

บทที่ 2

พจนานุกรมอิเล็กทรอนิกส์

ประวัติการพัฒนาพจนานุกรมอิเล็กทรอนิกส์

(เย็น ภู่วรรณและชัยยงค์ วงศ์ชัยสุวัฒน์, 2535) การพัฒนาพจนานุกรมอิเล็กทรอนิกส์ในต่างประเทศนั้นมีมานานแล้ว ในปี พ.ศ. 2510 US Library of Congress ได้พัฒนาโครงการ MARC (Machine Readable Catalog) โดยได้พัฒนาวิธีการและรูปแบบในการจัดเก็บข้อความไว้ในตัวกลางที่คอมพิวเตอร์จะอ่านได้ และยังได้ทำการจัดเก็บพจนานุกรมชื่อ Webster's Seventh Collegiate Dictionary ซึ่งมีคำอยู่ประมาณ 68,657 คำ ไว้ในตัวกลางแม่เหล็กที่จะทำการค้นหาได้ง่าย

ปี พ.ศ. 2527 ยีนและคณะได้ออกแบบและสร้างพจนานุกรมคำไทย โดยเก็บคำที่ใช้ในชีวิตประจำวันมากกว่า 140,000 คำ มารวมกลุ่มกันเพื่อหาความถี่ของการใช้คำ จากนั้นก็หาคำซ้ำและจัดเรียงตามความถี่ ซึ่งจะทำให้เหลือคำที่จะต้องจัดเก็บในพจนานุกรมประมาณ 6,000 คำ ซึ่งสามารถนำมาใช้ประโยชน์ในการตัดแบ่งแยกคำไทยออกจากประโยค และสามารถใช้ในระบบตรวจสอบตัวสะกดของโปรแกรมเวิร์ดโปรเซสเซอร์ได้

ปี พ.ศ. 2528 สมนึก ศิริโต และคณะได้พัฒนาโปรแกรมแปลภาษาอังกฤษเป็นไทย โดยใช้หลักการ Transfer Approach ด้วยเทคนิค Preference Semantics ในงานวิจัยดังกล่าวมีการสร้างพจนานุกรมขึ้นมาสำหรับงานแปลภาษาขึ้น โครงสร้างของพจนานุกรมยึดหลักการแทนความหมายระหว่างคู่ศัพท์ที่เกี่ยวข้องกับสิ่งแวดล้อมของประโยค

ปี พ.ศ. 2526 H.Tanaka เสนอหลักการของ Active Dictionary ที่ใช้ในงานแปลภาษา ซึ่งเก็บแต่ละ Entry ในโครงสร้างของ list

ผลงานพจนานุกรมอิเล็กทรอนิกส์ที่มีผู้ผลิตคิดค้นขึ้น

ปัจจุบันนี้มีผู้ผลิตพจนานุกรมอิเล็กทรอนิกส์ขึ้นหลายรายด้วยกัน ซึ่งต่างก็มีรายละเอียดการใช้งานแตกต่างกันไป ดังตัวอย่างต่อไปนี้

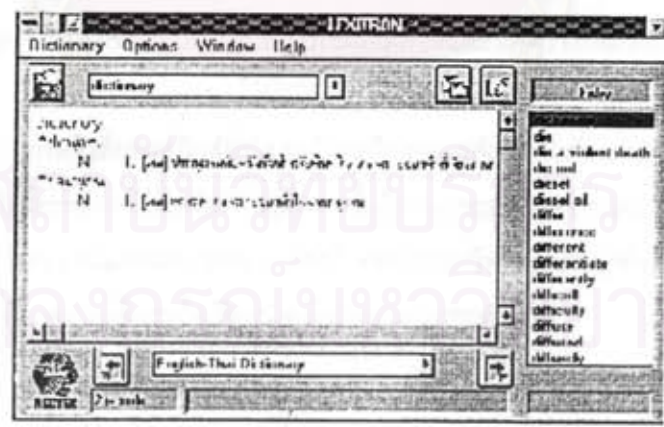
1. LEXITRON เป็นพจนานุกรมอิเล็กทรอนิกส์บนซีดีรอม พัฒนาโดยทีมงานจากห้องปฏิบัติการวิจัย

