**วิธี Soundex ของเทพพิทักษ์ การุญบุญญานันท์**

          ในปี  พ.ศ.  ๒๕๔๐  เทพพิทักษ์  การุญบุญ-ญานันท์  และคณะ  ได้เสนอวิธีการใหม่ในการแก้ปัญหาซาวน์เด็กซ์ภาษาไทย  โดยใช้  Nondeterministic  Finite  Automation  ในการแปลงรหัสเครื่องจักรจำลองดังกล่าว  จึงทำให้กฎอยู่ในรูปที่ง่ายขึ้น  มีรูปแบบตายตัว  และมีความเป็นไปได้ที่จะเขียนกฎให้ครอบคลุมหลักอักขรวิธีทุกข้อการแยกพยางค์และลงรหัสคำอ่านจึงมีความถูกต้องและละเอียดขึ้นมาก

          แบบจำลองดังกล่าวจะค้นหาและลงรหัสคำอ่านที่สามารถเป็นไปได้ทั้งหมด  ซึ่งจะมีทั้งคำที่อ่านถูกและผิด  แต่คำอ่านทุกคำถูกต้องตามหลักอักขรวิธี  เช่น  กมล  จะอ่าน  ได้ทั้ง  กะมนกะ - มะ - ละ   กะ - มน - ละ    กอ - มอ - ลอฯลฯ  คำอ่านทั้งหมดจะถูกกำจัดรายละเอียดเล็กน้อย  เช่นตัวควบกล้ำ  เสียงวรรณยุกต์  ความสั้นยาวของเสียงสระ  ก่อนที่จะถูกลงรหัส  รหัสจะถูกค้นหาระหว่างแปลงอย่างรวดเร็วในฐานข้อมูลด้วยโครงสร้างดัชนีแบบต้นไม้  คำอ่านที่พบว่าไม่มีกิ่งในต้นไม้  (กล่าวคือ  ไม่มีคำที่ออกเสียงเช่นนั้นในฐานข้อมูล)  จะถูกหยุดแปลงแต่เนิ่นๆ

          อย่างไรก็ตาม  ระบบนี้ถูกออกแบบมาเพื่อการใช้ซาวน์เด็กซ์ในการแก้ไขตัวสะกดเท่านั้น  ยังไม่เหมาะที่จะใช้กับฐานข้อมูลทั่วไป

**ปัญหา**

1. คำบางคำออกเสียงต่างกัน แต่รหัส Soundex coed ที่ได้เหมือนกัน ในขณะที่บางคำออกเสียงเหมือนกัน แต่รหัส Soundex code ที่ได้แตกต่างกัน เช่น
   1. สีส้ม, สีดำ (มี Soundex code เหมือนกัน)
   2. บุญญา / บุณยา (มี Soundex code ต่างกัน)
2. คำที่เป็นคำพ้องเสียง บางครั้งมีรหัส Soundex code ที่ต่างกัน เช่น
   1. เก้า / ก้าว
   2. จร / จอน
   3. ได้ / ด้าย
3. คำที่มีพยัญชนะตัวเดียวกัน แต่ออกเสียงต่างกัน เช่น
   1. ทราบ / ทรุด / พุทรา / จันทรา / ทรานซิสเตอร์ / ทรอมโบน
   2. บัณฑิต / มณฑป / บัณเฑาะว์ / มณโฑ / มณฑล / จุฑา

**เป้าหมายของระบบ**

* ตัดวรรณยุกต์ออก
* รหัส Soundex มีคุณลักษณะการออกเสียงต่างๆดังนี้:
* เสียงพยัญชนะต้น

|  |  |  |
| --- | --- | --- |
| 1. /ก/ ก | 8. /ด/ ด ฎ | 15.   /ฟ/    ฟ ฝ |
| 2. /ค/ ข ค ฆ (ฃ ฅ) | 9. /ต/ ต ฏ | 16.   /ม/    ม |
| 3. /ง/ ง | 10. /ท/ ท ฐ ถ ฑ ฒ ธ | 17.   /ร/    ร ล ฬ ฤ ฦ |
| 4. /จ/ จ | 11. /น/ น ณ | 18.   /ว/    ว |
| 5. /ช/ ช ฉ ฌ | 12. /บ/ บ | 19.   /ห/    ฮ ห |
| 6. /ซ/ ซ ศ ส ษ | 13. /ป/ ป | 20.   /อ/    อ |
| 7. /ย/ ญ ย | 14. /พ/ ผ พ ภ |

* เสียงสระ
  + ออ แทนสัญลักษณ์ Soundex code ‘c’
  + อะ/อา แทนสัญลักษณ์ Soundex code ‘a’
  + อิ/อี แทนสัญลักษณ์ Soundex code ‘i’
  + เอะ/เอ แทนสัญลักษณ์ Soundex code ‘e’
  + อุ/อู แทนสัญลักษณ์ Soundex code ‘u’
  + โอะ/โอ แทนสัญลักษณ์ Soundex code ‘o’
* เสียงพยัญชนะตัวสะกด

|  |  |  |
| --- | --- | --- |
| 1. ไม่มีตัวสะกด | แม่ ก กา | ฮ |
| 2. ก ข ค ฆ | แม่ กก | ก |
| 3. จ ด ต ถ ท ธ ฎ ฏ ฑ ฒ ช ซ ศ ษ ส | แม่ กด | ด |
| 4. บ ป พ ภ ฟ | แม่ กบ | บ |
| 5. น ณ ญ ร ล ฬ | แม่ กน | น |
| 6. ง | แม่ กง | ง |
| 7. ม อำ | แม่ กม | ม |
| 8. ย ไอ ใอ | แม่ เกย | ย |
| 9. ว เอา | แม่ เกอว | ว |

**การเข้ารหัส Soundex code**

* ใช้หลักการ Nondeterministic Finite Automaton เพื่อให้ได้ output
* กฎจะถูกเขียนแยกกันในนิพจน์ทั่วไปพร้อมเอาท์พุท \*\*แปลไม่ได้
* การสร้างโค้ด : เป็นการแปลงแบบ many–to–many แทนที่จะแปลงด้วย many–to–one เช่นงานวิจัยก่อนหน้า

**โมเดลสำหรับการเข้ารหัส Soundex**

สมมติว่าพยางค์ที่ถูกต้องคือ :

***\*\*ini-cons คือ พยัญชนะต้น , fnl-cons คือตัวสะกด***

* ini-cons fnl-cons เช่น กล, ต้น
* ini-cons 'า' [fnl-cons] เช่น กลาย, สาป
* 'เ' ini-cons 'า' เช่น เขา, เมา
* 'เ' ini-cons [fnl-cons] เช่น เอย, เกม