**การปิดบังข้อมูลที่ระบุตัวบุคคล**

**ณัฏฐณิชา ชัยศิริพานิช1 และ ประวิตรานันท์ บุตรโพธิ์2**

*1คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ*

*2คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ*

*Emails: 60070135@it.kmitl.ac.th, 60070148@it.kmitl.ac.th*

**บทคัดย่อ**

ในปัจจุบันเทคโนโลยีส่งผลให้การดำเนินชีวิตสะดวกขึ้น ซึ่งก็ส่งผลต่อพฤติกรรมการทำธุรกรรมกับทางธนาคารเช่นกัน กล่าวคือ ลูกค้ามักดำเนินการทำธุรกรรมออนไลน์ หรือดำเนินการทำธุรกรรมผ่านทางศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) ซึ่งรายละเอียดต่าง ๆ ที่ลูกค้าดำเนินการทำธุรกรรมผ่านศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์นั้นมีรายละเอียดข้อมูลส่วนบุคคลค่อนข้างมาก และทางธนาคารได้มีการบันทึกเสียงเพื่อใช้เป็นหลักฐานการระบุตัวตนลูกค้า และนำข้อมูลไปวิเคราะห์เพื่อพัฒนาประสิทธิภาพขององค์กร แต่ในกระบวนการวิเคราะห์นั้น หากยังมีข้อมูลส่วนบุคคลของลูกค้าอยู่ อาจส่งผลให้ผู้ที่ประสงค์ร้ายสามารถลักลอบนำข้อมูลส่วนบุคคลของลูกค้าไปเผยแพร่โดยไม่ได้รับอนุญาตได้ ดังนั้น การรักษาความลับและข้อมูลส่วนตัวของลูกค้าเป็นเรื่องที่ทางธุรกิจต้องพึงตระหนักเป็นอย่างมาก

ทางผู้จัดทำจึงได้สร้างโครงงานฉบับนี้ขึ้นโดยมีวัตถุประสงค์เพื่อปิดบังการสนทนาที่ประกอบด้วยข้อมูลส่วนบุคคลทั้งของลูกค้าและพนักงานผู้ให้บริการ โดยมีการสร้างแบบจำลองที่สามารถแปลงเสียงพูดให้อยู่ในรูปแบบของข้อความ และทำการตรวจจับรูปแบบของข้อมูลที่เป็นส่วนบุคคล จากนั้นทำการจับคู่เวลาที่มีข้อมูลส่วนบุคคล และปกปิดเสียงเหล่านั้นออกไป เพื่อที่องค์กรสามารถนำผลลัพธ์ที่ได้ไปวิเคราะห์และพัฒนาประสิทธิภาพทางธุรกิจ

***คำสำคัญ –*** *ข้อมูลส่วนบุคคล; ศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center); การประมวลผลภาษาธรรมชาติ (Natural Language Processing); นิพจน์ระบุนาม (Named Entities); การรู้จำเสียงพูด (Speech Recognition)*

**1. บทนำ**

ปัจจุบันการละเมิดข้อมูลส่วนบุคคลนั้นเกิดขึ้นได้หลายรูปแบบ ซึ่งการละเมิดข้อมูลส่วนบุคคลจากการบันทึกบทสนทนาการทำธุรกรรมกับทางธนาคารก็ถือเป็นหนึ่งในปัญหาการละเมิดสิทธิส่วนบุคคลเช่นกัน ทางผู้จัดทำได้เล็งเห็นถึงความสำคัญของการรักษาข้อมูลส่วนบุคคลของลูกค้าในการทำธุรกรรมกับทางธนาคารผ่านศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ (Call Center) โดยจะมีการทำการตรวจจับการสนทนาบางส่วนกับทางศูนย์บริการข้อมูลลูกค้าทางโทรศัพท์ โดยเฉพาะส่วนที่เป็นข้อมูลส่วนบุคคลของลูกค้า เช่น ชื่อ - นามสกุล วันเกิด เบอร์โทรศัพท์ เลขที่บัญชี และเลขหน้าบัตรเครดิต หรือเดบิต ก่อนจะนำข้อมูลการสนทนาเหล่านั้นส่งต่อไปสู่กระบวนการวิเคราะห์เพื่อใช้ในกระบวนการทางธุรกิจ โดยทางผู้จัดทำจะดำเนินการแปลงการสนทนานั้นให้อยู่ในรูปแบบข้อความ ตรวจจับเนื้อหาของข้อความว่าคำใดมีรูปแบบที่เป็นข้อมูลที่สำคัญหรือข้อมูลส่วนบุคคล จากนั้นดำเนินการจับคู่คำกับเวลาในไฟล์บันทึกเสียง และดำเนินการปกปิดข้อความในส่วนนั้นออกไป