ภาษาและวิทยาการความรู้ (LINKS) แห่งศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)

LEXITRON เวอร์ชัน 1.1 บรรจุบนซีดี-รอม 1 แผ่นกินเนื้อที่ 50 เมกะไบต์ (MB) บรรจุคำศัพท์ทั้งหมด 22,000 คำ แบ่งเป็นคำศัพท์ภาษาไทย 13,000 คำ คำศัพท์ภาษาอังกฤษ 9,000 คำ และความหมายคำกว่า 20,000 ความหมาย สามารถทำงานได้บนระบบปฏิบัติการเอ็มเอสเอสแอลเวอร์ชัน 3.1 หรือสูงกว่า และระบบปฏิบัติการไมโครซอฟต์วินโดวส์เวอร์ชัน 3.1 ไทยอิตีชั่นหรือสูงกว่านี้

ผู้ใช้งานสามารถค้นหาคำศัพท์ได้ 6 รูปแบบด้วยกัน คือ

- 1.1 ใช้เป็นพจนานุกรมทั่วไป (Thai General Dictionary)
- 1.2 ใช้เป็นพจนานุกรมการใช้ภาษาไทย (Thai Usage Dictionary) แสดงเฉพาะวิธีการใช้คำศัพท์ภาษาไทย ถ้าเป็นคำนามก็จะให้คำลักษณะนามของคำนามนั้น ถ้าเป็นคำกริยาก็นจะแสดงรูปแบบโครงสร้างการใช้คำกริยานั้น
- 1.3 ใช้เป็นพจนานุกรมไทย คำเหมือน-คำตรงข้าม (Thai Synonyms-Antonyms Dictionary)
- 1.4 ใช้เป็นพจนานุกรมไทย-อังกฤษ (Thai-English Dictionary) แสดงคำเทียบเคียงภาษาอังกฤษของคำศัพท์ภาษาไทยนั้นๆ
- 1.5 ใช้เป็นพจนานุกรมอังกฤษ-ไทย (English-Thai Dictionary) ที่แสดงเทียบเคียงภาษาไทยของคำภาษาอังกฤษนั้นๆ โดยจำแนกตามระเบียบวิธีใช้คำอย่างคำไทย
- 1.6 ใช้เป็นพจนานุกรมรวมกลุ่มคำไทย (Thai Word Group Dictionary) ที่แสดงกลุ่มคำที่อยู่ภายใต้ความหมายหลักเดียวกัน ช่วยให้ผู้ใช้เข้าใจความหมายของคำนั้นได้ดียิ่งขึ้น



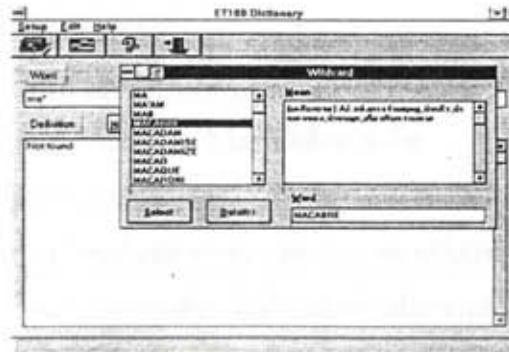
รูปที่ 2-1 การค้นหาคำศัพท์ของเล็กซิตรอน

2. โปรแกรมพจนานุกรม เป็นซอฟต์แวร์ดิจิทัลขั้นนำร่องที่พัฒนาโดยบริษัทไอบีเอส (2535) จำกัด โปรแกรมพจนานุกรมนี้ได้รวบรวมพจนานุกรมหลายประเภทไว้ดังนี้

- 2.1 พจนานุกรมอังกฤษ-ไทย และพจนานุกรมไทย-อังกฤษ
- 2.2 พจนานุกรมสำนวนอังกฤษ (idiom, slang, phrasal verb) โดย อ.เชียวรัชย์ เขียววรเมธ
- 2.3 พจนานุกรมศัพท์คอมพิวเตอร์ โดย ศ.ทักษิณา สนวนานนท์
- 2.4 พจนานุกรมคำเหมือนและคำตรงข้าม
- 2.5 พจนานุกรมไทย ฉบับคำพ้องและคำตรงข้าม

