ca
ca2
caa
cabann
cabbi
cabin
cabl
cableac
cablecomm
cablemedi
cabler
cabl-rate
cabl-tv
cac
cadiff
cafe
caffein-f
cage
cain
cairo
calcium
calcutta
calendar
calgari
calhoun
calib
calif
californi

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

calista
call
calla
callaghan
callahan
call-in
calm
calor
calori
calvin
camaro
came
camera
cameraman
camerawor
cameron
camilla
camillett
cammarata
camp
campaign
campbel
campion
camri
can
canada
canada/ra
canadian
canadian-
canal
canalsate
cancel
cancer
cancer-pr
candac
candi
candid
candl
candlelig
cane
cann
canopi
cantalupo
canton
cantona
canuck
canva
canvas
cap
capabl
capac
capaldi
capatch
cape
capeman
capit
capitol
capshaw
captain
captiv
captur
car
carbon
carcinoge
card
cardiac
cardin
cardiolog
care
career
caregiv
carei
caribbean
carl
carla
carlo
carlole
carlstad
carl-uw
carmedged
carmel
carmello
carmen
carniv
carol
carolco
carolin
carolina
carolyn
caron
carpal
carradin
carri
carrier
carrol
carrot-to
carsalad
carson
cart
carter
carticel
cartilag
cartland
cartman
cartoon
cartridg
cartwrigh
carver
carvil
case
casei
cash
cashless
casilla
casino
casket
casper
cassandra
cassavet
cassett
cassidi
cast
castl
castleber
casualti
cat
catalog
catalogu
catalunya
cataract
catcher
catherin
cathet
cathi
cathol
cathryn
catti
catwalk
caughei
caus
caution
cavern
caviti-in
cazalb
cb
cd
cdc
cdc25
cdma
cd-produc
cd-rom
cedar
ceil
cel
celeb
celebr
celebr-st
celesti
celgen
celia
celin
cell
cellar
cellular
cemeteri
censor
censorshi
censu
cent
cent-a-sh
center
centerpie
centimet
centocor
central
centralia
centuri
ceo
ceram
cereal
ceremoni
ceridwen
certif
certifi
cervic
cesarean
cesk
cessat
cessna
cf
cfc
cha-cha
chaimowit
chain
chairman
chalk
challeng
chamber
chamberla
champ
champaign
champion
champions
champlain
chan
chanc
chanei
chang
chango
channel
chao
chapt
charact
character
charg
chariti
charl
charli
charliz
charlott
charlotte
charlton
charm
charmer
chart
charter
chase
chat

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

chattanoo	chines	cigar	class	clout	co-host
chawla	chip	cigarett	classic	clover	coin
che	chirac	cincinnat	classif	clown	coincid
cheap	chiyoda	cindi	classmat	clozapin	coki
cheaper	chkl	cinema	classroom	clt	cola
check	chloe	cinemathe	claudia	clt/ufa	colantoni
checkerbo	chloroxyl	cinematog	claus	club	cold
checklist	cho	cineplex	clean	clubland	cole
checkout	choic	cineplex-	cleaner	clue	coli
cheek	choir	cinergi	clear	clueless	colicki
cheesi	cholera	cingeri	clerk	cme	colin
chef	cholester	cink	cleveland	cnet	coliti
chemic	chom	circadian	clever	cnn	collabor
chemother	choos	circl	clich	co	collaps
cher	chop	circu	client	coach	colleagu
chernin	chord	circuit	clifford	coalit	collect
cherri	choreogra	circuitri	climact	coars	collector
chess	choru	circul	climat	coast	colleg
chest	chose	circumcis	climax	coastal	collin
chester	chri	circumfer	climb	coastlin	collis
cheung	chriqui	cite	clinch	coat	colombia
chevali	christ	citi	clinic	co-author	colombian
chiapa	christian	citibank	clinician	cobb	color
chicago	christina	citicorp	clinton	coburn	colorado
chide	christma	citizen	clip	coca	colorect
chief	christo	citizensh	clip-on	coca-cola	columbia
child	christoph	citric	clock	cocain	columbia/
childbirt	chromosom	citytv	clohessi	cockroach	columbu
childhood	chronic	civil	clone	cocktail	column
children	chronicl	clai	cloonei	code	column/ja
chile	chronothe	claim	close	coetzer	column/rb
chilean	chrysler	clair	close-cir	co-exec	columnist
chill	chubut	clamp	close-nig	coffe	colvin
chilli	chuck	clan	closer	coffin	coma
chilton	chunk	clapton	closet	co-founde	comb
chin	church	clarenc	clot	cognit	combat
china	churchil	clark	cloth	cognoscen	combin
chinabyt	cider	clash	cloud	coher	combo

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

combust	compel	concuss	consolid	cool	costner
comcast	compens	condemn	consort	cooper	costum
come	compensat	condit	consortiu	coordin	cotsako
comedi	compet	condol	conspir	cop	couch
comedian	competit	condom	conspirac	cope	cough
comedi-dr	competito	conductor	constant	copeland	coulter
comedi-va	compil	cone	constin	copperfie	coulthard
comfort	complain	confer	constitut	coppola	council
comic	complaint	confess	constrain	co-produc	councillo
command	complemen	confid	construct	copul	councilma
commandeu	complet	confident	consult	copyright	councilwo
commemor	complex	configur	consum	core	counsel
commend	complexio	confin	contact	corest	count
comment	complianc	confirm	contain	corman	counterca
commentar	compliant	confirmat	contamin	corn	counterfe
commer	complic	conflict	contempl	cornbelt	counterpa
commerc	complimen	conform	contempor	corneal	counterpr
commerc-e	compon	confound	contend	cornelia	counti
commerci	compos	confront	contest	corner	countri
commiss	composit	confus	context	coronari	countrysi
commissio	compound	congest	contin	corp	coup
commit	compress	conglomer	continent	corpor	coupl
committe	compris	congress	conting	corps	courag
commod	compromis	congressi	continu	correct	courier
commodor	compusa	congressm	contracep	correspon	cours
common	compuserv	conjur	contract	corridor	court
commun	comput	conn	contracto	corrupt	courtesi
communist	comput-cl	connect	contraind	cort	courthous
compact	computer	connectic	contrast	corteg	courtnei
compani	comput-ge	conquer	contribut	cortex	cousteau
companion	comrad	conrad	control	cosbi	couturi
companywi	conceal	conscienc	controver	cosgrov	cover
compaq	concentr	conscious	convei	cosmet	coverag
compar	concept	consecut	convent	cosmonaut	covert
compart	concern	consensu	convict	cost	cow
compass	concert	conserv	convinc	co-star	cowan
compassio	concess	consid	convuls	costello	cowboi
compat	concret	consist	cook	costli	co-writer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

co-writte
co-wrote
cox
cozi
cpr
crack
crackdown
cracker
craft
craig
crane
cranki
crash
crave
crawford
crazi
cream
creat
creatacar
creation
creativ
creator
creatur
credenti
credibl
credit
creditor
creep
cremer
crepe
creutzfel
crew
cri
cricket
crime
crimin
crisi
crisp
criteria
critic
croasdal
croatia
crohn
cronkit
crooner
crop
cross
crosson
crossov
cross-pro
crow
crowd
crown
crtc
crucial
crude
crudup
cruel
cruelti
cruis
crummiest
cruz
cso
csx
ctd
cub
cuba
cuban
cubic
cuddl
cue
cuff
culmin
cult
cultur
culver
cum
cumul
cunanan
cup
curb
cure
currenc
curri
curriculu
currier
curs
curti
curto
curvebal
cusack
cushion
custodi
custom
cut
cutback
cute
cut-throa
cuyahoga
cw
cyber
cyberawar
cybercaf
cyberchat
cyber-rig
cyberspac
cyberstar
cycl
cyclist
cyst
cystic
czech
d.c.
da
dabb
dad
daddi
dade
daffi
dafo
dai
daida
dai-glo
daik
daili
daisi
dalai
dale
dall
dalla
dalla-for
dallowai
dalton
damag
dame
damien
damn
damon
damu
dan
danc
dancer
dandenaul
dane
danger
dangl
dani
daniel
danish
danni
danson
danza
darden
dare
dario
dark
darker
darkli
darl
darlen
darn
darnedest
darren
darryl
dashiel
data
databas
date
datelin
daughter
daulton
dave
davi
david
davidson
daylight
daypart
daytim
daytona
dead
deadli
deadlin
deadlock
deal
dealer
dealershi
dealt
dean
death
debat
debbi
debit
deborah
debrah
debt
debut
debutant
dec.-feb.
decai
decarlo
decenc
dechernei
decis
declar
declin
decod
decor
decre
decreas
decrimin
dedic
deep
deeper
deepli
defam
defeat
defend
defens
defensele
defensema
deferiet
defi
defiantli
defici
deficit
definit
deform
dega

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

degener	dentsu	detector	diari	director-	disneyfi
deguzman	denver	deterior	diarrhea	directori	disord
dehydr	denzel	determin	diaz	direct-to	dispar
dehydroep	depardieu	detour	dicaprio	directtv	dispatch
del	depart	detractor	dice	directv	displai
delai	departur	detriment	dick	dirk	dispos
delawar	depend	detroit	dicker	disabl	disput
delfino	depict	deutsch	dickinson	disadvant	disqualif
deliber	deplor	deutschem	die	disagr	disrupt
delight	deploy	devan	diego-bas	disagre	diss
deliv	deposit	devast	diekman	disappear	dissatisf
deliveri	depositar	develop	dielectr	disappoin	dissemin
dell	depot	devic	diesel	disast	dissent
della	depreci	devil	diet	disc	distanc
delroi	depress	deviou	dietari	discern	distant
delta	depth	devito	dietitian	discharg	distast
demand	deputi	devlin	differ	disciplin	distinct
dement	derek	dewei	difficult	disclaim	distraugh
dementia	derid	dexter	digest	disclos	distress
demi	dermatolo	dextrou	digit	disclosur	distrib
demo	dern	dfw	dilat	disco	distribut
democraci	derrick	dhaka	dill	discord	district
democrat	descend	dharma	dillon	discount	disturb
demograph	descenza	dhea	dimens	discov	ditch
demot	desecr	dhl	diminish	discoveri	ditzi
dengu	desert	di	dimond-ca	discrep	diuret
deni	deserv	diabet	dimwit	discrimin	diva
denis	design	di-affect	dinard	discuss	dive
denmark	desir	diagnos	dinner	disdain	divers
denni	desk	diagnosi	dinosaur	diseas	diversifi
denot	desmond	diagnost	dion	diseas-re	divestitu
denounc	desper	dialogu	dioxid	disgorg	dividend
densher	despic	dialysi	dip	dish	divin
densiti	despit	diamet	dipego	disinfect	divis
densmor	destin	diamond	diphtheri	disk	divorc
dent	destruct	dian	diplomat	dismiss	divx
dental	detail	diana	direct	disnei	dixon
dentist	detect	diana-the	director	disnei-ow	dizzi

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

dj	donor	dracula	dry	duti	eberl
dna	donut	draelo	dry-brush	dutton	ebert
dna-damag	doom	draft	dtb	duva	e-busi
dobson	door	drafte	dtp	duval	eccentr
docket	dopamin	dragon	duan	duwan	echo
doctor	dope	drain	dub	dvd	eckhard
doctor-pa	dor	drama	duchess	d-vt	eckhart
docudrama	dorff	dramat	duck	dwek	eclips
document	dorfman	drang	dud	dwell	econom
documenta	dori	draper	dudlei	dy	economi
dodg	dormant	draw	due	dylan	economist
dodger	dorothi	drawn	duel	dynasti	ed
dodi	dorsei	drawn-out	duet	dysfunct	eddi
dof	dose	dre	dug	dzubeck	edg
doff	doubl	dread	dukaki	eager	edgar
dog	doubl-bre	dreadlock	duke	eagerli	edgi
doherti	doubl-dig	dream	dull	eagl	edinburgh
dolan	doubledai	dreamwork	duma	ear	edit
dolc	doubli	dreari	dumb	ear-bite	editor
dole	doubt	dress	dumber	earl	editori
doll	doug	drew	duncan	earli	edmond
dollar	dougla	drill	dundi	earmark	edmonton
dollmak	dour	drink	dunk	earn	edmund
dolorosa	dous	drive	dunkin	earnest	eduard
dolphin	dove	drive-in	dunkirk	earth	eduardo
dome	dow	driver	dunleavi	earthl	educ
domecq	dowd	drivewai	dunn	eas	edutain
domest	down	droll	duo	easi	edward
domin	downhil	drop	durabl	easili	effect
dominican	download	drop-off	duran	easi-to-c	efficaci
dominik	downsid	drove	durango	east	effici
dominiqu	downsiz	drown	durat	easter	effort
donahu	downtown	drowsi	durban	eastern	egg
donald	downturn	drug	durham	eastman	egg-bound
donaldson	downward	drug-and-	durrant	eastwood	egoyan
donat	doyl	drug-resi	dust	eat	egypt
donatella	dozen	drum	dustin	eateri	egyptian
donohu	dr	drummer	dutch	eben	eichelber

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

eid-el-ke	elk	empir	enni	erad	european
eigeman	ell	employ	eno-byrn	ergonom	european-
eight-coi	ella	employe	enorm	eric	ev
eighth-ra	ellen	empow	enquir	eriq	eva
eight-tim	elliott	empti	enrico	erod	evad
eija	elswit	enabl	enrol	eros	evalu
einfeld	elton	enact	ensembl	errand	evan
eisner	elucid	enceph	ensu	erstwhil	evand
ejacul	elud	encephalo	ensur	erupt	evangelis
el	elus	encircl	enter	erythromy	evelyn
elabor	elvi	encompass	enterpris	escal	even
elain	e-mail	encor	entertain	escap	even-end
elastas	e-mail/di	encount	enthusias	espn	event
elbow-to-	embark	encourag	entic	espn/abcn	eventu
elder	embarrass	encrypt	entir	espn2	ever
elderli	embassi	encyclope	entiti	espnew	everhart
eldon	emboss	end	entranc	esposito	ever-incr
elect	embrac	endear	entrench	esrd	evid
electr	embri	endeavor	entri	essenti	evok
electra	embryo	endlink	environ	estat	evolut
electrici	emerald	endometri	environme	estefan	evolution
electrolu	emerg	endotheli	envis	estel	evolv
electron	emi	endow	enzi	estim	ex
eleg	emili	enema	enzym	estonia	exact
elektra	emir	enemi	epa	estrogen	exam
element	emiss	energi	epic	estrogen-	examin
eleph	emma	enforc	epidem	eszterha	exampl
elev	emmanuel	engag	epidemiol	eta	ex-beatl
elfman	emmer	engin	epidermi	ethan	ex-boyfri
elfont	emmerich	england	episod	ethic	exceed
elicit	emmerich-	englert	epp	ethiopian	except
elig	emmi	english	epstein	ethnic	exception
elimin	emmi-win	enhanc	equal	eu	excess
elisabeth	emmycast	enid	equip	eucalyptu	exchang
elista	emot	enjoi	equiti	eugen	excit
elit	empathi	enl	equival	euro	exclud
elizabeth	emphasi	enlarg	er	euro-mp	exclus
elizondo	emphysema	enlist	era	europ	excruci

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

excurs	extend	fairwai	father	fellini	fill
excus	extens	faith	father-ha	fellow	film
exec	extent	fake	fatigu	fellow-mu	filmmak
execut	extra	fall	fatma	felon	final
executon	extraordi	falloff	fatso	femal	finalist
exercis	extraterr	fallout	fatti	feminist	final-rou
exex	extravaga	fals	fave	femm	financ
exhaust	extrem	fame	favor	fender	financi
exhibit	extremist	famili	favorit	ferguson	find
exhibitor	exual	familiar	favreau	fernandez	fine
exhum	exud	famili-or	fawcett	ferrari	finest
exil	exum	famili-ru	faxon	fertil	finger
exist	ex-white	famin	fbi	fest	fingerpri
exit	ex-wife	famou	fcc	festiv	finish
ex-marin	exxon	fan	fda	fetch	finland
exorbit	ey	fanat	fear	fete	finnish
expand	eyemark	fanci	feast	fetisov	fior
expans	eyr	fao	featur	fetu	fire
expect	ezralow	farber	featurett	fever	firelight
expedit	fab	fare	featur-le	fewer	firewal
expenditu	fabien	farentino	februari	feydeau	firework
expens	face	farewel	fece	fiasco	firm
experi	facelift	farm	fed	fiber	firm;
experienc	face-off	farmer	feder	fibrosi	first
experimen	facial	farquhar	federico	fiction	first-cla
expert	facil	farrah	fedex	fide	first-eve
expertis	fact	farrel	fedora	fidel	first-han
expir	factor	farrow	fee	fiefdom	first-loo
explan	factori	fascin	feed	field	first-qua
explicit	fade	fashion	feedback	fienn	first-rou
explor	fahrenhei	fast	feeder	fierberg	firstrun
explos	fai	faster	fee-for-s	fierc	first-run
export	fail	fastest-s	feel	figaro	first-tim
expos	failur	fast-food	feet	figgi	first-yea
exposur	fair	fast-trac	fein	fight	fiscal
express	fairfax	fat	felic	fighter	fishburn
ex-soni	fairli	fatal	felicia	figur	fisher
ex-studen	fair-us	fate	fell	file	fit