**2. แนวคิด และเทคโนโลยีที่เกี่ยวข้อง**

**2.1 แนวคิดที่เกี่ยวข้อง**

*2.1.1 สิทธิความเป็นอยู่ส่วนบุคคล*

มีการบัญญัติรับรองสิทธิดังกล่าวในรัฐธรรมนูญแห่งราชอาณาจักรไทย พ.ศ. 2560 มาตรา 32 ว่า “สิทธิของบุคคลในครอบครัว เกียรติยศ ชื่อเสียง ตลอดจนความเป็นอยู่ส่วนบุคคล ย่อมได้รับความคุ้มครอง การกล่าวหรือไขข่าวแพร่หลายซึ่งข้อความหรือภาพไม่ว่าด้วยวิธีใดไปยังสาธารณชนอันเป็นการละเมิดหรือกระทบถึงสิทธิของบุคคลในครอบครัว เกียรติยศ ชื่อเสียง หรือความเป็นอยู่ส่วนบุคคล จะกระทำมิได้ เว้นแต่กรณีที่เป็น ประโยชน์ต่อสาธารณะ บุคคลย่อมมีสิทธิได้รับความคุ้มครองจากการแสวงประโยชน์โดยมิชอบจากข้อมูลส่วนบุคคลที่เกี่ยวกับตน ทั้งนี้ ตามที่กฎหมายบัญญัติ” [1]

**2.2 เทคโนโลยีที่เกี่ยวข้อง**

*2.2.1 การทำเหมืองข้อมูล (Data Mining)*

เป็นกระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลชุดนั้น โดยอาศัยหลักการทางสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ [2]

*2.2.2 การรู้จำเสียงพูด (Speech Recognition)*

เป็นสิ่งที่ช่วยให้โปรแกรมสามารถประมวลผลคำพูดของมนุษย์ให้อยู่ในรูปแบบลายลักษณ์อักษร โดยเน้นที่การแปลงเสียงพูดจากรูปแบบคำพูดเป็นข้อความ [3]

*2.2.3 Google Speech Recognition*

พิมพ์อธิบายรายละเอียดสั้น ๆ พร้อม ref

*2.2.4 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)*

การประมวลผลภาษาธรรมชาติ คือ หนึ่งในสาขาของวิทยาศาสตร์คอมพิวเตอร์ที่เกี่ยวข้องกับปัญญาประดิษฐ์ (Artificial Intelligence) และภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistics) เป็นศาสตร์ที่ศึกษาเกี่ยวกับการทำให้คอมพิวเตอร์สามารถสื่อสารโต้ตอบด้วยภาษาของมนุษย์ และทำให้คอมพิวเตอร์เข้าใจภาษามนุษย์มากขึ้น เช่น Siri, Google Assistant และ Alexa [4]

*2.2.5 Stanford Named Entity Recognizer (Stanford NER)*

เป็นการประยุกต์ใช้จากภาษาจาวา (Java) สำหรับการรู้จำนิพจน์ระบุนาม (Named Entity Recognizer: NER) ซึ่งเป็นการจัดประเภทของคำในข้อความ เช่น ชื่อสิ่งของ ชื่อบุคคล และบริษัท เป็นการกำหนดโครงสร้างการสกัดคุณสมบัติที่เหมาะสมสำหรับการรู้จำนิพจน์ระบุนาม (Named Entity Recognition: NER) [5]