คุณสมบัติของโปรแกรมพจนานุกรม

1. เรียกดูศัพท์ได้ทั้งอังกฤษ-ไทย และไทย-อังกฤษ ดูได้ทั้งศัพท์ทั่วไป ศัพท์คอมพิวเตอร์ คำโดดคำผสม สำนวน คำพ้อง-คำตรงข้าม (Idiom Synonym-Antonym)
 2. สร้างจากพจนานุกรม 7 ฉบับ ใช้เทคโนโลยีบีบอัดข้อมูลจึงใช้ฮาร์ดดิสก์ไม่เกิน 2.5 เมกกะไบท์
 3. มีรายละเอียด คำแปล ความหมาย หน้าที่ของคำ ตัวอย่างประโยค ครบถ้วน ตรงกับต้นฉบับที่พิมพ์จำหน่ายแพร่หลายในปัจจุบัน
 4. ระบบจับคำศัพท์อัตโนมัติ ดูคำแปลได้โดยไม่ต้องป้อนคำ
 5. ระบบวิเคราะห์ปัญหาออกย่อนของคำภาษาไทย
 6. ระบบค้นหาคำคล้ายและคำใกล้เคียง ทั้งภาษาไทยและภาษาอังกฤษ
 7. แสดงคำแปลอย่างเป็นระบบมีสีสันสวยงาม
 8. สามารถใช้งานได้บนระบบปฏิบัติการเอ็มเอสดีเอสและไมโครซอฟต์วินโดวส์
3. ET100 เป็นพจนานุกรมอิเล็กทรอนิกส์ที่พัฒนาโดยบริษัทซอฟต์แวร์ จำกัด สามารถทำงานได้บนระบบปฏิบัติการไมโครซอฟต์วินโดวส์เวอร์ชัน 3.1 ขึ้นไป โดยจะแสดงความหมายของคำศัพท์ภาษาอังกฤษเป็นภาษาไทย และแสดงความหมายของคำศัพท์ภาษาไทยเป็นภาษาอังกฤษ ผู้ใช้สามารถใช้งานพจนานุกรมเพียงอย่างเดียวหรือเรียกผ่านโปรแกรมประยุกต์อื่นได้

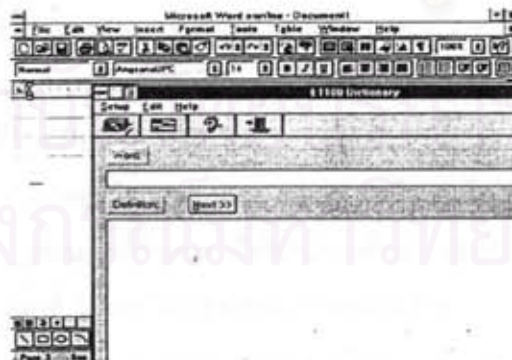


รูปที่ 2-2 แสดงการค้นหาความหมายของคำศัพท์ของ ET100

โดยการป้อนเฉพาะส่วนต้นของคำศัพท์



รูปที่ 2-3 แสดง ET100 สามารถเรียกใช้ได้ในขณะที่ใช้โปรแกรมประยุกต์อื่น



รูปที่ 2-4 แสดงการทำงานของ ET100 ในขณะที่ใช้โปรแกรมไมโครซอฟต์เวิร์ด

4. โปรแกรมพจนานุกรมไทย สอ.เสถบุตร 1.5

- 4.1 ทำงานได้ทั้งบนวินโดวส์ 95 และ วินโดวส์ 3.1X ใช้เนื้อที่บนฮาร์ดดิสค์ประมาณ 5 เมกะไบต์ จุดเด่น คือ มีพจนานุกรม 2 เล่มให้เลือกใช้ คือ
 - 4.1.1 พจนานุกรมอังกฤษเป็นไทย
 - 4.1.2 พจนานุกรมไทยเป็นอังกฤษของ สอ เสถบุตร ฉบับประจำโต๊ะ พ.ศ. 2537
- 4.2. การใช้งานเป็นลักษณะโปรแกรมเดี่ยว ไม่มีการติดตั้งให้ทำงานร่วมกับโปรแกรมประยุกต์อื่น แต่สามารถนำคำศัพท์ผ่านคลิปบอร์ดมาค้นหาได้
- 4.3 การแสดงคำแปลของคำศัพท์ที่ค้นหาพบมี 2 แบบ คือ แบบหนังสือ ซึ่งจะเหมือนกับการเปิดพจนานุกรมจริงๆ และผู้ใช้สามารถเลือกให้แสดงเฉพาะคำศัพท์โดยไม่ต้องแสดงความหมายได้