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

fitzgeral	flinder	fogerti	forev	fourth-bi	free-spee
five	fling	foggitt	foreword	fourth-hi	freez
five-fold	flint	fogl	forfeitur	fox	frei
five-game	flippant	foglia	forget	fox-liber	freight
five-hitt	float	fold	fork	fox-new	freilich
five-seco	flock	folei	form	fox-own	french
fix	flockhart	folic	formal	foxx	french-ca
fla	flood	folk	formaldeh	foyer	freno
flagship	floor	folk-blue	format	fraca	frequenc
flamboy	flop	folk-rock	former	fractur	frequent
flame	floral	follow	formerli	fragil	fresh
flamenco	florenc	follow-up	formid	fragment	freshest
flamingo	florida	fonar	formula	frai	freshman
flammabl	flourish	fond	for-profi	frame	fri
flanker	flow	fonda	forrest	fran	fridai
flap	flower	fondest	forster	franc	friedman
flash	flu	fontain	forsyth	franchis	friend
flashligh	flub-proo	fontana	fort	franci	friendshi
flat	fluctuat	foo	forthcom	francisco	fright
flatten	fluent	food	forti	frank	fring
flavor	fluf	food-born	fortif	frankenhe	front
flaw	fluff	foot	fortifi	frankenst	fruit
fled	fluid	footag	fortun	franklin	frustrat
fledgl	fluorid	footbal	forum	franz	fry
flee	fluorid-c	footstep	forward	frasier	fuddi-dud
fleet	fluorosi	footwear	fosamax	fratern	fudg
fleetwood	fluti	forai	fossil	fraud	fuel
flesh	fly	forb	foster	fraught	fujimori
fletch	fly-by-ni	forbid	fought	frawlei	fujisanke
fletcher	flyer	forbio	foul	frazer	fujitsu
flex	flynn	forc	foul-up	frazier	full
flexibl	fmf	forcibl	found	frecuenci	full-blow
flextech	fo	ford	foundat	fred	fulli
fli	foam	fordham	founder	frederick	fulwid
flick	focu	forecast	four11	free	fume
flicker	focus	foreign	four-cyli	freedom	fun
flier	foe	foreman	four-stro	freelanc	function
flight	fogel	forest	fourth	freemantl	fund

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

fundament	gambit	gem	giancarlo	glaxo	golin
fund-rais	gambl	gena	gianni	gleam	golota
funer	gambler	gender	giant	glenn	gomez
funerari	gambon	gene	gibson	glickman	gong
funk	game	gener	gibson/ju	glimmer	gonzal
funki	game-ty	genesi	gifford	glitch	gonzalez
funkifi	gandolfin	genet	gift	global	good
funni	gang	geneticis	gig	globalsta	goodson
funniest	gangster	geneva	gigabyt	globe	goodwil
fur	gao	geni	gil	glock	gooral
furi	gap	geniu	gilbert	glori	gore
furiou	garag	genius	gile	gloria	gorecki
furnish	garcia	genom	gillo	glove	gorg
furnitur	garcia-st	genr	gimmick	glover	gori
furyk	garden	gentrifi	gina	glow	gospel
futil	gari	genuin	gingerbre	gm	goss
futur	garner	geoff	ginni	go	gossip
futurist	garrett	geograph	ginsburg	go-ahead	gossip-mo
ga	garri	geographi	gipsi	goal	got
gabanna	garson	georg	girl	god	gotham
gabbana	garth	georgetow	girl-dete	godless	gottlieb
gabelli	gartner	georgia	girlfrien	godzilla	gould
gabrel	gase	georgia-b	gisel	godzilla-	goursaud
gabriel-i	gasolin	georgian	gittin	goe	gov
gaddi	ga-tanker	geraldo	giuliani	goeth	gove
gadget	gate	gerard	give	gogh	govern
gaff	gatewai	german	giveawai	goizueta	governmen
ga-haul	gather	germani	glad-hand	gold	governor
gai	gattaca	germann	gladli	goldberg	govern-ow
gai-in-mi	gauntlet	gernot	glamor	golden	gown
gail	gazett	gershwin	glamour	goldeney	gra
gain	ge	gerstman	glanc	goldenvoi	grab
gala	gear	gestur	gland	goldman	grace
gallbladd	gedd	get	glaser	goldmin	grade
gallego	gee	ghost	glasgow	goldsmith	grader
galleri	geek	ghostwrit	glasier	goldthwai	graduat
galthi	geiberg	ghoulish	glass	goldwyn	graf
gam	geldof	giacomo	glavin	golf	graft

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

graham	gregori	guilt	halo	hardcor	haugland
grai	grei	guilti	halt	hardcov	haul
grainger	grief	guitar	halv	hardlin	haunt
gramerci	grier	guitarist	ham	hard-line	haven
grammi	griffin	gulf	hama	hard-live	havoc
grammi-wi	griffith	gum	hamburg	hardship	hawk
granada	grissom	gumbel	hamburg-r	hardtop	hawthorn
grand	gritti	gumbo	hamdan	harkin	hay
grandchil	grocer	gummo	hamil	harlei	hazard
grandmoth	groceri	gumption	hamilton	harm	haze
grandson	groin	gun	hamlet	harmless	hbo
grant	groom	gurvitz	hammer	harmoni	hbv
granulocy	groov	gusher	hammond	harmonica	hdl
graphic	gross	guthridg	hampton	harnel	hdl-2b
grappelli	grosser	guttenber	hamsher	harold	head
grappl	ground	gutter	han	harper	headach
grate	group	guttman	hana-bi	harpercol	headhunt
grave	grove	gwyneth	hand	harperpap	headlin
gravedigg	grow	gymnast	handgun	harpist	headquart
graveyard	grown-up	gynecolog	hand-held	harpo	headset
graviti	growth	habit	handi	harrel	health
great	growth;	hacker	handl	harrelson	healthi
greater	grune	hackman	handov	harrer	healthier
greco	gt	haft	handphon	harri	heap
greec	gte	haggerti	handsom	harrison	hear
greed	guangdong	ha-ha	hand-writ	harrow	hearst
greek	guarante	haight	handyman	harsh	heart
green	guard	hail	hang	harsher	heartbrea
greenberg	guccion	haim	hangout	hart	hearti
greenest	guerrilla	hain	hank	hartenbau	heat
greenfiel	guest	hair	hanselman	hartford	heat/colu
greenhous	guevara	hairdress	hansi	hartman	heather
greenligh	guid	hal	hanson	harvard	heavi
greenspan	guidanc	half	hanson-he	harvei	heavier
greenwich	guidelin	hall	happi	hasek	heavili
greet	guilak	halloween	harangu	hat	heavyweig
greg	guild	hallucin	harass	hatch	hebert
gregg	guilder	halmi	hard	hate	hech

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hecht	hepburn	highest	hirsch	homeless	horvitz
heckler	herald	highest-p	hirschfel	home-made	hose
hector	herb	highest-r	hirschhor	homer	hospic
hedaya	hercul	high-fly	hirst	homevid	hospit
hedg	hereford	high-freq	hispan	homicid	hossein
hedonist	herk	highli	histor	homo	host
heel	herman	highlight	histori	homosexu	hostag
heft	hernandez	high-octa	historian	honcho	hostil
hefti	hero	high-prof	hit	honda	hot
hegedu	heroin	high-qual	hitchcock	honest	hotard
height	heroism	high-scho	hitler	honeyboi	hotel
heighten	herp	high-spee	hiv	honeymoon	hoteli
heineken	herpesvir	high-stak	hiv-posit	hong	hotlin
heinrich	hershei	high-tech	hk97	honor	hotmail
heiress	herv	hightman	hmo	honorari	hot-pant
held	hesit	higuera	ho	honore	hot-sell
helen	heston	hijack	hoax	hood	hottest
helena	heterosex	hike	hobbi	hoodlum	hound
helicopt	hethering	hilar	hoch	hoof	hour
hell	hewitt	hill	hockei	hook	hourli
helm	hewlett-p	hillari	hoffman	hooker	hourlong
helmer	heydai	hiller	hofmann	hookworm	hous
helmet	heyward	hillerman	ho-hum	hoop	household
helmut	hh/cdc	hilligoss	hoist	hoover	housem
help	hhv-6	hilltop	holbrook	hop	hous-sena
helpl	hib	hilt	hold	hope	houston
helsinki	hick	hilton	holder	hopkin	hove
hemispher	hickei	himalaya	holdov	hopper	howard
hemophilu	hidden	hincklei	hole	horac	howel
hen	hideki	hine	holidai	horib	howi
henchman	hieroglyp	hing	hollywood	hormon	hpi
henderson	high	hingi	holobyt	horn	hpv
hendrick	highbrow	hint	holocaust	horner	hr
hendrix	high-calo	hip	holyfield	horribl	hrt
henri	higher	hip-hop	homag	horrifi	htcc
hensela	higher;	hippocamp	home	horror	hub
heparin	higher-le	hipster	homegrown	hors	hudson
hepat	higher-pr	hire	homeland	horton	huei

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

hug	hyler	iii	inactiv	ineffect	ink
hugh	hype	ill	inadequ	ineffectu	ink-stain
hui	hyperact	illeg	inan	inertia	inmat
hull	hypersens	illicit	incarn	infam	inn
human	hyper-sin	illinoi	incent	infant	inner
human-bas	hypertens	ilo	incestu	infect	innoc
humanitar	hyperther	imag	inch	infecti	innocent
humbert	hypervita	imagin	incid	inferior	inoper
humberto	hypothes	immatur	includ	infest	inquiri
humbl	hypothesi	immedi	inclus	infin	insan
humid	hypothet	immens	incom	inflamm	insecur
humor	hypoxia	immers	incomplet	inflat	insert
humphrei	i/o	immigr	inconsist	inflow	insid
hundr	iab	immin	increas	influenc	insight
hundt	iagara	immun	incred	influenti	insignifi
hungari	ian	immunodef	increment	influenza	insist
hungarian	iata	immunolog	indec	influx	insomnia
hungaroto	ibaraki	impact	indecis	inform	inspect
hunger	ibf	impair	independ	infrastru	inspir
hungri	ibm	impart	index	infring	instal
hunki	ic	impeach	indi	infus	instanc
hunt	icahn	impel	india	ingersol-	instant
huntingto	ichakawa	impermiss	indian	ingest	institut
huntlei-b	ick	implant	indic	inglewood	instruct
hurd	icm	implaus	indict	ingmar	instructo
hurdl	icon	impli	indiffer	ingredi	instrumen
hurrican	idaho	implic	indigen	inhabit	insulin
hurt	ideal	import	indign	inher	insur
husband	ident	impos	indirect	inherit	intak
husband-a	identifi	impress	indiscree	inhous	integr
hush	ideolog	impressio	individu	in-hous	intel
hussein	idiom	imprint	individu-	inhuman	intellect
huston	iffm	imprison	indonesia	initi	intellig
hut	ifpi	improp	indoor	inject	intel-wsj
hutchison	ight	improprie	induc	injunct	intend
hve	iglesia	improv	industri	injur	intens
hybrid	ignit	improvis	industri-	injuri	intensifi
hydrocarb	ignor	inact	industryi	injuri-pr	intent

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

interacit	iowa	italian	japan	jessi	jone/will
interact	iq	itchi	japanes	jessica	jone-will
intercept	ir	item	japanes-a	jesu	jonz
intercour	ira	itn	japanes-l	jet	joplin
intercut	irabu	itv	jarnefelt	jew	jordan
interest	irani	iv	jarrett	jewel	jordanian
interfer	iranian	ivan	jason	jewish	jose
inter-gov	iraq	ivf	java	jewison	joseph
intern	iraqi	ivori	jave	jibe	joss
internet	ireland	ix	javier	jim	journal
interpers	iri	jab	jaw	jimenez	journalis
interplai	iridium	jack	jazz	jimi	journei
interpret	irish	jacket	jealousi	jimmi	journeyma
interst	irk	jacki	jean	joan	journo
interv	iron	jack-in-t	jean-bapt	joaquin	jovi
intervent	ironi	jackson	jean-clau	job	joyless
interview	irradi	jacksonvi	jean-jacq	jobe	judd
intim	irrit	jaclyn	jean-mari	jobless	judg
intimid	irv	jacob	jean-mich	jocelyn	judgment
intl	irvin	jacobi	jeann	jock	judici
intra-squ	irwin	jacor	jeanniot	jodi	judith
intraven	isaac	jacott	jeff	joe	juic
intrigu	isabella	jacqu	jefford	joel	juilliard
introduc	isford	jacquelin	jeffrei	johannesb	juli
intrud	islam	jagger	jellyvis	johansson	julia
intrus	island	jai	jen	john	juliann
intuit	ismail	jail	jeni	johnni	juliett
invad	isobel	jake	jenni	johnson	julio
invas	isol	jakob	jennif	johnston	jump
invent	isomedix	jam	jen-peter	joi	jungl
inventori	isra	jama	jeopard	join	junior
invest	israel	jamaican	jeremi	joint	junket
investig	issei	jame	jerk	joke	junki
investor	issu	jami	jerom	jollei	jupit
involv	istvan	jamiroqua	jerri	jon	juri
in-your-f	isuzu	jana	jersei	jonathan	jurras
inzerillo	it	jane	jerusalem	jonbenet	jusczyk
ioc	itali	jani	jess	jone	just

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

justic
justifi
justin
k.d.
k6-mmx
kabc-am
kacd
kacd-fm
kagan
kai
kaiser
kaleidosc
kalevi
kalmykia
kamil
kangaroo
kansa
kaphan
kaplan
karachi
karaszews
karg
karloff
karma
karolinsk
karpov
kar-wai
kasparov
kastner
kate
katharin
katherin
kathi
kathleen
kathrin
kathryn
kati
katlin
katt
katz
katzenber
kayoko
kazan
kbp
kcal
kcb
keanu
keaton
keaton/ge
keenen
keep
keeton
kehl
kei
keitel
keith
kelantan
kellei
keller
kelli
ken
kenna
kennedi
kenneth
kennett-b
kenni
kensingto
kentucki
kerri
kesei
kessman
kevin
keyboard
keyboardi
keyword
kfi-am
khan
kiarostam
kibosh
kick
kickback
kicker
kid
kidman
kidnei
kid-orien
kidron
kidvid
kiefer
kill
killer
kilmer
kilo
kilomet
kilt
kim
kimmel
kimmelman
kimura
kind
kinda
king
kingdom
kingpin
kingsmil
kinsella
kinski
kirch
kiri
kirk
kirkpatri
kirsti
kishimoto
kiss
kit
kitano
kitchen
kitti
kkbt
kkbt-fm
klass
klein
klesko
klezmer
klif
kline
klo
klo-fm
klsx-fm
klurfeld
klve-fm
knee
kneecap
knee-leng
knick
knight
knighthoo
knock
knockoff
knopf
knotti
know
knox
knuckl
kobzon
koch
kombat
kommunika
kong
konstanc
konstanti
kool-aid
kooper
koran
korea
korin
kortnei
kosevo
koss
kost-fm
kosuk
kota
kotto
kovitz
kozak
krane
kranner
kraus
krauss
krenek
kri
krieger
kristen
kristi
kristin
kristoffe
kroger
krolock
kroq/gold
kroq-fm
krth-fm
krzyzewsk
ktwv-fm
ktzn
kubrick
kuerten
kuilenbur
kuito
kumamoto
kuralt
kurdish
kurki
kurosawa
kurtz
kurtzman
kusturica
kweti
kxa-tv
kyoto
l.a.
la
lab
labarrer
label
labor
laborator
labut
lace
lachlan
laci
lack
lacroix
ladd
ladi
lafac
lahti
lai
laid
laid-back
lam
lama
lamb
lame
lamont
lamp
lampoon
lan
land