*2.2.6 Natural Language Toolkit (NLTK)*

เป็นแพลตฟอร์มที่นิยมในโปรแกรมภาษาไพทอน (Python) เพื่อทำงานกับข้อมูลภาษาของมนุษย์ พร้อมกับชุดของไลบรารีที่ช่วยในการประมวลข้อความ แบ่งประเภทของคำ (Classification) การแบ่งโทเค็นของคำ (Tokenization) การตัดคำ (Stemming) การติดแท็กคำ (Tagging) และการแยกวิเคราะห์คำ (Parsing) [6]

*2.2.7 spaCy*

เป็นไลบรารีสำหรับการทำการระมวลผลภาษาธรรมชาติขั้นสูงในภาษาไพทอน (Python) โดยที่ spaCy ถูกออกแบบมาสำหรับการประยุกต์ใช้งานจริง และช่วยสร้างแอปพลิเคชันที่สามารถประมวลผล และทำความเข้าใจข้อความจำนวนมาก สามารถใช้ในการดำเนินการสกัดข้อมูล (Information Extraction) หรือระบบการทำความเข้าใจภาษาธรรมชาติเพื่อดำเนินการประมวลผลข้อความล่วงหน้าสำหรับการเรียนรู้เชิงลึก (Deep Learning) [7]

*2.2.8 Regular Expressions*

เป็นสัญลักษณ์ที่ใช้ระบุชุดของอักขระตัวอักษร เมื่อชุดของอักขระตัวอักษรที่เฉพาะเจาะจงนั้นอยู่ในชุดอักขระตัวอักษรที่มีการกำหนดให้เป็น Regular Expressions โดยทั่วไปแล้วจะใช้สัญลักษณ์ “\*”, “+”, “?”, “()” และ “|” ในการกำหนดเงื่อนไขของชุดตัวอักษร [8]

*2.2.9 …*

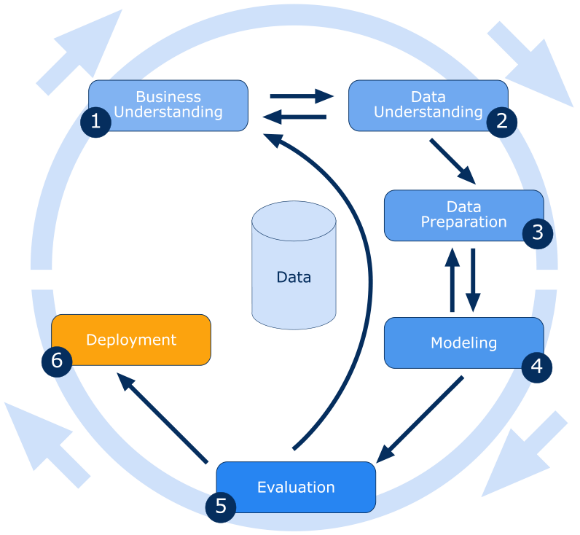
*2.2.10 Jaccard’s Coefficient Similarity*

เป็นสถิติประยุกต์แนวคิดในทฤษฎีเซตเพื่อนำมาใช้เปรียบเทียบความคล้ายคลึงและความหลากหลายของกลุ่มตัวอย่าง แนวคิดของค่าสัมประสิทธิ์ Jaccard’s Coefficient Similarity คือ การวัดค่าความคล้ายคลึงระหว่างกลุ่มประชากร 2 กลุ่ม โดยคำนวณจากขนาดของประชากรที่ทั้งสองกลุ่มมีตัวอย่างร่วมกัน (อินเตอร์เซกชันในทฤษฎีเซต) หารด้วยขนาดของประชากรทั้งหมดจากทั้งสองกลุ่มตัวอย่าง (ยูเนียนในทฤษฎีเซต) [10] ดังสูตรที่ 1

(1)