การค้นหาคำศัพท์

1. การค้นหาแบบมีรูปแบบ พิมพ์คำศัพท์ที่ต้องการหาคำแปลในช่องคำศัพท์ โปรแกรมจะแสดงคำแปลของคำศัพท์ที่ค้นพบ
2. การค้นหาแบบเร็ว ซึ่งคำศัพท์จะถูกเลื่อนขึ้นมาแสดงตามตัวอักษรในคำศัพท์ที่ผู้ใช้พิมพ์เข้าไป และเมื่อกดเอ็นเทอร์ โปรแกรมจะแสดงคำแปลสำหรับค้นหาผ่านช่องคำศัพท์
3. สามารถนำคำศัพท์จากคลิปบอร์ดมาใส่ในช่องคำศัพท์
4. สามารถนำคำศัพท์ที่เคยค้นหาในครั้งก่อน ซึ่งจะเก็บไว้ในรายการขึ้นมาหาคำแปลใหม่
5. สามารถกดเอสเคป เพื่อลบคำศัพท์เดิมที่อยู่ในช่องค้นหาคำศัพท์ได้
6. การค้นหาคำศัพท์ จากการแสดงแบบเลื่อนขึ้นลงได้ โดยผู้ใช้เลือกคำศัพท์ที่ต้องการจากรายการคำศัพท์แล้ว ดับเบิลคลิกที่คำนั้น โปรแกรมจะแสดงคำแปลในช่องทางขวามือ การค้นหาวิธีนี้จะสะดวกขึ้น ถ้าเลือกวิธีค้นหาแบบรวดเร็ว วิธีนี้เมื่อผู้ใช้เริ่มพิมพ์ตัวอักษรในคำศัพท์ รายการคำศัพท์ก็จะเลื่อนไปแสดงคำศัพท์ที่ใกล้เคียง
7. การค้นหาคำศัพท์โดยค้นผ่านไดอะล็อก Search by Pattern ? หรือ * เพื่อให้โปรแกรมแสดงกลุ่มคำศัพท์ตามเงื่อนไขตัวอักษรทั้งหมดที่กำหนด เมื่อโปรแกรมค้นหากลุ่มคำศัพท์ขึ้นมาแสดงแล้ว ผู้ใช้จึงเลือกคำที่ต้องการแล้วคลิกที่ Select ให้โปรแกรมแสดงคำแปลอีกที
8. การค้นหาโดยเลือกที่แท็บตัวอักษรแรกของคำศัพท์ก่อน แล้วค่อยๆเปิดพจนานุกรมไปที่ละหน้า วิธีนี้จะใช้ได้กับการแสดงผลแบบหนังสือเท่านั้น การค้นหาจะคล้ายกับการเปิดพจนานุกรมจริงๆ



รูปที่ 2-5 แสดงการค้นหาคำศัพท์ของสอง เสดบุตร



รูปที่ 2-6 แสดงการค้นหาคำศัพท์ในรูปแบบหนังสือ

ประเภทและองค์ประกอบของพจนานุกรม

พจนานุกรมเป็นแหล่งที่เก็บคำศัพท์ ความหมายของคำและข้อมูลทางด้านภาษาศาสตร์อื่นๆ เนื่องจากการประมวลผลด้วยคอมพิวเตอร์ทำงานได้อย่างรวดเร็ว แม่นยำและสามารถจัดเก็บข้อมูลได้เป็นจำนวนมาก จึงได้มีการจัดเก็บพจนานุกรมลงบนคอมพิวเตอร์ เพื่อให้การสืบค้นข้อมูลเป็นไปอย่างรวดเร็ว การจัดเก็บข้อมูลลงในคอมพิวเตอร์จึงต้องมีการจัดหมวดหมู่ตามลักษณะของการนำไปใช้งาน (เย็น ภู่วรรณ และชัยยงค์ วงศ์ชัยสุวัฒน์, 2535) ได้กล่าวถึงการจำแนกพจนานุกรมออกเป็นประเภทต่างๆดังต่อไปนี้