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

landfal	latvia	leagu	lennon	liar	linda
landfil	lau	leahi	lennox	libel	lindo
landi	laud	leak	lens	liberti	lindro
landmark	lauda	lean	lenz	librari	line
landown	laudadio	leann	leo	licenc	lineag
landscap	lauder	learn	leon	licens	lineback
landslid	lauderdal	leas	leonard	license	line-item
lane	lauff	leav	leonardo	lichtenst	lineman
lang	laugh	leavin	leopard	lichtman	linesman
langella	launch	lectur	leprosi	lid	lineup
languag	launder	lederman	leptin	lidstrom	link
lanka	lauper	lee	leptospir	lie	linklett
lanni	laura	leed	lesbian	lieberfar	linnei
lansburi	laureat	left	leshner	life	linu
lao	laurenc	left-hand	lesion	lifeguard	lion
lap	laurent	left-wing	leslei	life-save	lionel
lapierr	lausann	lefurgi	lesson	lifestyl	lipid
laptop	lauter	leg	less-stre	life-thre	lipstick
lar	lavinia	legaci	lest	lifetim	liqueur
larg-scal	lavishli	legal	lester	lift	liquid
lark	law	legend	let	light	lira
larri	lawmak	legendari	lethal	lighter	lisa
laru	lawrenc	leggi	lett	lightn	list
lasal	lawsuit	legion	letter	lightstor	listen
laser	lawyer	legisl	letterman	likabl	listen-to
lash	lax	legit	leukaemia	likelihoo	lit
last	layer	legitim	leukemia	like-mind	literari
last-chan	layoff	lehman	level	likewis	literatur
laszlo	layout	leicest	leverag	lila	lithgow
latch	lcd	leigh	levi	lillehamm	lithotrip
late	le	leipzig	levin	limbaugh	lithuania
late-nigh	lea	leisur	levinson	limber	litig
latest	lead	leisz	levitch	limit	littl
latex	lead-base	leland	lew	limit-run	live
latin	leader	lemk	lewi	limousin	live-acti
latina	leadershi	len	lhakpa	linag	livent
latino	lead-in	lend	li	linchpin	liver
latter	leaf	lenert	liabil	lincoln	liverpool

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

livewir	longest-r	lover	luther	magistr	malign
liz	longev	lovett	luxembour	magnat	malik
lizandro	long-miss	low	luzhkov	magnet	maling
lloyd	long-run	low-birth	ly	magnitud	mall
lo	long-stan	low-budge	lymphoma	maguir	mallett
load	long-term	low-cost	lynch	magyar	malloi
loan	longtim	lower	lynda	maharajah	malnutrit
loath	look	loweri	lyne	mahmoud	malon
lobbi	look-alik	lower-tha	lynn	mahmud	mambo
lobbyist	lookout	lowest	lyon	mahonei	mammal
local	loom	low-kei	lyonnai	mai	man
locat	loonei	lowri	lyric	maiden	manag
lock	loopi	low-speed	mac	mail	manageri
lockhart	loos	lozeng	macau	main	managua
lodg	loot	lso	macbeth	maina	manchest
lodz	lopez	luaka	macharia	mainland	mandalai
loew	loral	lucci	machin	mainli	mandarin
loftier	lorca	lucent	machineri	mainlin	mandatori
lofton	lord	luci	maci	mainstrea	mandela
logic	lori	luciano	macintosh	maintain	mandrel
logo	lose	lucil	macisaac	maj	mangan
log-roll	loser	lucki	maclean	majandra	mangold
lolita	losh	lucretia	macneil	major	mangold-h
lollo	loss	ludwig	macpherso	major-lab	manhattan
lollobrig	lost	lufkin	macrophag	major-lea	manipul
lomax	lott	lugar	mad	make	man-made
lomb	lotteri	lugosi	maddock	maker	manner
lombard	lotti	lui	made-for-	makeup	manni
lombardo	lou	lukewarm	madigan	make-up	mano
lon	loughlin	lullabi	madison	makharadz	manor
london	loui	lumber	madonna	malacarn	mansion
london-ba	louima	luminari	madrid	malaria	manson
lone	louis	lumpi	mae	malaysia	mantl
loneli	louisiana	lunch	mafia	malaysian	manton
long	louisvil	luncheon	magazin	male	manufactu
long-awai	love	lunden	magda	male-bias	manz
long-dist	loveli	lung	magdalena	male-driv	manzarek
longer	lovemast	lure	maggi	maleeva	map

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

mapl	maroza	materi	mcgovern	medicar	menthol
map-make	marque	matern	mcgrath	medicin	merc
maradona	marri	mathieson	mcgraw	medic-inf	merced
marathon	marriag	matoian	mcguir	mediocr	mercer
marbl	marriott	matraca	mci	mediterra	merchandi
marc	marshal	matrix	mckai	medium	merchant
marcelo	mart	matt	mckean	meek	mercier
march	martha	mattea	mckellan	meer	merck
marchand	marti	mattel	mckendric	meet	mercuri
marci	martial	matter	mckenna	megahertz	meredith
marcia	martin	matter-of	mckinlei	megastor	merg
marcu	martina	matthew	mckinnei	megil	merger
mardi	marv	mattiolo	mclean	meharri	meridian
mare	marvel	maud	mcleish	mel	mering
marenostr	marvin	maximum	mcmichael	melani	merit
margaret	marx	maxin	mcnamara	melanoma	merri
margin	maryland	maxwel	mcname	melbourn	merril
mari	masaya	mayor	mcnaught	melissa	mersei
maria	masayuki	mazar	mcveigh	mello	merton
marian	mascot	mazurski	mdp	mellon	meryl
mariann	masculin	mbe	mdrd	melros	mess
mariano	mashantuc	mca	meadow	melt	messag
marijuana	mask	mcafe	meagan	member	messier
marilyn	maslon	mcalari	meal	membershi	metal
marina	maso	mcartor	mealtim	membran	meteor
mario	mason	mcat	mean	memento	meteorit
marisa	mass	mcbeal	meaningle	memo	meteorolo
marit	massachus	mcbride	meanwhil	memoir	method
mark	massag	mccain	measur	memor	metro
marker	massiv	mccartnei	meat	memori	metro-gol
market	mass-mark	mccaw	mechan	men	metropoli
marketpla	master	mcclure	med	mena	mexican
market-st	mastercar	mcconaugh	medal	mend	mexico
marlei	masur	mccurri	media	mendt	meyer
marlin	mat	mcdonald	median	menopaus	mg/dai
marlon	matalin	mcdormand	mediat	menstruat	mgm
marlow	match	mcduffi	medic	mental	mi
marold	matchmak	mcginni	medicaid	mental-he	miami

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

miandad
mice
michael
micheaux
michel
michelang
michelob
michigan
michon
mick
mickei
mickelson
micro
microbiol
micron
micropros
microsoft
midcon
middl
middl-ag
middl-lev
midem
midg
midler
midseason
midst
midwest
midyear
mielli
mighti
miguel
mike
mikita
milan
mild
mile
milit
militant
militari
militari-
militia
milk
mill
millenniu
miller
milli
milligram
million
millionai
million-s
milo
milwauke
mimi
mimic
mind
mine
miner
mini
minim
miniseri
mini-seri
miniskirt
minist
ministri
minivan
mink
minneapol
minnesota
minor
minstrel
minu
minut
mir
mira
mirag
miramax
miranda
mirror
mischer
misconcep
misconduc
misdemean
misdiagno
miseri
misguid
misportra
miss
missel
missens
mission
missionar
mississip
mistak
mit
mitchel
mite
mithe
mitosi
miu
mix
miyak
miyazaki
mln
mln/dai-c
mmp
mmp-9
mob
mobi
mobil
mob-theme
mochiach
mock
model
modem
moder
modern
modern-da
modest
modestli
modif
modifi
modin
moesha
mogadishu
mogul
moham
mohammad
mohammadi
moistur
mold
molecul
molecular
molli
mom
moment
mo-mix
monarchi
monash
mondex
monei
monei-los
monei-win
monetari
monica
monitor
monk
monkei
monoclon
monopoli
monr
monro
monster
monster-m
mont
montana
montel
month
monthli
monti
monticell
montreal
montreal-
montserra
monument
mood
moon
moonlight
moor
moorer
moorhous
moorish
moral
moratoriu
morello
moreov
moretta
morgan
moriarti
mormon
morn
morn-long
morocco
morphin
morri
morrison
mortal
morton
moscow
mose
moshontz
mosquito
moss
mossad
mossberg
mother
mother-ch
motion
motiv
motor
motorcycl
motorola
motown
mottola
mount
mountain
mourn
mous
mouskouri
mouth
mouthguar
move
movement
movi
moviego
moviemak
movi-goer
mowad
mozilla
mr
mrap
mraz
mri
ms
msn
msnbc
mtm
mtm-sb
mtv

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

mubarak	musgrav	nap	nd	network	nhl
much-anti	mush	napoleon	nea	networld/	niagara
muddi	music	naqvi	near	neuman	nicaragua
mug	music-bas	narada	nearbi	neurolog	nich
mugshot	musician	narr	nearer	neurosurg	nichol
muhammad	musicvide	narrat	nearest	neurot	nichola
mui	muslim	narrow	nearli	neurotox	nicholson
mula	musselman	nasal	near-term	neurotran	nick
mulder	mussolini	nascar	nebraska	nevada	nickel
muldoon	muster	nasdaq	nec	neve	nickelode
mull	mutant	nashvil	neck	neveen	nicki
mulli	mutat	nashvil-b	necklac	neverthel	nickla
mullin	mutual	nassau	necrophil	new	nicknam
mulo	mutv	nastassja	ned	newborn	nicol
multi-app	mycobacte	nasti	nednei	newburi	nicola
multicent	myer	nat	need	newcom	nicotin
multi-cha	myra	nathan	needl	new-found	nida
multimedi	myriad	nation	needl-fre	newfoundl	niedermay
multi-mil	mysteri	nationwid	neg	newhart	nielsen
multipl	myth	nativ	neglig	newli	nietzsch-
multi-scr	mytholog	nato	negoti	newman	nifti
multi-yea	n.y.	nato-led	neighbor	new-model	nigel
multiyear	nab	nattili	neighborh	newscast	nigh
multo	nabokov	natur	neil	newsgath	night
mumbl	nacho	natwest	nell	newsmag	nightclub
mummi	nader	nausea	nelson	newsmagaz	nightmar
munich	nafta	naveen	neon	newspap	nightspot
murder	nagano	navig	neonat	newsprint	nighttim
murderess	nagi	navin	nephrolog	newsweek	nih
murdoch	nahoko	naymark	nerv	newswir	nik
murphi	nair	nazi	nervou	newsworth	nike
murrah	nakamura	nba	nestl	newton	nimitz
murrai	nam	nbc	net	next	nina
muscl	nama	nbc-monda	netcar	next-gene	nind
muscular	name	nbc-tuesd	netherlan	nextgener	nine-mile
muse	nana	nc-17	netscap	nfc	nino
museeuw	nanci	ncaa	netsurf	nfl	nippon
museum	naomi	ncr	netwar	ngawarla	nirmala

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

niro
nishi
nissan
nitrogen
nitti-gri
nix
nixon
nl
nlc
noa
noah
nobel
nobilo
nobl
nobodi
nobuhiro
no-contes
nocontest
nod
no-fly
no-hitter
noir
nois
nolt
nom
nomin
nomine
nomo
nondur
non-engli
nonethele
nonexclus
nonexist
non-ficti
non-gover
non-hodgk
non-holid
nonlinear
non-live
non-mainl
non-micro
non-music
non-perfo
non-profi
non-ring
non-road
nonsmok
non-smoke
non-stop
non-tradi
noon
normal
norman
norri
nortel
north
northam
northeast
northern
northwest
norton
norwai
norweb
norwegian
no-sale
nose
nostalg
nostalgia
notabl
notch
note
notebook
noteworth
notic
notif
notorieti
notr
notwithst
nov
nov.-jan.
nova
novel
novella
novemb
novic
novotna
now
nra
nrg
ntl
nuclear
nucleu
nude
nuditi
nui
numb
number
numer
nun
nurburgr
nurs
nutrient
nutrit
nutrition
nutti
ny/la
nyff
nymphet
nynex
nypd
o.j.
oak
oakland
oat
oaxaca
obes
object
oblig
obscen
obscur
observ
obsess
obsess-co
obst
obstacl
obstetr
occas
occasion
occident
occidenti
occipit
occup
oceanogra
oconnor
oct
octob
odd
oddbal
odonnel
oedip
offbeat
off-beat
offenc
offens
offer
offfend
offic
offici
off-kilte
off-label
off-netwo
off-road
off-spinn
offspr
offstag
ogea
ohio
ohio-base
oil
oiler
ok
oklahoma
oko
oktoberfe
old
older
oldi
old-time
oliv
olivi
olivia
olymp
olympia
omar
omnipres
on
onassi
onboard
on-board
on-camera
onc-domin
onetim
on-game
ongo
onlin
onlin.not
on-line
on-name
onoran
on-ramp
onsal
onset
onstag
on-third
on-time
onward
on-yard
open
openli
oper
opera
ophthalmo
opinion
oppenheim
oppon
opportun
oppos
opposit
oprah
opt
optic
opu
oracl
oral
orang
orbison
orchestra
order
ordin
ordinari
oregon
orenstein
organ
orgazmo
origin
oriol
orion
orlean

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

orphan
orphanag
orrin
ortt
osaka
oscar
oscar-cal
oscar-nom
oscar-win
osgood
osmo
osteoporo
osugi
oth
othello
other
ottawai
oul
oust
out
outback
outbid
outbreak
outcast
outclass
outcom
outdon
outdoor
outer
outfit
outh
outlai
outlaw
outlet
outlin
outlook
out-of-co
outpati
outperfor
outpost
outpour
output
outrag
outscor
outsid
outskirt
outsourc
outspoken
outstand
outstrip
outtak
ovat
over
over-40
over-achi
overal
overcom
overcrowd
over-diag
overdon
overdos
overdr
overh
overhaul
overnight
overpopul
over-pres
overrun
overse
oversea
oversight
overstai
overt
overtli
overtur
overturn
overweigh
overwhelm
ovitz
ovum
ow
owen
own
owner
ownership
oxid
oyster
ozwald
pace
pacif
pacino
pack
packag
packer
pact
pad
padr
page
pageant
pager
pai
paid
pain
paint
painter
pai-per-v
pair
pakistan
pal
palac
palai
palazzo
pale
palestini
palisad
palm
palma
paltrow
pam
pamela
panama
pancreat
pandem
panel
panelist
panini
panish
pant
pantheon
panther
papa
paparazzi
paper
paperback
paperboar
papp
par
par-5
par-71
par-72
parabl
parad
paradi
paradorn
parallel
paralysi
paramount
parapleg
parasit
pardon
parent
parenthes
par-five
pari
park
parker
parkinson
parlanc
parliamen
parlor
parnevik
parodi
parol
parretti
parri
parrot
parson
part
parti
partial
particip
particl
particul
partner
partnersh
partygo
pasadena
pass
passeng
passer-by
passion
passiv
passport
past
pasta
pasteur
pastor
pat
patagonia
patch
patent
paterno
path
pathogen
pathwai
patient
patricia
patrick
patriot
patron
pattern
patterson
pattiz
patton
paul
paula
pauli
paulin
pavarotti
pave
pawn
pawpaw
paxson
paxton
payback
paycheck
paydai
payer
payment
paymer
payout
paz
pb
pbx
pc
pca
pcb
pco

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

pdt	penseroso	persona	photo	pitch	pleasant
pdufa	pension	pertain	photograp	pitcher	pleasantv
peac	pentagon	peruvian	phrase	pitt	pleasur
peacekeep	penthous	pervers	physic	pittsburg	pledg
peacemak	pentium	pestel	physician	pix	plenti
peach	penultim	pet	physiolog	pizza	plight
peak	peopl	pet-born	pianist	placard	plod
peanut	peopl/arc	pete	piano	place	ploi
pearc	peoplesof	peter	piccoli	placebo	plot
pearson	pequot	petersbur	piceanc	plagu	plu
peck	perceiv	petersen-	pick	plai	plug
peddler	percent	peterson	pickart	plain	plumb
pediatr	percentil	petrikin	picker	plainfiel	plung
pediatric	percept	petroleum	picki	plaintiff	plymouth
pedophili	perdita	petti	pickl	plan	pm
peel	perez	pew	pickup	plane	pneumonia
peer	perfect	pfeiffer	pictuir	planet	pocket
peerless	perfor	pg-13	pictur	planner	pocketnet
pelvi	perform	pga	pie	plant	pod
pelvic	perhach	phantom	pierc	plant-der	podgi
pen	period	pharmaceu	pilch	plaqu	poe
penalti	periodico	pharmaci	pill	plasma	poem
penc	periosteu	phase	pilot	plastic	poet
pend	perk	phenomeno	pin	platform	poetri
pendergas	perki	phere	pinch-hit	platinum	poignant
pendulum	perkin	phil	pink	platt	point
penelop	perlich	philadelp	pinochet	plausibl	pointcast
penetr	perman	philand	pinstrip	playboi	point-of-
penguin	permiss	philip	pin-up	playdat	poir
penicilli	permit	philipp	pioneer	player	pois
penn	peroxid	philippin	pipe	playmak	poison
pennant	perplex	phillimor	pippen	playoff	poker
pennebak	perri	phillip	piraci	playwrigh	poland
pennel	perriman	philosoph	pirat	plaza	polanski
penner	per-scree	phoenix	pisark	plc	pole
penni	persecut	phone	piss	plea	polic
penningto	persian	phone-in	piston	plead	policeman
pennsylva	person	phonetap	pit	pleas	polici