**3. ขั้นตอน และวิธีการดำเนินงานวิจัย**

**3.1 กระบวนการการทำเหมืองข้อมูล (Data Mining Process)**



**3.1.1 การทำความเข้าใจธุรกิจ (Business Understanding)**

เมื่อเข้าสู่ยุคที่มีการแข่งขันสูง หลาย ๆ ธนาคารเริ่มนำเทคโนโลยีต่าง ๆ เข้ามาประยุกต์ใช้ในการให้บริการเพื่อเพิ่มความสะดวกสบายต่อลูกค้า รวมถึงต้องนำความพึงพอใจจากลูกค้า หรือปัญหาต่าง ๆ ทั้งทางออนไลน์ และการสนทนาผ่านโทรศัพท์ มาดำเนินการวิเคราะห์เพื่อเพิ่มประสิทธิภาพขององค์กรให้ดีที่สุด ข้อมูลส่วนบุคคลของลูกค้าจึงจำเป็นต้องมีการปกปิดก่อนจะเข้าสู่กระบวนการวิเคราะห์นั้น เพื่อป้องกันการละเมิดสิทธิส่วนบุคคลของลูกค้า และเพิ่มความน่าเชื่อถือขององค์กร

**3.1.2 การทำความเข้าใจข้อมูล (Data Understanding)**

เมื่อเข้าสู่ยุคที่มีการแข่งขันสูง หลาย ๆ ธนาคารเริ่มนำเทคโนโลยีต่าง ๆ เข้ามาประยุกต์ใช้ในการให้บริการเพื่อเพิ่มความสะดวกสบายต่อลูกค้า รวมถึงต้องนำความพึงพอใจจากลูกค้า หรือปัญหาต่าง ๆ ทั้งทางออนไลน์ และการสนทนาผ่านโทรศัพท์ มาดำเนินการวิเคราะห์เพื่อเพิ่มประสิทธิภาพขององค์กรให้ดีที่สุด ข้อมูลส่วนบุคคลของลูกค้าจึงจำเป็นต้องมีการปกปิดก่อนจะเข้าสู่กระบวนการวิเคราะห์นั้น เพื่อป้องกันการละเมิดสิทธิส่วนบุคคลของลูกค้า และเพิ่มความน่าเชื่อถือขององค์กร

**4. เนื้อเรื่อง**

เพื่อให้บทความที่ตีพิมพ์ในวารสารเทคโนโลยีสารสนเทศลาดกระบังเป็นไปในรูปแบบเดียวกัน โปรดใช้รูปแบบและขนาดอักษรดังนี้ในการพิมพ์บทความภาษาไทย ในส่วนของเนื้อหาหลักใช้รูปแบบอักษร TH Sarabun New ขนาด 13 พอยท์ ในส่วนของคำบรรยายภาพและตาราง ใช้รูปแบบอักษร TH Baijam ขนาด 12 พอยท์

สำหรับย่อหน้าแรกภายใต้หัวข้อ **ไม่ต้อง**เว้นระยะหนึ่งแท็บจากขอบซ้ายของคอลัมน์ ในย่อหน้าต่อไปภายใต้หัวข้อจึงเว้นระยะหนึ่งแท็บจากขอบซ้ายของคอลัมน์ (ดังเช่นที่แสดงในหัวข้อที่ 4 นี้)

**5. หัวข้อหลัก**

หัวข้อหลัก ตัวอย่างเช่น “1. บทนำ” ต้องพิมพ์ด้วยอักษร TH Sarabun New ขนาด 16 พอยท์ แบบตัวหนา จัดให้อยู่ชิดขอบซ้ายของบรรทัดในคอลัมน์ ใส่เครื่องหมายจุด (“.”) หลังจากตัวเลขกำกับหัวข้อหลัก เว้นระยะ 12 พอยท์ระหว่างหัวข้อหลักกับบรรทัดก่อนหน้า และเว้นระยะ 3 พอยท์ระหว่างหัวข้อหลักกับบรรทัดใต้หัวข้อหลัก หลังจากหัวข้อหลักแล้วอาจแบ่งหัวข้อย่อยภายในหัวข้อหลักได้ ซึ่งการจัดรูปแบบของหัวข้อย่อยนั้นจะกล่าวถึงในลำดับถัดไป