พจนานุกรมคอมพิวเตอร์สามารถแบ่งออกได้หลายลักษณะคือ

1. แบ่งตามลักษณะการใช้งาน
2. แบ่งตามลักษณะคู่ภาษา
3. แบ่งตามคุณสมบัติของข้อมูลที่เก็บ
4. แบ่งตามลักษณะที่ให้นักผู้ใช้หรือเครื่องจักรใช้

การแบ่งในแต่ละลักษณะมีรายละเอียดดังต่อไปนี้

1. แบ่งแยกตามลักษณะการใช้งาน โดยยึดหลักตามผู้ใช้ ดังนี้
 - 1.1 พจนานุกรมทั่วไป เป็นพจนานุกรมที่เก็บคำศัพท์สำหรับการใช้งานทั่วไป
 - 1.2 พจนานุกรมศัพท์เทคนิค เป็นพจนานุกรมที่เก็บข้อมูลคำศัพท์เฉพาะสาขาไว้ เช่น ศัพท์ทางการแพทย์ ศัพท์ทางด้านวิศวกรรมไฟฟ้า ศัพท์ทางกฎหมาย ฯลฯ
 - 1.3 พจนานุกรมศัพท์เฉพาะงาน เป็นแหล่งเก็บข้อมูลที่ใช้เฉพาะงาน พจนานุกรมแบบนี้นำมาใช้เฉพาะกับซอฟต์แวร์ต่างๆ เช่น ใช้สำหรับตรวจสอบตัวสะกด เก็บความหมายในลักษณะมโนทัศน์เพื่องานแปลภาษาด้วยคอมพิวเตอร์ที่เรียกว่า คอนเซ็ปต์ ดิกชันนารี เป็นต้น
2. แบ่งแยกตามลักษณะคู่ภาษา จะแบ่งตามประเภทของภาษาที่ใช้ในการเก็บข้อมูล
 - 2.1 พจนานุกรมภาษาเดียว เป็นการเก็บคำศัพท์และความหมายหรือข้อมูลประกอบในลักษณะที่เขียนขึ้นด้วยภาษาเดียวกัน เช่น พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2525
 - 2.2 พจนานุกรมสองภาษา เป็นพจนานุกรมที่ใช้คำศัพท์และความหมายเป็นสองภาษา เช่น พจนานุกรมภาษาอังกฤษ-ไทย
 - 2.3 พจนานุกรมหลายภาษา เป็นพจนานุกรมที่เก็บข้อมูลไว้หลายภาษา มักไม่พบในรูปของหนังสือ เนื่องจากมีขีดจำกัดในการจัดรูปเล่ม แต่อาจพบในสินค้าในรูปของอิเล็กทรอนิกส์ เช่น พจนานุกรมที่มีขนาดเล็กคล้ายเครื่องคิดเลขที่เก็บคำศัพท์ไว้หลายภาษา

3. แบ่งแยกตามคุณสมบัติของข้อมูลที่เก็บ เป็นการแบ่งตามลักษณะของข้อมูลที่เก็บอยู่ภายใน ได้แก่

- 3.1 คลังคำ เป็นการเก็บข้อมูลเกี่ยวกับคำ มีไว้สำหรับใช้ในการตรวจสอบตัวสะกด การแบ่งแยกชนิดคำ คำศัพท์โบราณ
- 3.2 พจนานุกรม เป็นพจนานุกรมที่เก็บข้อมูลเกี่ยวกับภาษา ปัจจุบันมีการจัดทำและจัดเก็บข้อมูลแบบพจนานุกรมกันเกือบทุกภาษา จัดทำไว้เพื่อใช้ในงานทั่วไปหรือใช้เฉพาะงานข้อมูลและรายละเอียดต่างๆจะเก็บรวบรวมไว้เป็นหมวดหมู่ ข้อมูลที่จัดเก็บอาจขึ้นกับวัฒนธรรม สังคมและกาลเวลา
- 3.3 อรรถาภิธาน เป็นพจนานุกรมชนิดหนึ่งที่เกี่ยวข้องกับความหมายของคำ โดยเก็บคำที่มีความหมายคล้ายคลึงกัน เช่น โรเจอร์ ที่ซอร์ส
- 3.4 สารานุกรม เป็นการเก็บรวบรวมความรู้ของมนุษย์เกี่ยวกับเรื่องราวต่างๆ หลายๆด้าน ข้อมูลและรายละเอียดที่จัดเก็บจึงต้องมีมาก