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

polish	port	power-lin	prenat	primetim	produc-so
polish-bo	porter	powerpc	prepar	prime-tim	product
polish-fo	portfolio	power-pla	pre-produ	primu	professio
polit	portion	pph	prequel	princ	professor
politican	portland	pprom	preschool	princess	profici
politicia	portli	ppv	prescrib	princip	profil
poll	portrai	practic	prescript	principl	profit
polli	portrait	pragu	presenc	print	profit-sh
pollock	portug	prais	present	prion	profound
pollut	pose	prank	preserv	prior	progestin
polygram	posei	prayer	presid	prioriti	program
polygram-	posit	preacher	president	prison	program/c
polystyre	poss	pre-broad	preslei	privaci	programm
polyunsat	possesori	pre-cance	pre-sold	privat	progress
pontecorv	possess	precaut	press	privat-ru	prohibit
pontel	possibl	preced	press-ent	privileg	project
pontiac	post	precinct	pressur	prix	prolif
poof	postcard	preclud	preston	prize	prolifer
pooh	poster	precursor	pretax	prize-win	prolong
pooki	postman	predat	preterm	pro	promin
pool	post-mark	predecess	pre-treat	proactiv	promis
poor	post-op	predict	prevail	probabl	promo
poorer	postpon	predictor	preval	probat	promot
poorest	post-prod	predomina	prevent	probe	prompt
pop	postseaso	preemi	preview	problem	prone
pope	post-surg	preempt	price	probucol	prop
poppa	postur	preemptiv	price/wee	proce	propag
poppi	pot	prefer	price-fix	procedur	propagand
popul	potato	pregnanc	pricei	proceed	propan
popular	potent	pregnant	pride	proceedur	proper
pop-up	potenti	prei	priest	process	properti
porcelain	potter-ri	prejudic	primaci	processor	propon
porcelli	poultri	prelimina	primari	proclaim	proport
porgi	pound	prematur	primarili	prod	propos
porn	pour	premier	primat	prodi	proscan
pornograh	powder	premio	prime	prodigi	prosecut
pornograp	powel	premis	prime-md	produc	prosecuto
porsch	power	premium	primer	produc-cr	prosk

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

prospect	puff	que	rain	raven	receiv
prostat	pulitz	queen	rainbow	rawl	recept
prostitut	pull	queer	rais	raymond	receptacl
protect	pulp	quentin	ralli	raytheon	receptor
proteg	puma	question	ralph	rb	recess
protein	pumpkin	questionn	ramazan	rca	re-challe
protest	punch	quiet	ramei	reach	recip
protocol	punch-up	quill	rami	react	recipi
proudli	punish	quilti	ramif	reaction	recit
prove	punit	quinetet	ramirez	read	reckless
proven	pup	quinlan	ramona	reader	reckon
provid	pupil	quinn	rampl	readi	recogn
provis	purchas	quinta	ramsei	readi-to-	recognit
provoc	pure	quit	rand	reagan	recollect
prp	puri	quot	randal	real	recommend
prudenc	purport	qvc	randi	realiti	reconcil
prudent	purpos	rabin	randolph	realnetwo	reconsid
prusin	pursu	rabindran	random	realplay	reconstru
pseudonym	pursuit	race	rang	realti	reconven
psi	purveyor	rachel	ranger	real-time	record
psychedel	pusan	rachin	rank	reappear	recount
psychiatr	push	rachtman	rank/seas	rear	recoup
psychic	put	racial	rank-and-	reason	recov
psycholog	putt	racism	rap	reassert	recoveri
psychopat	pyke	rack	rape	reassur	recruit
psychosi	pyramid	radiat	rapid	rebat	recur
pten	pyrin	radio	rapidli	rebecca	recut
pub	pyrotechn	radiosurg	rapper	rebeck	recycl
public	pyscholog	rae	raptor	rebel	red
public-af	q3	raechel	rare	rebirth	red-carpe
public-in	quaid	raffl	rarefi	reborn	redempt
publicist	qualifi	rage	rashad	rebound	redesign
public-pr	qualiti	ragsdal	rat	rebuff	redford
publish	quart	ragtim	rate	rebuild	redgrav
puddl	quarter	rai	ratif	recal	red-hors
pudnei	quarterba	raid	ratio	recapit	redistric
puent	quarterfi	raider	ratner	recchi	redmond
puerto	quash	railwai	ravag	receipt	redrock

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

reduc	regularli	remot	request	resurrect	rhesu
reduck	regular-s	remov	requir	retail	rheumatoi
reduct	regulator	ren	rerun	retain	rhino
redund	rehabilit	renaiss	reschedul	retent	rhythm
redux	rehears	renal	rescu	rethink	rhythmic
redwood	reid	re-narrow	research	retir	riaa
reed	reimburs	render	resembl	retool	riana
ree-jone	reiner	rendit	reserv	retreat	ribbon
reel	reinforc	rene	reshap	retun	riccardo
reemerg	reinvigor	renegoti	reshuffl	return	ricci
rees	reiss	renew	resid	reunion	rice
reev	reissu	renna	residenti	reunit	rich
re-examin	reiter	reno	residu	reuter	richard
refer	reith	renov	resign	reuter/va	richenber
refere	reject	rent	re-sign	reveal	richer
refin	rekindl	reopen	resist	revel	richest
reflect	rel	reorgan	resolut	revenu	rick
reform	relat	rep	resolv	reverber	rickel
refriger	relations	repair	reson	revers	rickman
refug	relaunch	repeat	resort	review	rico
refund	relax	repeatedl	resourc	review/ca	rid
refurbish	releas	repel	respect	review/fi	ride
refus	reli	repetit	respirato	review/pe	rifl
regal	reliastar	replac	respond	review/st	rigbi
regard	relief	replic	respons	review/te	riggi
regardles	religi	report	rest	review/ti	right
regenc	religion	reportedl	restart	review/vi	right-han
regga	relish	repres	restaur	revis	right-thi
regi	reloc	represent	restenosi	revision	right-win
regicid	reluct	repress	restitut	revit	rigor
regim	remain	reprint	restor	reviv	rijnsburg
regimen	remak	repris	restraint	revok	riker
regina	remar	reproduct	restrict	revolt	riki
region	remark	repubblic	restructu	revolut	rilei
regist	rembrandt	republ	result	revolv	rim
regret	rememb	republica	resum	reward	rimantadi
regul	remind	repurchas	resurfac	rewrit	rime
regular	remiss	reput	resurg	reynold	rimini

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

riminicin	rochon	room	rubber	saban	san
ring	rock	roommat	rubin	sabin	sanchez
rinker	rocker	roon	rudi	sabr	sanchini
rins	rocket	roosevelt	rudnick	sach	sanction
rio	rocki	rooster	rudolf	sack	sand
riordan	rockin	root	rudolph	sacr	sander
riot	rockplex	rosario	rudraneil	sacrament	sandra
rip	rod	rose	rue	sacrifici	sandrich
ripe	roddi	rosemari	ruff	sad	sandrin
rise	rodger	rosenberg	rugbi	safe	sanguin
risk	rodham	rosenblat	rugrat	safeguard	sanofi
riski	rodkei	rosencran	rule	safer	santa
rita	rodman	rosi	ruler	safeti	santana
rittenban	rodnei	ross	rumbel	safewai	santeria
ritual	rodrigo	roster	rumbl	sag	santiago
ritz	rodriguez	rotat	rumor	saga	santini
ritz-carl	roger	roth	run	sahara	santo
riva	rogu	roug	run-in	sai	sapien
rival	roi	rough	runnerup	said	sappi
rivalri	roland	round	runner-up	sail	sarah
river	role	roundabou	rupert	saint	sarajevo
rivera	roll	rourk	rupp	saint-and	sarandon
rivet	rollin	rous	ruptur	salari	sarcomer
rivlin	rollov	rout	rural	sale	sardinia
roach	roman	router	rush	salient	sargent
road	romanc	routin	rush/high	salim	sartor
roar	romania	rout-reca	rushdi	saliva	satan
rob	romant	rove	russel	salivari	satcher
robard	romantic	row	russia	salon	satellit
robb	rome	rowland	russian	salsa	satir
robber	romeo	roxann	russo	salt	satisfact
robert	romp	royal	rustl-pro	salut	saturdai
robeson	ron	royalist	rut	sam	saudi
robin	rona	royalti	ryan	samantha	saunder
robinson	ronald	royalti-f	ryavec	samet	sausag
robust	ronan	rpm	ryder	sampl	sausag-lo
rocha	roof	r-rate	sa	sampra	sauvag
rochest	rooki	rtl	saatchi	samuel	savag

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

savard	scholarli	screenpla	second	self-cong	sequel
save	school	screenwri	second-ac	self-este	sequel/pr
saver	schoolboi	scribe	secondari	self-made	sequenc
savvi	schoolhou	scripp	second-ha	self-proc	sequin
sawamatsu	schrager	script	secondhan	self-regu	serafini
sax	schulman	scripter	second-hi	self-reli	serbian
saxophoni	schulz	scrum-hal	second-la	self-repo	sergeant
sayl	schumach	scrutini	second-mo	sell	sergio
sbano	schwartz	scudder	second-pl	selleck	seri
scald	schwarzen	sculptor	second-ye	seller	serial
scale	schwarzma	sculptur	secret	sellout	seriou
scalia	scienc	scum	secretari	selznick	serious
scam	scienc-fi	sea	secretli	semest	sero
scan	scientif	seafood	sect	semicondu	seroconve
scandal	scientist	seagal	section	semifin	serotonin
scandinav	scientolo	seagram	sector	seminar	serpent
scarc	sci-fi	seahawk	secur	semi-reta	serv
scenario	scion	seamanshi	sedan	semi-reti	server
scene	sciorra	seami	sedona	semler	servic
sceneri	sclerosi	sean	see	sen	servic-pr
scent	scold	seana	seed	senat	session
schaefer	scoop	seanc	seek	send	set
schaffer	scope	sear	seeker	sendup	setback
schatzsch	score	search	seem	sengupta	settl
schedul	scorses	searchabl	seen	senior	settlemen
scheider	scotch	search-fo	segment	sens	set-top
scheme	scotia	searchlig	segreg-er	sensat	seuss
schevardn	scotland	seascap	segu	sensibl	sever
schiff	scott	seasid	segura	sensit	sevigni
schiffer	scottish	season	seidel	sensori	sex
schiller	scrambl	season-en	seidemann	sens-perc	sexi
schizophr	scranton	season-op	seinfeld	sentenc	sexual
schlamm	scrapi	seat	seiz	seoul	seydoux
schlessin	scratch	seatbelt	seizur	separ	seymour
schlosber	scratcher	seattl	seldom	separatis	sg1
schmidt	scream	seaworthi	select	sepsi	sgt
schmitz	screen	sebac	selenium	sept	shacklefo
scholar	screen-an	sebastian	self	septemb	shadow

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

shahravan	sheer	short-att	sigarm	sister	sleepi
shake	sheet	shorter	sight	sisterhoo	sleepless
shakespea	sheffield	shortli	sign	sit	sleeveles
shakeup	sheila	short-liv	signatur	sitcom	slept
shaki	shelbi	short-ter	signific	site	slick
shakur	shelburn	short-tim	significa	situat	slide
shalhoub	sheldon	shot	silenc	six-membe	slight
shalom	shell	shoulder	silent	size	slip
sham	shelter	show	silicon	skate	slogan
shaman	shelton	showbiz	silk	skelton	slot
shampoo	shelv	showcas	silva	skew	slovenia
shanahan	shepard	show-clos	silver	skg	slow
shandong	shepherd	showdown	silverman	ski	slower
shane	sheppard	shower	similar	skid	slowli
shanghai	sherak	showtim	simmon	skill	sluggish
shannen	sheridan	shrine	simon	skin	slum
shape	sheriff	shrink	simoom	skinner	slump
shapiro	sheryl	shriver	simpleton	skinni	small
share	shi	shu	simplic	skin-to-s	small-cap
shareen	shield	shuai	simplifi	skip	smaller
sharehold	shift	shue	simpson	skirt	smallpox
shark	shindl	shuffl	simul	skull	small-tow
sharman	shine	shutout	simulcast	sky	smart
sharon	ship	shuttl	simultan	skynet	smartcard
sharp	shipment	shy	sinatra	skyscrap	smarter
sharpli	shipper	shyamalan	sincerest	slab	smash
sharp-ton	shipwreck	sib	sinclair	slack	smell
shaun	shirlei	sibeliu	sing	slacker	smelt
shave	shirt	sibl	singapor	slain	smile
shaver	shoah	sichuan	singer	slam	smilei
shaw	shochiku	sick	singer-so	slam-dunk	smith
shawn	shock	sickl	singh	slash	smithe
shawne	shocker	sid	singl	slate	smith-hem
shear	shoenberg	side	sinis	slater	smithsoni
sheba	shoot	sidelin	sink	slave	smog
sheehan	shop	sidewalk	sip	slavic	smog-ridd
sheelah	shore	sidnei	sir	slavin	smoke
sheep	shortag	siegfri	siren	sleep	smoke-fre

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

smoker	solemn	sour	specialti	spokeswom	stacei
smokin	solicit	sourc	specif	spongifor	staci
smoother	solid	south	spectacul	sponsor	stack
smoothli	solo	southeast	spectrum	sponsorsh	stadium
smuggl	soloist	southern	specul	spontan	staff
smuggler	solomon	southpaw	speculosc	spoof	staffer
snag	solti	southwest	speech	sport	stage
snake	solut	southwick	speed	sport/uti	stagnant
snap	solv	sovereign	spell	sport-ori	stagnat
snare	somali	soviet	spenc	sportscas	stahl
snider	somber	soyinka	spencer	sportswri	stai
snippet	somoza	spa	spend	sportszon	stain
snivel	son	space	spender	spot	stake
snob	sondheim	spacei	spent	spotlight	stalei-lo
snoop	song	spaciou	sperm	spotmedia	stalk
snow	songstres	spain	sperm-bor	spous	stalker
snowboard	songwrit	span	spice	sprai	stall
snub	soni	spanek	spici	sprain	stallon
soap	sonni	spaniard	spider	spread	stalwart
soapi	sonoran	spanish	spiderman	sprightli	stamford
soar	sophist	spanish-l	spielberg	spring	stamp
sobi	sophomor	spar	spike	springboa	stamped
soccer	soprano	sparc	spin	springfie	stan
social	sore	spare	spina	sprinkl	stanc
societi	sorrow	spark	spinal	spur	stand
sociologi	sort	sparkasse	spine	spurt	standard
socko	sorvino	sparkl	spinner	squabbl	standard-
sodomi	sosa	spastic	spinoff	squad	stand-in
soft	sought	spawn	spin-off	squall	standout
softer	soul	spe	spirit	squar	stand-up
soft-thro	soul-sear	speak	spirlea	squar-dan	stanford
softwar	sound	speaker	spit	squeez	stanlei
soho	soundbit	spear	splash	squirrel	staph
soil	soundchec	spearhead	split	sri	staphyloc
sold	soundmak	spec	spoke	st	stapl
soldier	sound-onl	speci	spoken	stab	stapleton
sold-out	soundscan	special	spokesman	stabil	star
sole	soundtrac	special-e	spokesper	stabl	stardom