**5.1. หัวข้อย่อย**

หัวข้อย่อยให้ใช้อักษร TH Sarabun New ขนาด 14 พอยท์ แบบตัวหนา จัดชิดขอบซ้ายของบรรทัดในคอลัมน์ เว้นระยะ 12 พอยท์ระหว่างหัวข้อย่อยกับบรรทัดก่อนหน้า และเว้นระยะ 3 พอยท์ระหว่างหัวข้อย่อยกับบรรทัด ใต้หัวข้อย่อย

*5.1.1. หัวข้อย่อยลำดับที่สอง*

ควรหลีกเลี่ยงการใช้หัวข้อย่อยลำดับที่สอง อย่างเช่นที่แสดงในย่อหน้านี้ อย่างไรก็ตามหากมีความจำเป็นต้องใช้จริงๆให้พิมพ์ชื่อหัวข้อย่อยลำดับที่สองด้วยตัวอักษร TH Sarabun New ขนาด 13 พอยท์ แบบตัวเอียง จัดชิดขอบซ้ายของบรรทัดในคอลัมน์ เว้นระยะ 6 พอยท์ระหว่างหัวข้อย่อยลำดับที่สองกับบรรทัดก่อนหน้า และเว้นระยะ 3 พอยท์ระหว่างหัวข้อย่อยลำดับที่สองกับบรรทัดใต้หัวข้อย่อยลำดับที่สอง

ตัวอย่างของรูปภาพ

**รูปที่ 1.** คำบรรยายภาพต้องอยู่ใต้ภาพ

**ตารางที่ 1.** คำบรรยายตารางต้องอยู่เหนือตาราง

|  |  |  |
| --- | --- | --- |
| Name | Age | Weight (kg) |
| Alice | 12 | 30 |
| Bob | 13 | 28 |
| Charlie | 14 | 35 |

**6. รูปภาพและตาราง**

รูปภาพและตารางจะต้องจัดให้อยู่ภายในพื้นที่การพิมพ์ที่กำหนด รูปภาพ และตารางสามารถมีความกว้างมากกว่าหนึ่งคอลัมน์ได้ ถ้าหากสิ่งเหล่านี้มีความกว้างมากกว่าหนึ่งคอลัมน์ต้องจัดสิ่งเหล่านี้ให้อยู่กึ่งกลางของความกว้างของคอลัมน์ทั้งสอง

รูปภาพและตารางต้องมีคำบรรยายและหมายเลขกำกับเสมอ เขียนคำบรรยายรูปภาพไว้ใต้รูปภาพ เขียนคำบรรยายตารางไว้ที่ส่วนบนของตาราง ในส่วนของคำบรรยายภาพและตาราง ใช้รูปแบบอักษร TH Baijam ขนาด 12 พอยท์

**7. สมการ**

ควรหลีกเลี่ยงการใช้สมการในการเขียนบทความวิชาการ ในหนึ่งบทความไม่ควรมีสมการเกินกว่า 3 (สาม) สมการ เขียนหมายเลขกำกับสมการทุกสมการที่ปรากฏในบทความ โดยใส่หมายเลขสมการในเครื่องหมายวงเล็บ จัดหมายเลขสมการชิดขอบขวาของคอลัมน์ (ดังที่แสดงด้านล่างนี้)

*a* + *b = c* + 2 (1)

เว้นระยะหนึ่งบรรทัดก่อนสมการและเว้นระยะหนึ่งบรรทัดหลังสมการ จัดสมการให้อยู่กึ่งกลางบรรทัดในคอลัมน์ ขนาดตัวอักษรและสัญลักษณ์ที่ใช้ในการเขียนสมการจะต้องมีขนาดเทียบเท่ากับตัวอักษร TH Sarabun New ขนาด 13 พอยท์

**8. หมายเลขหน้า**

โปรดอย่าใส่หมายเลขหน้าลงในบทความ หมายเลขหน้าจะถูกใส่หลังจากที่บทความจะถูกบรรจุในวารสารเทคโนโลยีสารสนเทศลาดกระบังแล้วเท่านั้น บทความภาษาไทยต้องมีความยาวไม่เกิน 15 (สิบห้า) หน้า สำหรับบทความภาษาอังกฤษผู้เขียนต้องปฏิบัติตามคำแนะนำที่ให้ไว้ในวิธีการจัดรูปแบบบทความภาษาอังกฤษ (โปรดดูใน Template ของบทความภาษาอังกฤษที่ให้ไว้ในอีกไฟล์หนึ่ง)