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

stare	steepest	stole	strict	su	sullivan
stargat	steer	stolen	strident	suav	sum
stark	steffi	stoltz	strike	subcommit	summer
starr	steiger	stone	string	subcontra	summit
starrer	stella	stoop	strip	subject	sun
star-stud	stellar	stop	strip-dow	subkoff	sunda
start	stent	store	strive	sublim	sundai
startac	step	stori	stroe	sub-noteb	sundai-ni
starter	stepdaugh	stori-vis	stroke	subordin	sundanc
startl	stepfath	storm	stroll	subpoena	sunglass
startlink	stephan	stormi	strom	subscrib	sunlight
startup	stephen	storylin	strong	subscript	sun-senti
stat	stepson	stp	strong;	subsequ	sunset
state	stereotyp	straight	stronger	subsidiar	suo-helm
statement	steril	strain	strongest	subspecia	super
staten	stern	strait	strongli	substanc	super-abi
state-of-	sterngold	strand	struck	substanti	superb
state-run	stethosco	strandber	structur	substitut	superbl
statesman	steve	strang	struggl	subtl	supergrou
station	steven	strang-be	strut	subtleti	superhigh
statist	stewart	stranger	stuart	subtyp	superior
statu	stewart-t	strap	stub	suburb	superman
statuett	stieger	strateg	stubborn	suburban	supermark
statut	stiffen	strategi	stuck	succe	supermode
staunch	stifl	strategis	student	succeed	supernatu
stayer	stigma	stratocas	studi	success	superspee
std	stiller	strauss	studio	successor	superstar
steadi	stillman	straw	stuf	suck	superstor
steadili	stimul	streak	stuff	sucker	supervis
steal	sting	stream	stuffi	sue	superviso
steam	stink	streep	stumbl	suffer	supplemen
stecalov	stinki	street	stump	suffici	suppli
steckler	stint	streisand	stupid	suffolk	supplier
steeb	stir	strength	sturm	sugar	support
steel	stock	strengthe	stutter	suggest	suppos
steeler	stockard	stress	style	suicid	suppress
steen	stocker	stress-re	stylist	suit	suppresso
steep	stockwel	stretch	stylu	suitcas	suprem

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

sure-fire	swedish	syphili-i	tampa	tb	telethon
surf	sweenei	syracus	tampopo	tci	teletubbi
surfac	sweep	system	tan	tea	televis
surfer	sweepstak	systemwid	tang	teacher	telewest
surg	sweet	szabo	tanglewoo	teahous	tell
surgeon	sweethear	szell	tango	team	tellurid
surgeri	sweetwat	szulc	tank	teammat	tempera
surgic	swell	szymborsk	tanker	teamster	temperatu
surpass	swept	tab	tanna	tear	templ
surplu	swicord	tableaux	tanner	teaser	tempt
surpris	swim	tablet	tantamoun	tech	temptat
surround	swimmer	tabloid	tantrum	technic	tenant
survei	swimsuit	tackl	tao	technicia	tendenc
surveil	swindl	taco	tap	technicol	tendon
surviv	swindler	tactic	tape	techniqu	teng
survivor	swing	tactil	tapeworm	techno	tenn
susan	swiss	tag	tar	technolog	tennant
suscept	switch	tagor	tara	ted	tennesse
suspect	switzerla	tailor	tarak	teddi	tenni
suspend	swoon	taiwan	tarantino	teem	tenor
suspens	swoosi	taiwanes	targan	teen	tension
suspicion	sword	taka	target	teenag	tent
sustain	sydnei	take	tars	teen-ag	tenth
sutherlan	syke	taken	tartar	teenybopp	tenur
sutton	syllabl	takeoff	tartikoff	teeth	teresa
suzann	sylvest	takeov	taskforc	teheran	term
swaffer	symbol	takeshi	tast	telco	termin
swain	symcox	tale	tast-perc	telecast	term-pape
swallow	sympathet	talent	tattoo	telecom	terrain
swamp	symphoni	talk	taught	tele-comm	terrestri
swan	symposium	talker	taupin	telecommu	terri
swanson	symptom	talkin	tauru	teleconfe	terribl
swap	syndi	talladega	tavern	telefilm	territori
swarm	syndic	tamagotch	tax	telefood	terror
swashbuck	syndrom	tamahori	tax-free	telegraph	terrorist
sweat	synergi	tamara	taxi	telekom	test
sweathog	synergist	tame	taxman	telephon	testicl
swede	syphili	tami	taylor	telephoni	testicula

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

testifi
testimoni
testoster
tetsu
tewksburi
texa
tex-mex
text
textend
textil
textur
tfl
thai
thailand
thalidomi
thame
thanksgiv
thapsigar
tharp
thaw
the
theater
theatr
theatric
theft
theirs
themat
theme
then-head
then-owne
then-pres
theori
therapi
therapist
thereaft
theresa
thermo
theron
thesp
thewli
thick
thick-fil
thierri
thiev
thigh
thin
thing
think
third
third-hig
third-lar
third-par
third-qua
third-rou
thom
thoma
thoma-ani
thompson
thomson
thorn-smi
thornton
thorough
thoroughl
thousand
thrashi
threat
threaten
three-goa
thrift
thrill
thriller
thrive
throat
throat-sl
thrombosi
throw
thrown
thrust
thun
thunderbi
thunderbo
thurman
thurmond
thursdai
thwart
ti
tiara
tibet
tibetan
ticket
ticketmas
ticotin
tide
tie
tiebreak
tie-in
tier
tiger
tight
tighten
tighter
tightli
tile
till
tilt
tim
timberlan
time
timecop
timeslot
timmi
timpani
tin
tina
tingstad
tip
tiptop
tirad
tire
tisha
tissu
titan
titl
tito
tm
tmm
tnn
tnt
tobacco
todai
toddler
toho
tohoku
toi
toil
toilet
tokuma
tokyo
toll
tolochena
tom
tomei
tomita
tomkin
tomlin
tommi
tone
toner
tongu
toni
tonight
t-onlin
tool
tooth
toothbrus
toothless
toothpast
top
top-10
topaz
top-level
topograph
top-rate
top-sell
torch
torn
tornado
toronto
toronto-c
tortur
toshiba
total
totalitar
toti
touch
touchback
touchdown
touchston
tough
toughen
tougher
tour
tourism
tourist
tournamen
tournei
tout
towel
tower
town
townsend
townsfolk
townspeop
toxic
toymak
toyota
tp
trace
tracei
tracinda
track
tract
tractor
trade
trademark
trader
tradit
tradition
traffic
tragedi
tragic
trail
trailer
train
trainer
trait
traitor
trak
tran
transact
transcrip
transfer
transfus
transit
translat
transmiss
transmitt
transpar

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

transplan
transport
trauma
travail
travel
travers
travolta
treach
treacher
treasur
treat
treati
treati-re
treatment
tree
trei
trek
tremend
tremor
trend
trendi
trepid
tress
trevino
tri
trial
tribal
tribe
tribul
tribut
trick
tricki
tricycl
trie
trigger
triglycer
trillion
trilogi

trimark
trimest
trinidad-
trio
trip
trippleho
trisha
tristar
triton
triumph
triumphan
trivial
troi
troop
trooper
trophi
trotman
troubl
troubl-sh
troup
truck
trucker
true
true-life
truman
trumpet
trust
trustbust
truste
truth
try
trypsin
tsamcho
t-shirt
tsumura
tub
tubal
tuberculo

tucci
tuck
tucker
tuckett
tucson
tumbl
tummler
tumor
tune
tuner
tunja
tunnei
tunnel
tur
turbul
turek
turkei
turku
turmoil
turn
turner
turntabl
turquois
turtl
turturro
tuskege
tussl
tutelag
tuzla
tv
tv2
tv3
tvn
tv-scandi
twa
tweak
twentysom
twice

twin
twin-engi
twist
two
two-drug
two-paren
two-run
two-stori
two-third
two-time
ty
tycoon
tyler
tymchyshy
type
type-a
typewrit
typic
tyson
tyson-eva
tyson-hol
u.k.
u.n.
u.s
u.s.
u.s.a.
u2
uaw
uaw-gm
uchima
ucla
udal
uhri
uk
ukrain
ullman
ultim
ultimatum

ultrasoun
ultraspar
ultraviol
uma
umax
ummer
unaccept
unacquir
unadjust
unaffect
unalt
unanim
unappet
unattract
unauthor
unbeaten
unbleach
unblink
unbound
unc
uncensor
uncertain
unclaim
unclear
uncoat
unconsci
uncov
und
undeni
underag
under-dia
underdog
underesti
undergo
undergon
undergrou
undermin
undersid

underst
understan
understat
undertr
underw
underwai
underweig
underwood
underworl
undevelop
undisclos
undon
undress
uneduc
unemploi
unemploy
unfair
unfairli
unfamilia
unfocus
unforesee
unfortun
unfractio
unfunni
unhappi
unhealthi
unheard
uni
unifi
uniform
uninteres
union
unionist
uniqu
unit
univers
unix
unknown

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

unlaw	upfront	vacat	variant	vespucio	villella
unleash	upgrad	vacation	variat	vessel	villeneuv
unlik	upheld	vaccin	varieti	veteran	vilniu
unload	uphold	vagin	vascular	veto	vinc
unmark	upi	vagu	vast	veysset	vincent
unnecessa	upn	vain	vatican	vh1	vineyard
unnerv	upper	vajna	vaudevil	viabil	vintag
unopen	uproar	val	vaughn	viabl	vintner
unpaid	upset	valdez	v-chip	viachesla	violat
unparal	upsw	valdina	vega	viacom	violenc
unprofit	uptick	vale	veget	viasat	violenc-g
unpromis	ur	valedicto	vegetaria	vibe	violent
unprotect	urban	valenti	vehement	vibrant	violin
unreleas	urbona	valentin	vehicl	vibrio	violinist
unresolv	ureter	valentino	vein	vice	viral
unsatisfa	ureterosc	valet	veinott	vice-chai	virgin
unseat	urg	vall	velodrom	vicent	virginia
unsmart	uri	vallegran	velvet	vice-pres	virtual
unsolicit	uria	vallei	venardo	viciou	virtual-p
unspecifi	urin	valour	vendor	victim	virtuoso
unspool	urinari	valu	venic	victor	viru
unsur	urquhart	valuabl	venou	victori	virul
unsustain	us	valu-clar	ventur	victoria	virus
untitl	usa	vampir	venu	vid	visa
unusu	user	van	verdi	video	vishwa
unveil	usoc	vancomyci	verdict	videodisc	visibl
unw	uss	vancouv	vere	videotap	vision
unwant	usual	vandal	verifi	videotron	visit
unwieldi	uta	vandamm	verio	vienna	visitor
up	utah	vander	verit	viennes	visual
up!	uteru	vanessa	vermont	vietnam	viswanath
up-and-co	utian	vanish	verona	view	vita
upbeat	util	vann	vers	viewer	vital
upbring	utterli	vanska	versa	viewershi	vitamin
upcom	u-turn	var	versac	vigor	vittorio
updat	uunet	varanu	version	villafran	vivian
uperman	uup	varga	vertebra	villag	vivica
upernatur	uzdelewic	vari	verv	villain	vividli

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

vivienn	waco	warfar	weaken	weiss	wharton
vixenish	waddl	warm	weaker	weissman	whatev
vladimir	wade	warmer	wealth	weisz	wheel
vlsi	waga	warn	wealthi	weitz	wheeler
vnu	wage	warner	wealthies	weizman	wheeler-d
vocal	wagner	warplan	weapon	welch	whelan
vocalist	wahlberg	warsaw	wear	welcom	wherebi
voe	wai	wartim	wearer	welfar	wherev
vogel	wail	wash	weari	well	whiblei
vogu	waistcoat	washingto	weather	wellcom	whiskei
voic	wait	wash-up	weatherca	well-kept	whiski
voiceov	waiter	wasim	weaver	well-oil	whisper
voic-reco	wake	wasserman	web	welshman	whit
void	wakeham	wast	web;	welt	whitak
volatil	waldman	watanab	web-base	wemba	white
volcan	waldo	watch	webber	wemblei	whitewash
volcano	waldorf	watchdog	websit	wendel	whitewat
volic	wale	watcher	webster	wender	whitfield
vollhardt	walk	water	webtv	wendi	whitford
volunt	walker	watercraf	wed	wendt	whitnei
voluntari	walk-per-	waterfron	wedgewort	wept	whittak
volusia	wall	waterg	week	werb	who
vomit	wallac	waterhors	weekdai	weslei	whoever
von	wallet	water-ski	weekend	wesson	wholesal
vonnegut	wal-mart	waterston	weekli	west	whoopi
vornado	walston	waterwai	weeklong	westaim	whowher
vote	walt	watt	weep	westchest	wi
voucher	walter	wave	wegman	western	widen
vow	walton	wavephor	weigh	westgat	wider
voyag	wang	wax	weight	westin	wide-rang
v-roi	wangchuk	wayan	weightlif	westingho	widesprea
vsd	wanna	wayland	weiland	westminst	widow
vulgar	wannab	wayn	weinberg	weston	wiener
vy	want	waynesvil	weinert	westwood	wiest
w	war	wb	weinrich	weyman	wife
wabc	ward	wba	weinstein	wga	wigan
wachowski	wareh	we	weintraub	wgnx	wild
wacki	warehous	weak	weird	whaler	wilder

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

wildli	winterbot	woodfolk	wright	yakusho	younger
wildlif	winwood	woodi	wrist	yakuza	youngest
wile	wipo	woodstock	write	yanc	youngster
wilhoit	wire	woolf	writedown	yanke	youngwort
wilk	wireless	woosnam	write-in	yanko	youssou
will	wisconsin	word	writer	yannick	youth
willem	wisdom	word-of-m	writer/di	yaphet	yuan
willi	wise	wore	writer-di	yard	yugoslav
william	wishmast	work	writer-in	yardag	yuko
willing	wislawa	workdai	written	yarn	yuliet
willock	wisot	worker	wrn	yassin	yura
wilmingto	wit	workplac	wrong	y-bear	yuri
wilson	witchcraf	world	wrongdo	y-chromos	zach
wim	withdraw	world-cla	wsj	ye	zack
win	withdrawn	worldcom	wsp	yeah	zahn
winchel	withdrew	worldcom-	wto	year	zaid
wind	witherspo	world-fam	wurtz	year-end	zaillian
windfal	withstand	worldmark	wvue	yearli	zakiya
windolf	witti	worldmus	ww2	yearlong	zanamivir
window	wive	worldnet	wyle	yearn	zanuck
window/ma	wizard	worldwid	wyli	yeast	zargaran
window95	wjro	wormwood	wyom	yell	zealand
windshiel	wmvp	worn	x-bear	yellow	zealou
windsor	wnba	worri	x-chromos	yeltsin	zeisz
wine	woefulli	worship	xena	yesteryea	zeitgeist
winemak	woessner	wort	xenia	yiddish	zellweg
winfrei	wole	worth	xerox	yield	zen
wing	wolf	worthi	xerox-bra	yimou	zenz
wink	wolfgang	wozniak	x-file	ying	zero
winkler	wolv	wpix	xiao	yitzhak	zhang
winkless	woman	wrai	x-inactiv	yoga	zhejiang
winless	women	wrap	x-link	yoko	ziegler
winner	won	wrapper	x-rai	yonker	zienowicz
winner-ta	wonder	wrather	x-sperm	york	ziff
winni	wong	wreak	yacht	york-base	ziggi
winninges	woo	wreck	yahoo	yorker	zippel
winston	wood	wrestl	yahoo!	yorn	zmg
winter	woodard	wretch	yakker	young	zone

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

zoom	zuckerman	zuniga	zurinaga
zorro	zunder	zurich	zylka

ที่กล่าวมาทั้งหมดคือคำสำคัญที่ได้จากการเรียนรู้ของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ซึ่งเลือกชุดการทดลองที่ดีที่สุดของชุดข้อมูลต่างๆ คือชุดข้อมูล CSTR มีจำนวนคำสำคัญ 1,555 คำ ชุดข้อมูล Reuters-Top10 มีจำนวนคำสำคัญ 5,085 คำ และชุดข้อมูล K-dataset มีจำนวนคำสำคัญ 9,614 คำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค.

ผลงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์และได้รับการตีพิมพ์

1. Worapoj Kreesuradej and Puangpaka Kunasit, "Text Processing Simplified ARTMAP Neural Network," WSEAS Transactions on Information Science and Applications, Issue 2, Volume 2, Austria, February, 2005, pp.116-120.
2. พวงผกา คุณาสิทธิ์ และ วรพจน์ กรีสุระเดช, "การจำแนกประเภทเอกสารโดยใช้โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ (Text Categorization using a Text Simplified ARTMAP Neural Network)," Data Mining : Proceeding of the National Computer Science and Engineering Conference (NCSEC 2005), Bangkok, October, 2005.

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS

Issue 2, Volume 2, February 2005

ISSN 1790-0832 http://www.wseas.org

A Proposed Multiagent System Utilizing Swarm Intelligence for Collective Learning and Cooperation in a Networked Environment 65
Gregory Julock

e-banking Prediction using Data Mining Methods 71
Vasilis Aggelis, Panagiotis Anagnostou

Feature Selection Algorithms in Classification Problems: An Experimental Evaluation 77
Michael Doumpos, Athina Salappa

Improving the Knowledge Performance Using "Ternary Grid" Knowledge Acquisition and Model 83
Yuliadi Erdani, Axel Hunger, Stefan Werner

A Method of Small Data Set Learning for Early Knowledge Acquisition 89
Fengming Michael Chang, Ming-Yuan Chiu

Hybrid Expert System 95
Vaclav Jrsik, Petr Honzik

An Efficient Feature Selection using Multi-Criteria in Text Categorization for Naive Bayes Classifier 98
Son Doan, Susumu Horiguchi

Learning from Noise Data with the Help of Logic Programming Systems 104
Elena Bauer, Gabriella Kokai

Approximate Reasoning in Structural Funds Knowledge Management System 110
Tadeusz A. Grzeszczyk

Text Processing Simplified ARTMAP Neural Network 116
Worapoj Kreesuradej, Puangpaka Kunasit

ANFIS Networks Design using Hybrid Genetic and SVD Methods for Modelling of the Level Variations of the Caspian Sea 121
A. Mehrdad, N. Nariman-zadeh, A. Jamali, A. Teymoorzadeh

A Fuzzy Extension in ALC Description Logics 127
V. Tzouvaras, P. Zoitos, G. Stamou

A Review of Different Approaches for Measuring Knowledge Value in Organizations 132
Kamran Fatahi, Abbas Afrazeh

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Toward a Multi-Tier Index for Information Retrieval System 251
Emad S. Elabd, Ahmed Z. Emam, Nabil A. Ismail, Fawzy A. Torkey

A solution for displaying medical data models on mobile devices 258
Fabrizio Lamberti, Andrea Sanna

Optimized Multiple Wavetable Interpolation 265
Jonathan Mohr and Xiaobo Li

Hiding Information in Side-Match VQ Compressed Images 274
Shinfeng D. Lin, Shih-Chieh Shie, Chih-Ming Fang

White Noise Reduction of Audio Signal using Wavelets Transform with Modified Universal Threshold 279
Matko Saric, Luki Bilicic, Hrvoje Dujmic

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

116 WSEAS TRANS. ON INFORMATION SCIENCE & APPLICATIONS Issue 2, Volume 2, February 2005 ISSN: 1790-0832

Text Processing Simplified ARTMAP Neural Network

WORAPOJ KREESURADEJ and PUANGPAKA KUNASIT
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Ladkrabang, Bangkok 10520
THAILAND
worapoj@it.kmitl.ac.th and s5066009@kmitl.ac.th

Abstract: - This paper proposes text processing simplified ARTMAP neural network. The algorithm works directly on textual information without transforming to numerical value. The input layer of the neural network can directly receive a qualitative value without mapping the qualitative value into numerical value. Then, based on simplified fuzzy ARTMAP neural network and the concept of similarity measure for symbolic objects, the proposed neural network can assigns class labels to the objects correctly.

Key-Words: - Text mining, Text classification, Text categorization, Document classification, Document categorization, Simplified ARTMAP neural network and Neural networks.

1 Introduction

Several classification techniques for objects whose feature values are numerical values are well known. Several neural networks such as backpropagation neural networks, ARTMAP, fuzzy ARTMAP and simplified fuzzy ARTMAP neural network are proposed for classification. Recently, the classification problems are extends for document classification. To classify a document by using typical neural networks, a document has to be mapped onto a document representation that has quantitative features. The vector-space model is the most widely-used document representation [1]. Then, the typical neural networks or classifiers can be applied for classifying documents. However, the utilization of the vector-space model may led to a very high dimensional feature space. In addition, this feature space is generally not free from correlation.

Unlike conventional classifiers, the proposed algorithm works directly on textual information without mapping documents onto some numeric document representations. The proposed neural network, text processing simplified ARTMAP neural network, is based on the architecture of the simplified fuzzy ARTMAP neural network and the concepts of symbolic objects similarity measure [2],[3] and [4]. The experimental results show that the text processing simplified ARTMAP neural network can assigns class labels to documents correctly. This paper is introduced as following. In the 2nd section, we present similarity measure for symbolic objects. In the 3rd and 4th section, the text processing simplified ARTMAP neural network is introduced. In the 5th section, we present the concepts of document representation. In the 6th section, experimental results are presented. Finally, the 7th section is conclusions.

2 Similarity Measure

According to K.C. Gowda [3], the definition of similarity between two symbolic objects A and B are written as Cartesian product of A_k and B_k features as:

$$A = A_1 * A_2 * \ldots * A_d, \quad (1)$$

$$B = B_1 * B_2 * \ldots * B_d. \quad (2)$$

The similarity between two symbolic objects A and B is written as:

$$S(A,B) = S(A_1,B_1) + S(A_2,B_2) + \ldots + S(A_d,B_d). \quad (3)$$

For the k^{th} feature, $S(A_k,B_k)$ is defined using the following three components:

1. the similarity component due to position is $S_p(A_k,B_k)$,
2. the similarity component due to span is $S_s(A_k,B_k)$,
3. the similarity component due to content is $S_c(A_k,B_k)$.

The similarity component due to "position" arises only when the feature type is quantitative. It indicates the relative position of two feature values on real line. Since this paper is only dealing with text data, the similarity component due to position is neglected. The similarity component due to span

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

WSEAS TRANS. ON INFORMATION SCIENCE & APPLICATIONS Issue 2, Volume 2, February 2005 ISSN: 1790-0832 117

indicates the relative sizes of the feature values without referring to common part between them. The similarity component due to content is a measure of common parts between two feature values [3],[4] and [5].

The similarity component due to span is defined as:

$$S_s(A_k,B_k)=\frac{(l_a+l_b)}{2.l_s}, \quad (4)$$

and the similarity component due to content defined as:

$$S_c(A_k,B_k)=\frac{inters}{l_s}, \quad (5)$$

where

l_a = length of A_k or number of elements in A_k,

l_b = length of B_k or number of elements in B_k,

$inters$ = length of the intersection of A_k and B_k,

l_s = span length of A_k and B_k or number of elements in $A_k \cup B_k$.

Therefore, the net similarity between A_k and B_k is

$$S(A_k,B_k)=S_s(A_k,B_k)+S_c(A_k,B_k). \quad (6)$$

3 The Text Processing Simplified ARTMAP Neural Network

The proposed neural network consists of three layers: input layer (F_1), output layer (F_2) and category layer (F_3). Each neural node in the F_1 layer is connected to each neural node in the F_2 layer by two weighted pathways. The F_1 neural node, i.e., X_i, is connected to F_2 neural node by bottom-up weights, b_{ij}. Similarly, the F_2 neural node, i.e., Y_j, is connected to the F_1 neural node by top-down weights, t_{ji}. The F_2 layer is connected to some neural nodes in the F_3 layer if a category is associated with the winning output node. Unlike conventional neural networks, the F_1 layer can receive qualitative values. The bottom-up weight, b_{ij} and top-down weight, t_{ji} contain qualitative values and degrees of association of the qualitative values is defined as [5]:

$$b_{ij}=\{(A_{1ij},e_{1ij}),(A_{2ij},e_{2ij}),(A_{3ij},e_{3ij}),...,(A_{pij},e_{pij})\}. \quad (7)$$

A_{pij} is the p^{th} qualitative value of the weight and e_{pij} is the degree of association of this qualitative value to the weight. The value of e_{pij} has values between 0 to 1. e_{pij} equals to zero, i.e., $e_{pij}=0$, if A_{pij} is not a part of the weight. While e_{pij} equals to one, i.e., $e_{pij}=1$, if the qualitative value has strong association with the weight.

$$t_{ji}=\{(B_{1ji},e_{1ji}),(B_{2ji},e_{2ji}),(B_{3ji},e_{3ji}),...,(B_{pji},e_{pji})\}. \quad (8)$$

B_{pji} is the p^{th} qualitative value of the weight and e_{pji} is the degree of association of this qualitative value to the weight. The value of e_{pji} is between 0 to 1. e_{pji} equals to zero, i.e., $e_{pji}=0$, if B_{pji} is not a part of the weight. While e_{pji} equals to one, i.e., $e_{pji}=1$, if the qualitative value has strong association with the weight. The architecture of the proposed neural networks is shown in figure 1.

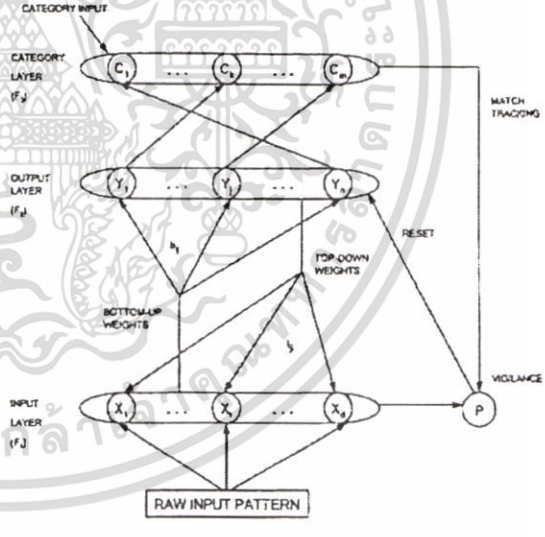


Fig. 1 The architecture of text processing simplified ARTMAP neural network

4 Learning Algorithm

The learning algorithm of The text processing simplified ARTMAP neural network algorithm is based on that of the simplified Fuzzy ARTMAP neural network and the concept of symbolic objects similarity measure as described in the prior section. The algorithm is summarized as below:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

118 WSEAS TRANS. ON INFORMATION SCIENCE & APPLICATIONS Issue 2, Volume 2, February 2005 ISSN: 1790-0832

0th Step: Set parameters values.

1st Step: While stopping condition is false, do the 2nd -14th step.

2nd Step: For each input vector.

$X = (X_1, X_2, ..., X_d)$, do the 3rd -13th step. (9)

3rd Step: Present the input pattern to the F_1 layer and present the category input to the F_3 layer simultaneously.

4th Step: If output node = 0, then

-Create an output node on the F_2 layer to encode this training pattern.

-Initialize bottom-up weights, b_{ij} and top-down weights, t_{ji} by choosing from the training data set arbitrarily and setting degrees for each element of the weight.

-Set the category of the new output node to that of the input pattern.

-Update the bottom-up weight and the top-down weight, do the 13th step.

5th Step: If output node ≠ 0, do the 6th -13th step.

6th Step: Set the vigilance to the baseline vigilance.

7th Step: Evaluate the activation function between X and bottom-up weight for all output nodes.

$$Y_j = (\sum_{i=1}^{d} S(X_i, b_{ij})) / (\sum_{i=1}^{d} S(b_{ij}, b_{ij})), \quad (10)$$

where d is number of feature values.

8th Step: Select the j^{th} highest activated output node.

$$Y_j = \max\{Y_j : j = 1..n\}. \quad (11)$$

9th Step: Evaluate the match function between X_i and the top-down weight that connects to the winning output node, i.e., J.

$$V = (\sum_{i=1}^{d} S(X_i, t_{Ji})) / (\sum_{i=1}^{d} S(X_i, X_i)), \quad (12)$$

where d is number of feature values.

10th Step: Test mismatch reset

$$V \geq \rho. \quad (13)$$

-If $V < \rho$, suppress activation of the current winning output node, i.e., J, do the 8th step until $V \geq \rho$.

-If $V \geq \rho$, do the 11th step.

11th Step: Test category mismatch between the network category and the category input.

-If the network category and the category input are the same category, do the 13th step.

-If the network category and the category input are not the same category, set vigilance to the match value of the winning output node plus a small value. Then, suppress activation of the current winning output node J and do the 8th -11th step.

12th Step: If all output nodes can not match the category input, then

-Create an output node on the F_2 layer to encode this training pattern and

-Initialize bottom-up weights, b_{ij} and top-down weights, t_{ji}.

13th Step: Update the winner bottom-up weights and top-down weights as following:

$$b_{iJ}^{(new)} = b_{iJ}^{(old)} \cap X$$

$$e_{niJ}^{(new)} = \begin{cases} f(e_{niJ}^{(old)} + \beta) & if \quad A_{nij} \in b_{iJ} \cap X, \\ f(e_{niJ}^{(old)} - \beta) & if \quad A_{nij} \notin b_{iJ} \cap X, \\ \beta_0 & otherwise \end{cases} \quad (14)$$

$$t_{ji}^{(new)} = t_{ji}^{(old)} \cap X$$

$$e_{nji}^{(new)} = \begin{cases} f(e_{nji}^{(old)} + \beta) & if \quad B_{nji} \in t_{ji} \cap X, \\ f(e_{nji}^{(old)} - \beta) & if \quad B_{nji} \notin t_{ji} \cap X, \\ \beta_0 & otherwise \end{cases} \quad (15)$$

where $f(.)$ is defined as below:

$$f(x) = \begin{cases} x & if \quad 0 \leq x \leq 1 \\ 0 & if \quad x < 0 \\ 1 & if \quad x > 1 \end{cases} \quad (16)$$

14th Step: Continue with the 2nd -14th step until the stopping condition is true.

5 Document Representation

A document, i.e., D, can be written as Cartesian product of specific values of its features D_k's as [6]:

$$D = D_1 * D_2 * ... * D_d. \quad (17)$$

Unlike a vector-space model, the features have qualitative values which are a set of words that describe the features. As an example, a document can be written as Cartesian product of Title feature and Keyword feature as:

$$Doc = Title * Keyword. \quad (18)$$

Where the values of the Title feature are a set of words that describe the title of the document and the

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

WSEAS TRANS. ON INFORMATION SCIENCE & APPLICATIONS Issue 2, Volume 2, February 2005 ISSN: 1790-0832 119

values of the Keyword feature are a set of keywords of the document.

6 Experimental Results

In this section, the experiment results of the text processing simplified ARTMAP neural network is presented. We use a training dataset of 200 documents that consist of 3 class. Each document in the training dataset can be represented by title feature and keyword feature according to equation (18).

Each feature of a document is qualitative value. Here, Some English alphabets are used to represent value of each feature. The values of title feature and the value of keyword feature for each class are shown in figure 2 and figure 3 respectively.

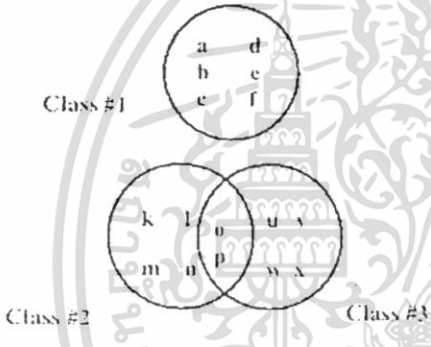


Fig. 2 Venn Diagram of Title feature

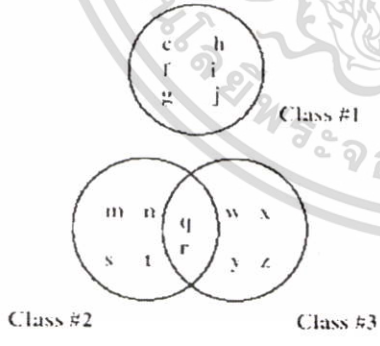


Fig. 3 Venn Diagram of Keyword feature

Some documents of the training dataset are shown in table 1.

Table 1 Some documents of the training dataset

No.	Data Values		Category Input
	Titles	Keywords	
1	l,n,o,p	m,n,q,r,t	2
2	a,d,e,f	c,f,g,h	1
3	k,m,n	m,r,s,t	2
4	o,p,u,w	q,y,z	3
5	v,w,x	r,w,x,y,z	3
6	b,c,d	g,h,i,j	1

The training data consist of 67 documents from the class number 1, 66 documents from the class number 2 and 67 documents from the class number 3.

Then, we use 5 testing datasets of 1000 documents, as shown in table 2, to evaluate the performance of the proposed neural network.

Table 2 The testing dataset

Dataset No.	Members		
	Class 1	Class 2	Class 3
1	330	336	334
2	328	338	334
3	332	333	335
4	332	336	332
5	330	332	338

The evaluate measurement of the proposed neural network is the accuracy rate (r) defined as:

$$r = [(\sum_{i=1}^{c} doc_i)/n] * 100, \qquad (19)$$

where c is a number of data that are the same class, n is a number of all data.