**9. เชิงอรรถ**

เพื่อช่วยให้ผู้อ่านสามารถอ่านบทความได้อย่างต่อเนื่องโดยไม่สะดุด โปรดพยายามหลีกเลี่ยงการใช้เชิงอรรถ (Footnote) ถ้าหากมีความจำเป็นต้องใช้เชิงอรรถ ให้วางเชิงอรรถไว้ที่ส่วนท้ายของคอลัมน์ในหน้าที่อ้างอิงเชิงอรรถนั้น ใช้ตัวอักษร TH Sarabun New ขนาด 12 พอยท์ ในการพิมพ์เชิงอรรถ

**10. รูปแบบการอ้างอิง**

ใส่เอกสารอ้างอิงทั้งหมดที่ส่วนท้ายสุดของบทความโดยให้อยู่ภายใต้หัวข้อ “เอกสารอ้างอิง” ไม่ต้องใส่หมายเลขหัวข้อ ภายใต้หัวข้อเอกสารอ้างอิงให้จัดลำดับเอกสารตามลำดับการอ้างอิงถึงในบทความโดยใช้ระบบการอ้างอิงแบบตัวเลขเท่านั้น ตัวอย่าง**วิธีการเขียนอ้างอิงเอกสารภาษาไทย**ประเภทต่างๆนั้น สามารถหาดูได้จากหน้าเว็บไซต์ของหอสมุดแห่งชาติ

(<http://www.nlt.go.th/cm/Bibliography.htm>)

การเขียนบรรณานุกรมให้ใช้ตัวเลขอารบิกแบบสากลเท่านั้น **อย่าใช้เลขไทย** เมื่อมีการอ้างอิงถึงเอกสารในบทความให้อ้างอิงโดยใช้หมายเลขในเครื่องหมายก้ามปู ดังที่แสดงไว้ในส่วนท้ายของประโยคนี้ [1] ห้ามใช้รูปแบบการอ้างอิงแบบอื่นๆ เช่น นาม-ปี โดยเด็ดขาด **การอ้างอิงเอกสารภาษาอังกฤษ** ให้ใช้การอ้างอิงตามรูปแบบขอ IEEE (http://www.ieee.org/documents/ieeecitationref.pdf) ตัวอย่างวิธีการอ้างอิงเอกสารที่ใช้บ่อยๆ ได้แสดงไว้ในหัวข้อเอกสารอ้างอิงของเอกสารแนะนำนี้แล้ว

**เอกสารอ้างอิง**

[1] A. B. Green, C. D. Black, and E. F. White, “Article Title,” *Journal*, vol. 100, no. 1, pp. 1-10, Dec. 2000.

[2] C. D. Black, A. B. Green, and E. F. White, *Book Title*, 3rd ed. New York: McGraw-Hill, 2001.

[3] สมชาย สกุลดี. “ชื่อบทความ”. **ชื่อวารสาร** ปีที่ 10, ฉบับที่ 2 (10 กุมภาพันธ์ 2553). หน้า 10-15.

[4] สมหญิง เจริญดี. **ชื่อหนังสือ.** พิมพ์ครั้งที่ 2. กรุงเทพฯ: สำนักพิมพ์เจริญทัศน์, 2553.

[5] J. K. Pink, “Article Title,” in *Proc. International Conference on Green Computing*, Paris, France, Jan. 2012, pp. 50-55.

[6] สมศักดิ์ มงคล. “**ชื่อวิทยานิพนธ์**”. (วิทยานิพนธ์ปริญญามหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร, 2543).

[7] สมศรี บุญมาก. “ชื่อบทความ”. ชื่อการประชุมวิชาการ. 2549. หน้า 45-48.

[8] R. Good. (2011, Feb 10). Computers (2nd ed.) [Online]. Available: http://www.computers.com

[9] J. Better, “How to Write,” Ph.D. dissertation, Dept. Elect. Eng., Amazing University, Cambridge, MA, 2003.