The results from table 3 show 100% accuracy rate for all data set. The experimental results show good performance of the proposed neural network.

Table 3 The experimental results

Data Set No.	Members			Experimental Result			Accuracy Rate
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	
1	330	336	334	330	336	334	100%
2	328	338	334	328	338	334	100%
3	332	333	335	332	333	335	100%
4	332	336	332	332	336	332	100%
5	330	332	338	330	332	338	100%

7 Conclusion

In this paper, we applied the proposed neural network for text classification. The text processing simplified ARTMAP neural network can receive qualitative value directly without transformation. According to the experimental results, the proposed neural network has good performance in classifying documents. In the future, we will conduct some experimentations on the Reuter-21578 news article [7].

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

References:

[1] G. Salton, *Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer*, Addison Wesley Publishing Company, New York, 1989

[2] T. Kasuba, Simplified Fuzzy ARTMAP, *AI Expert*, Vol.8, 1993, pp. 18-25.

[3] K.C. Gowda and E. Diday, Symbolic Clustering Using a New Similarity Measure, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol.22, No.2, 1992, pp. 368-378.

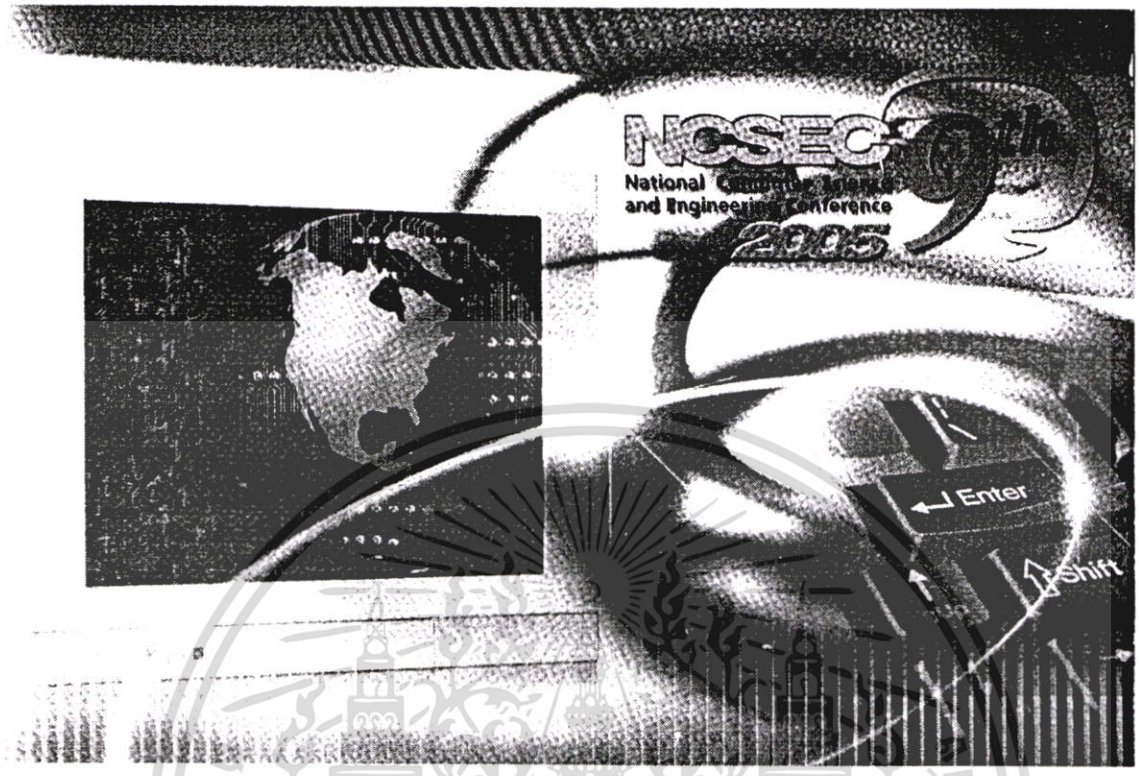
[4] T.V. Ravi and K.C. Gowda, Clustering of Symbolic Objects Using Gravitational Approach, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol.29, No.6, 1999, pp. 888-894.

[5] W. Kreesuradej, N. Chantasut and W. Kruaklai, Clustering Text Data Using Text ART Neural Network, *The WSEAS Trans. on Systems*, Vol.3, I.1, 2004, pp. 200-205.

[6] Y. El-Sonbaty and M.A. Ismail, Fuzzy Clustering for Symbolic Data, *IEEE Trans. on Fuzzy Systems*, Vol.6, No.2, 1998, pp. 195-204.

[7] D.D. Lewis, Reuters-21578 Text Categorization Test Collection Distribution 1.0, available on http://www.daviddlewis.com/resources/testcollections/reuters21578

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



The 9th National Computer Science and Engineering Conference

October 27-28, 2005

University of Thai Chamber of Commerce, Bangkok Thailand

Organized by:

Department of Computer Engineering, School of Engineering,
University of Thai Chamber of Commerce

In Cooperation with:

Electrical Engineering; Electronics, Computer, Telecommunications and Information Technology Association of Thailand (ECTI)

IEEE Communications Society, Thailand Chapter

Sponsored by:

University of Thai Chamber of Commerce

National Electronics and Computer Technology Center (NECTEC)

Sun Microsystems (Thailand)

CS Loxinfo Public Company Limited

Pearson Education Indochina Limited

The OGA Group

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Text Categorization using a Text Simplified ARTMAP Neural Network
การจำแนกประเภทเอกสารโดยใช้โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

NCSEC 9th
National Computer Science and Engineering Conference
2005

พวงผกา คุณาสิทธิ์ และ วรพจน์ กรีสุระเดช
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง
Email: s5066009@kmitl.ac.th, worapoj@it.kmitl.ac.th

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Text Categorization using a Text Simplified ARTMAP Neural Network
การจำแนกประเภทเอกสารโดยใช้โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

พวงผกา คุณาสิทธิ์[1] และ วรพจน์ กรีสุระเดช[2]
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
Email: s5066009@kmitl.ac.th[1], worapoj@it.kmitl.ac.th[2]

บทคัดย่อ

งานวิจัยนี้นำเสนอ การจำแนกประเภทเอกสารโดยใช้โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ซึ่งไม่เหมือนกับโครงข่ายประสาทเทียมอาร์ทแม็พโดยทั่วไปที่รับข้อมูลเข้าได้เฉพาะค่าเชิงตัวเลข โครงข่ายประสาทเทียมนี้สามารถรับข้อมูลเข้าที่เป็นข้อความได้โดยตรง โดยไม่ผ่านขั้นตอนการแปลงข้อมูลจากค่าเชิงคุณภาพไปเป็นค่าเชิงตัวเลข โดยยึดหลักการของโครงข่ายประสาทเทียมซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ และนำแนวความคิดของการวัดความคล้ายคลึงกันของซิมโบลิคอ็อบเจ็คมาประยุกต์ใช้ ผลการทดลอง การทำงานของโครงข่ายประสาทเทียมนี้สามารถจำแนกประเภทเอกสารได้ถูกต้อง

Abstract

This paper proposes text categorization using text simplified ARTMAP neural network. Unlike conventional ARTMAP which use numerical values as inputs, this neural network works directly on textual information without mapping onto some representations which have qualitative value into numerical value. The proposed neural network is based on simplified fuzzy ARTMAP neural network and the concept of similarity measure for symbolic objects. The proposed neural network had satisfactory performance in categorizing text data.

Key Words: Text mining; Text categorization; Text classification; Simplified ARTMAP neural network.

1. บทนำ

ปัจจุบันข้อมูลอิเล็กทรอนิกส์มีปริมาณเพิ่มขึ้นอย่างรวดเร็ว ข้อมูลเหล่านั้น มีบางส่วนเป็นข้อมูลประเภทข้อความ (Text) ซึ่งส่วนใหญ่จะเป็นข้อมูลแบบกึ่งโครงสร้าง (Semi-structured Data) และข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data) ข้อมูลนี้จะถูกจัดเก็บไว้ตามแหล่งเก็บต่างๆ เช่น Text Database เป็นต้น การจะนำความรู้ที่ถูกซ่อนอยู่ในข้อมูลเหล่านี้มาใช้ประโยชน์ คือ การทำ Text Mining ซึ่งวิธีการที่นิยมในการทำ Text Mining ได้แก่ การจัดกลุ่มเอกสาร (Text Clustering) และ การจำแนกประเภทเอกสาร (Text Classification หรือ Text Categorization)

การจำแนกประเภทเอกสารมีหลายวิธีเช่น โครงข่ายประสาทเทียม (Neural networks) นาอีฟ เบย์ (Naive Bayes) เค-เนียเรส เนเบอร์ (K-Nearest Neighbor) ซัพพอร์ท เวกเตอร์ แมชชีน (Support Vector Machines) ในที่นี้จะกล่าวถึงวิธีการจำแนกประเภทเอกสารด้วยโครงข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสาทเทียม โดยทั่วไปจะมีการรับข้อมูลที่เป็นตัวเลข ซึ่งการรับข้อมูลเข้า จะต้องทำการแปลงค่าให้เป็นค่าเชิงตัวเลขเสียก่อน โครงข่ายประสาทเทียมที่นำมาใช้ในการจำแนกประเภทเอกสาร มีการเรียนรู้แบบมีการชี้นำ (Supervised Learning) ที่รู้จักกันดี คือ โครงข่ายประสาทเทียม Backpropagation ซึ่งการรับข้อมูลเข้า เป็นค่าเชิงตัวเลข จึงต้องมีการแปลงค่าก่อน

การแปลงค่าของข้อมูลที่เป็นข้อความให้อยู่ในลักษณะของค่าเชิงตัวเลขเพื่อให้สามารถนำข้อมูลเข้าสู่แบบจำลองได้นั้นเรียกว่า การแทนเอกสาร (Document Representation) วิธีการหนึ่งที่ใช้กันอย่างแพร่หลาย คือ วิธีการ Vector Space Model นำเสนอโดย Salton G. [1] มีลักษณะข้อมูลเป็น Matrix Array 2 Dimensions แต่การทำ Document Representation โดยวิธีการ Vector Space Model อาจทำให้เกิดปัญหา High Dimensions ซึ่งมีผลกับการทำงานของแบบจำลอง ที่จะต้องใช้เวลาในการคำนวณมากขึ้น และอาจทำให้ข้อมูลสูญเสียความหมายในตัวมันเอง

เนื้อหาของงานวิจัยนี้ประกอบไปด้วย 6 หัวข้อ ได้แก่ หัวข้อที่ 1 บทนำ หัวข้อที่ 2 การวัดค่าความคล้ายคลึงกันของเอกสาร หัวข้อที่ 3 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ หัวข้อที่ 4 การเรียนรู้ของอัลกอริทึม หัวข้อที่ 5 การทดลองและผลการทดลอง และหัวข้อที่ 6 สรุปผล

2. การวัดความคล้ายคลึงกันของเอกสาร (Similarity Measure)

การวัดความคล้ายคลึงของเอกสารนั้น [2] จะต้องผ่านขั้นตอนการแทนเอกสารต่างๆ ให้อยู่ในรูปแบบที่สามารถนำมาใช้วัดค่าได้ เอกสารต่างๆ ประกอบด้วยคุณสมบัติต่างๆ ซึ่งสามารถแทนให้อยู่ในรูปแบบของ Cartesian Product ได้ดังนี้ [3]

$$Doc = D_1 \times D_2 \times D_3 \times \ldots \times D_d \quad (1)$$

d คือจำนวนคุณสมบัติของเอกสาร ซึ่งมีลักษณะเป็นข้อความ

ตัวอย่าง คุณสมบัติ (Features) ของเอกสารที่มี 2 คุณสมบัติ คือคุณสมบัติชื่อเรื่อง (Title Feature) และคุณสมบัติคำสำคัญ (Keyword Feature) โดยสามารถแทนให้อยู่ในรูปแบบของ Cartesian Product ได้ดังนี้

$$Doc = Title \times Keyword \quad (2)$$

เมื่อ Title คือคำที่ใช้อธิบายชื่อเรื่องของเอกสาร

Keyword คือคำที่ใช้อธิบายคำที่พบบ่อยในเอกสาร หรือคำสำคัญของเอกสาร

การเปรียบเทียบความคล้ายคลึงของเอกสาร ระหว่างเอกสาร A และเอกสาร B ตามแนวความคิดของ El-Sonbaty [3] สามารถเขียนเป็นสมการได้ดังนี้

$$S(A,B) = \sum_{k=1}^{d} S(A_k, B_k) \quad (3)$$

ในการเปรียบเทียบความคล้ายคลึงของ $S(A_k,B_k)$ มีการเปรียบเทียบ 2 ส่วนย่อยคือ ส่วนที่เป็นขนาดของคุณสมบัติเรียกว่า Span $S_s(A,B)$ และส่วนที่เป็นเนื้อหาของคุณสมบัติ เรียกว่า Content $S_c(A,B)$ ซึ่งนิยามสมการไว้ดังนี้

ส่วนของการเปรียบเทียบความคล้ายคลึงของ Span มีนิยามตามสมการที่ 4

$$S_s(A_k,B_k) = \frac{(l_a + l_b)}{2l_s} \quad (4)$$

ส่วนของการเปรียบเทียบความคล้ายคลึงของ Content มีนิยามตามสมการที่ 5

$$S_c(A_k,B_k) = \frac{inters}{l_s} \quad (5)$$

นิยามของสัญลักษณ์ในสมการมีดังนี้

l_a = จำนวนสมาชิกทั้งหมดในคุณสมบัติของเอกสาร A

l_b = จำนวนสมาชิกทั้งหมดในคุณสมบัติของเอกสาร B

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

inters = จำนวนของสมาชิกทั้งหมดที่ Intersection กันระหว่างคุณสมบัติของเอกสาร A และคุณสมบัติของเอกสาร B

l_k = จำนวนของสมาชิกทั้งหมดที่ Union กันระหว่างคุณสมบัติของเอกสาร A และคุณสมบัติของเอกสาร B

เมื่อเปรียบเทียบความคล้ายคลึงกันครบทั้งสองส่วนย่อยแล้ว ให้ทำการรวมค่าผลลัพธ์จากสองส่วนย่อยเข้าด้วยกัน ดังสมการ Net Similarity ดังนี้

$$S(A_k, B_k) = S_s(A_k, B_k) + S_c(A_k, B_k) \quad (6)$$

ค่าความคล้ายคลึงกันของ Net Similarity นี้มีค่าอยู่ระหว่าง [0.5,2]

3. สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

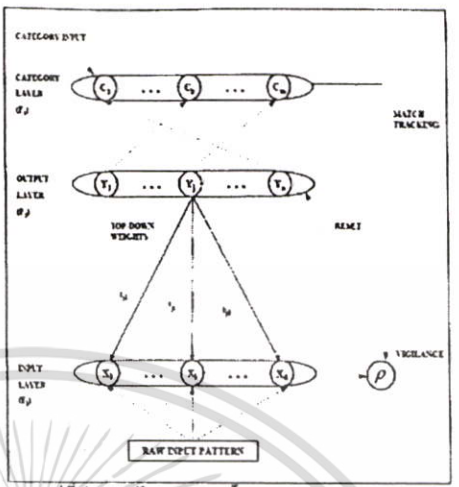
สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ ซึ่งได้ประยุกต์จากโครงข่ายประสาทเทียมซิมพลิไฟด์ฟัซซี่อาร์ทแม็พ [4] ประกอบด้วยชั้นการทำงาน 3 ชั้น (Layer) คือ 1. Input Layer (F_1) 2. Output Layer (F_2) และ 3. Category Layer (F_3) ดังรูปที่ 1

ในส่วนของ Input Layer จะมีการรับข้อมูลเข้า (Input Pattern) ที่เป็นข้อความ ซึ่งนิยามดังนี้

$$D = (D_1, D_2, ..., D_d) \quad (7)$$

โดยที่ d คือจำนวนคุณสมบัติของ D

ระหว่าง Input Layer และ Output Layer มีชุด Weights เรียกว่า Top-down Weight ซึ่งจะเชื่อมต่ออย่างสมบูรณ์และส่งข้อมูลระหว่างโหนดที่ i ใน F_1 Layer กับโหนดที่ j ใน F_2 โดยที่ Top-down Weight จะเก็บค่าใน Weight เป็น Qualitative Value และค่าแสดงความเป็นสมาชิก (Degree) ของ Weight เพื่อให้โครงข่ายประสาทเทียม สามารถหาความคล้ายคลึงของข้อมูลเข้าที่เป็น Qualitative Value ซึ่งในที่นี้คือข้อความได้ [5] ดังสมการที่ 8



รูปที่ 1. สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

$$t_{ji} = \{(A_{1ji}, e_{1ji}), (A_{2ji}, e_{2ji}), ..., (A_{pji}, e_{pji})\} \quad (8)$$

กำหนดให้ A_{pji} คือ ค่า Qualitative Value ของ Top-down Weight และ e_{pji} คือค่า Degree ซึ่งแสดงความเป็นสมาชิกของ A_{pji}

ค่า e_{pji} ของ A_{pji} ในโครงข่ายประสาทเทียมนี้ มีค่าอยู่ระหว่าง 0 ถึง 1 ซึ่งค่านี้เป็นค่าที่ให้ระดับความเป็นสมาชิกของ A_{pji} กับข้อมูลเข้า ถ้า Degree ของ A_{pji} มีค่าเท่ากับ 0 ให้ความหมายว่า Qualitative Value ของ A_{pji} ไม่ได้เป็นส่วนหนึ่งของข้อมูลเข้าลำดับที่ i ถ้า Degree ของ A_{pji} มีค่าเท่ากับ 1 ให้ความหมายว่า Qualitative Value ของ A_{pji} เป็นสมาชิกของข้อมูลเข้าลำดับที่ i อย่างสมบูรณ์

ใน Output Layer นี้ จะหาโหนดที่ชนะ โดยใช้หลักการหาค่าความคล้ายคลึงกันระหว่างข้อมูลเข้า กับค่า Qualitative Value ของ Weight และถ้าโหนดที่ชนะมีค่ามากกว่าหรือเท่ากับค่า Vigilance ถือว่าระดับความคล้ายคลึงกันของการจำแนกกลุ่มข้อมูลมากพอ และตรวจสอบว่ากลุ่มที่จำแนก ตรงกับกลุ่มที่รับเข้ามา (Category Input) หรือไม่ ถ้าตรงกัน ถือว่าการจำแนกกลุ่มสำเร็จและ Output Node ชี้ไปที่ Class ใน Category Layer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่ Category Layer มีการส่ง Category Input พร้อมๆ กับ ข้อมูลเข้า และในขั้นนี้จะบอกว่าคำตอบคือ Class ใด โดยการชี้คำตอบมาจาก Output Node ใน Output Layer

4. การเรียนรู้ของอัลกอริทึม (Learning Algorithm)

การเรียนรู้ของอัลกอริทึมโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ มีรายละเอียด ดังรูปที่ 2

Step 0 กำหนดค่าพารามิเตอร์ต่างๆ ดังนี้

ค่า Vigilance Parameter (ρ) อยู่ในช่วง 0-1

ค่า Learning Rate (β) ให้มีค่าน้อยๆ

ค่า Small Value (ε) ให้มีค่าน้อยๆ

Step 1 ยังไม่ครบเงื่อนไขในการหยุด (จำนวน Epoch) ทำ Step 2-14

Step 2 เลือกข้อมูลเข้า

$$D = (D_1, D_2, \ldots, D_d) \quad (9)$$

ทำ Step 3-13

Step 3 ส่งค่าข้อมูลเข้า ไปยังชั้น F_1 พร้อมทั้งส่งค่า Category Input ไปยังชั้น F_2

Step 4 ถ้า Output Node เท่ากับ 0

-ให้สร้าง Output Node ใหม่ เป็นโหนดที่ 1 ในชั้น F_2

-กำหนดให้ Top-down Weight ของ Node ใหม่เหมือนกับข้อมูลเข้า และกำหนดค่า Degree เริ่มต้น

-กำหนดให้ Output Node ใหม่นี้เป็นตัวแทนของ Class ที่ข้อมูลเข้าชุดนี้เป็นสมาชิก

Step 5 ถ้า Output Node ไม่เท่ากับ 0

ทำ Step 6-13

Step 6 กำหนด Vigilance Parameter ให้เป็น Baseline Vigilance $(\overline{\rho})$

รูปที่ 2. อัลกอริทึมของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ

Step 7 คำนวณค่าตาม Activation Function ของแต่ละโหนดใน Network จนครบทุกโหนด

$$Y_j = (\sum_{i=1}^{d} S(X_i, t_{ji})) / (\sum_{i=1}^{d} S(t_{ji}, t_{ji})) \quad (10)$$

d คือ จำนวนของคุณสมบัติ

Step 8 เปรียบเทียบโหนด J ที่มีค่ามากที่สุดเป็นโหนดที่ชนะ (Winning Node)

$$Y_J = \max\{Y_j : j = 1..n\} \quad (11)$$

Step 9 คำนวณค่าตาม Match Function ที่เชื่อมต่อกับ Winning Node J

$$V = (\sum_{i=1}^{d} S(X_i, t_{Ji})) / (\sum_{i=1}^{d} S(X_i, X_i)) \quad (12)$$

d คือ จำนวนของคุณสมบัติ

Step 10 ทดสอบ Mismatch Reset (Vigilance Criterion)

$$V \geq \rho \quad (13)$$

-ถ้า $V < \rho$ ให้ขับยั้งโหนด J ทำ Step 8 ใหม่จนกว่า $V \geq \rho$

-ถ้า $V \geq \rho$ ทำ Step 11

Step 11 ตรวจสอบว่า Category หรือ Class ที่ตอบตรงกับ Category Input ที่รับเข้ามาหรือไม่

-ถ้า Output Node ชี้ตรงกับ Category Input

-ทำ Step 13

-เซต Indicator เป็น No Mismatch (Match)

-ถ้า Output Node ชี้ไม่ตรงกับ Category Input

-ให้เพิ่ม Vigilance ของโหนด J เล็กน้อย โดยให้

$$\rho = V + \varepsilon \quad (14)$$

-ขับยั้งโหนด J

-เซต Indicator เป็น Category Mismatch

-ทำ Step 8-11 ต่อไปเรื่อยๆ จนกว่าจะครบทุกโหนด

รูปที่ 2. (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Step 12 เมื่อเช็คจนครบทุกโหนดแล้วไม่ผ่าน Step 10 (Mismatch) หรือไม่ผ่าน Step 11 (Category Mismatch)
-ให้สร้าง Output Node ใหม่ ในชั้น F_2
-กำหนดให้ Top-down Weight ของ Node ใหม่ให้เหมือนกับข้อมูลเข้า และกำหนดค่า Degree เริ่มต้น
-กำหนดให้ Output Node ใหม่นี้เป็นตัวแทนของ Class ที่ข้อมูลเข้าชุดนี้เป็นสมาชิก

Step 13 ปรับปรุงค่า Top-down Weight ของโหนด J

$$t_{Ji}^{(new)} = t_{Ji}^{(old)} \cap X$$

$$e_{n,Ji}^{(new)} = \begin{cases} f(e_{n,Ji}^{(old)} + \beta) & if \quad A_{ni} \in t_{Ji} \cap X, \\ f(e_{n,Ji}^{(old)} - \beta) & if \quad A_{ni} \notin t_{Ji} \cap X, \\ \beta & otherwise \end{cases} \quad (15)$$

กำหนดให้ f(.) นิยามดังนี้

$$f(x) = \begin{cases} x & if \quad 0 \le x \le 1 \\ 0 & if \quad x < 0 \\ 1 & if \quad x > 1 \end{cases} \quad (16)$$

Step 14 ทดสอบเงื่อนไขในการหยุด ถ้าครบจำนวนรอบ (Epoch) ให้หยุด

รูปที่ 2. (ต่อ)

5. การทดลองและผลการทดลอง

การทดลองของงานวิจัยนี้ ได้ใช้ข้อมูลที่เป็นเอกสาร Computer Science Technical Report (CSTR) [6] ซึ่งเป็นข้อมูลที่ใช้ในการจำแนกประเภทเอกสาร (Text Categorization) และการจัดกลุ่มเอกสาร (Text Clustering) โดยข้อมูล CSTR ที่ใช้ในการทดลองเลือกข้อมูลที่อยู่ระหว่างปี ค.ศ. 1991 ถึง ค.ศ. 2002 ซึ่งมีเอกสารทั้งหมด 476 เอกสาร และมี 4 กลุ่มเอกสาร [7] โดยใช้คุณสมบัติ 2 คุณสมบัติคือ Title และ Keyword รายละเอียดของแต่ละกลุ่มเอกสารแสดงในตารางที่ 1

ตารางที่ 1. จำนวนข้อมูลแต่ละกลุ่มเอกสาร

ชื่อกลุ่ม	จำนวนเอกสาร
AI	103
Robotics and Vision	74
Systems	172
Theory	127

แต่ละกลุ่มเอกสารอยู่ในรูปแบบแฟ้มข้อมูล HTML ตัวอย่างของเอกสารแสดงดังรูปที่ 3

Hemaspaandra, Lane A., "Teaching Computational Complexity: Resources to Treasure", TR543, Computer Science Dept., U. Rochester, November 1994.
94.tr543.Teaching_computational_complexity.ps.Z

Keywords: computational complexity theory; education.

This note describes and discusses textbooks, monographs, and collections that are excellent resources from which to teach courses on computational complexity theory.

รูปที่ 3. ตัวอย่างของเอกสาร

5.1 ขั้นตอนการเตรียมเอกสาร

ขั้นตอนในการเตรียมเอกสาร เพื่อให้สามารถนำเข้าสู่โครงข่ายประสาทเทียมแบบเพิ่มพลิไฟค์อาร์ทแม็พได้ มีขั้นตอนดังนี้

1.แยกเอาเฉพาะข้อความที่อยู่ในเครื่องหมาย " " เป็นข้อมูลของ Title และ ข้อความในส่วนของ Abstract หลังกลุ่มข้อความของ Keywords เป็นข้อมูลของ Keyword ของทุกๆ เอกสารในแต่ละกลุ่ม ซึ่งจะได้ข้อมูล Title และ Keyword ดังรูปที่ 4 และรูปที่ 5 ตามลำดับ

2.นำตัวข้อมูลที่ได้จาก Keyword ไปหาคำสำคัญ (Keyword) ด้วยโปรแกรม Copernic Summarizer โดยการหาคำสำคัญของเอกสารแต่ละชุด ได้กำหนดจำนวนคำไว้ที่ 20 คำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.นำข้อมูลในส่วนของ Title ไปตัด Stop word โดยใช้มาตรฐาน[8]

4.นำข้อความทั้งใน Title และ Keyword ไปหา Stemming Word โดยใช้ Porter's Algorithm [9]

Teaching Computational Complexity: Resources to Treasure

รูปที่ 4. ตัวอย่างข้อมูลของ Title

This note describes and discusses textbooks, monographs, and collections that are excellent resources from which to teach courses on computational complexity theory

รูปที่ 5. ตัวอย่างข้อมูลของ Keyword

หลังจากนั้นนำข้อมูลที่ได้จากการขั้นตอนเตรียมเอกสาร มาแบ่งเป็นข้อมูลสำหรับเรียนรู้ (Training) 70% และข้อมูลสำหรับทดสอบ (Testing) 30% ซึ่งรายละเอียดของจำนวนข้อมูลแต่ละกลุ่มดังตารางที่ 2

ตารางที่ 2. ข้อมูลสำหรับเรียนรู้และทดสอบ แต่ละกลุ่ม

ชื่อกลุ่ม	จำนวนเอกสารที่ใช้	
	เรียนรู้	ทดสอบ
AI	73	30
Robotics and Vision	52	22
Systems	121	51
Theory	89	38
รวมทั้งหมด	335	141

5.2 การวัดประสิทธิภาพของอัลกอริทึม

การวัดประสิทธิภาพการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟค์อาร์ทแม็พ ได้ใช้ตัววัดอัตราความถูกต้อง (Accuracy Rate) ตามสมการที่ 17

$$r = [(\sum_{i=1}^{n} doc_i)/n] * 100 \qquad (17)$$

เมื่อ c คือ จำนวนของข้อมูลที่ตอบถูกต้อง

n คือ จำนวนของข้อมูลทั้งหมด

5.3 ผลการทดลอง

นำข้อมูลที่ใช้ในการเรียนรู้เข้าสู่โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟค์อาร์ทแม็พ เพื่อให้ระบบเรียนรู้ หลังจากนั้นนำข้อมูลที่ใช้ในการทดสอบ มาทดสอบการทำงานของระบบว่าถูกต้องมากน้อยเพียงใด ซึ่งผลการทดลองแสดงในตารางที่ 3

ตารางที่ 3. ผลการทดลอง

จำนวนเอกสารที่ใช้		อัตราความถูกต้อง (Accuracy Rate)
เรียนรู้	ทดสอบ	
335	141	92.20 %

จากผลการทดลอง อัตราความถูกต้องในการจำแนกประเภทเอกสารของโครงข่ายประโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟค์อาร์ทแม็พ ให้ผลความถูกต้อง 92.20%

5. สรุปผล

งานวิจัยนี้นำเสนอการจำแนกประเภทเอกสารโดยใช้โครงข่ายประสาทเทียม ซึ่งเป็นวิธีการหนึ่งของการทำ Text Mining

จากการวัดประสิทธิภาพการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟค์อาร์ทแม็พ ได้ผลเป็นที่น่าพอใจในระดับหนึ่ง

การดำเนินการในลำดับต่อไป นำข้อมูลมาตรฐานชุดอื่นๆ เช่น Reuters-Top10 [10] และ K-dataset [11] ซึ่งเป็นชุดข้อมูลที่ใช้ในการทดสอบการจำแนกประเภทเอกสาร และการจัดกลุ่มเอกสาร มาทดสอบการทำงานของโครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟค์อาร์ทแม็พ เพื่อให้โครงข่ายประสาทเทียมนี้ มีความน่าเชื่อถือยิ่งขึ้น

6. เอกสารอ้างอิง

[1] Salton, G., Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer, Addison Wesley Publishing Company, New York, 1989.

[2] K.C. Gowda, and E. Diday, "Symbolic Clustering Using a New Similarity Measure," IEEE Trans. on System Man, and Cybernetics, Vol. 22, 1992, pp. 368-378.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[3] Y. El-Sonbaty, and M.A. Ismail, "Fuzzy Clustering for Symbolic Data," IEEE Trans.on Fuzzy Systems, Vol 6, 1998, pp. 195-204.
[4] T. Kasuba, "Simplified Fuzzy ARTMAP," AI Expert, Vol. 8, 1993, pp. 18-25.
[5] W. Kreesuradej, N. Chantasut, and W. Kruaklai, "Clustering Text Data Using Text ART Neural Network," The WSEAS Trans. on Systems, Vol. 3, pp. 200-205, 2004.
[6] http://www.cs.rochester.edu/trs/.
[7] T. Li, S. Zhu, and M. Ogihara, "Efficient Multi-way Text Categorization via Generalized Discriminant Analysis," Proceedings of the Twelfth International Conference on Information and Knowledge Management, 2003, pp. 317-324.
[8] http://www-fog.bio.unipd.it/waishelp/stoplist.html.
[9] http://www.tartarus.org/~martin/PorterStemmer/.
[10] http://www.reeltwo.com/datasets.html.
[11] ftp://ftp.cs.umn.edu/dept/users/boley/PDDPdata/.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล นางพวงผกา คุณาสิทธิ์

วัน เดือน ปีเกิด 7 พฤศจิกายน 2516 ที่จังหวัดขอนแก่น

ที่อยู่ 80/56 ถนนนครสวรรค์ ตำบลตลาด อำเภอเมือง จังหวัดมหาสารคาม 44000

ประวัติการศึกษา ปีพุทธศักราช 2538 สำเร็จการศึกษา วิทยาศาสตรบัณฑิต (วิทยาการ-คอมพิวเตอร์) จากสถาบันราชภัฏลำปาง

ประสบการณ์ทำงาน

ปีพุทธศักราช 2538-ปัจจุบัน อาจารย์ สถาบันราชภัฏมหาสารคาม (มหาวิทยาลัยราชภัฏมหาสารคาม)

ผลงานวิจัย

- Worapoj Kreesuradej and Puangpaka Kunasit, "Text Processing Simplified ARTMAP Neural Network," WSEAS Transactions on Information Science and Applications, Issue 2, Volume 2, February, 2005, pp.116-120.
- พวงผกา คุณาสิทธิ์ และ วรพจน์ กรีสุระเดช, "การจำแนกประเภทเอกสารโดยใช้โครงข่ายประสาทเทียมแบบเท็กซิมพลิไฟด์อาร์ทแม็พ (Text Categorization using a Text Simplified ARTMAP Neural Network)," Data Mining : Proceeding of the National Computer Science and Engineering Conference (NCSEC 2005), Bangkok, October, 2005.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้