4. แบ่งตามลักษณะที่สำหรับให้มนุษย์ใช้หรือเครื่องจักรใช้ เป็นการแบ่งตามจุดประสงค์ให้ใครเป็นผู้ใช้

- 4.1 พจนานุกรมสำหรับให้มนุษย์ใช้ เป็นพจนานุกรมที่ต้องมีการเรียกค้น และใช้อ้างอิงโดยมนุษย์ขนาดของข้อมูลที่จัดเก็บจะเป็นตัวแบ่งแยกชนิดของพจนานุกรม เช่น ฉบับนักเรียน ฉบับประชาชน และฉบับห้องสมุด เป็นต้น
- 4.2 พจนานุกรมสำหรับคอมพิวเตอร์ เป็นพจนานุกรมที่มีรูปแบบของการจัดเก็บข้อมูล เหมาะสำหรับการประมวลผลด้วยคอมพิวเตอร์ อาจเก็บข้อมูลในรูปแบบของรหัสหรือสัญลักษณ์สำหรับใช้เฉพาะงาน เช่น งานทางด้าน การประมวลผลภาษาธรรมชาติ

ข้อมูลที่มักจัดเก็บไว้ในพจนานุกรมประกอบด้วย

1. คำหลัก เป็นข้อมูลที่จะใช้สำหรับค้นหา หรือเป็นคำหลักที่ผู้ใช้จะเรียกค้นหาข้อมูลประกอบอื่นๆ
2. ข้อมูลการแบ่งคำหรือการแบ่งพยางค์ ข้อมูลคำบางส่วนเป็นคำผสม การตัดแบ่งคำย่อยจึงเป็นข้อมูลประกอบของส่วนต่างๆ
3. การสะกดที่แตกต่างออกไปหรือตัวย่อ เป็นข้อมูลที่เกี่ยวข้องกับคำบางคำที่เขียนได้หลายแบบ

หรือนิยมใช้ในลักษณะคำย่อหรืออักษรย่อ

4. ข้อมูลการใช้เฉพาะพื้นที่ คำบางคำเป็นคำที่ใช้เฉพาะท้องถิ่น หรือใช้กันเฉพาะในบางพื้นที่
5. การผันคำ ในบางภาษาสามารถผันแปรได้ ไปตามลักษณะการใช้ เช่น ภาษาอังกฤษคำจะผันแปรไปเมื่อใช้ในรูปแบบเอกพจน์หรือพหูพจน์
6. การออกเสียง เป็นข้อมูลที่แสดงการอ่าน หรือการออกเสียงที่ถูกต้อง
7. ชนิดของคำ เช่น คำนาม คำกริยา หรือชนิดของคำที่มีการจัดแบ่งชั้นเฉพาะ
8. ความถี่ของการใช้ เป็นข้อมูลที่แสดงถึงว่า คำๆ นั้นมีการใช้มากน้อยแค่ไหน อาจแสดงในรูปของความถี่ของการใช้โดยระบุเป็นเปอร์เซ็นต์ของการใช้
9. คำศัพท์โบราณ สมัยใหม่ คำหายาก หรือคำที่ใช้เป็นทางการ เป็นข้อมูลที่แสดงถึงการใช้ คำศัพท์ว่ามีการใช้ในลักษณะใด
10. รหัสข้อมูล เป็นรหัสที่มีไว้สำหรับเชื่อมโยงหรือให้คอมพิวเตอร์ใช้เชื่อมข้อมูล
11. ความหมาย เป็นข้อมูลเกี่ยวกับคำอธิบายหรือความหมายของคำศัพท์
12. โครงสร้างไวยากรณ์ เป็นข้อมูลเกี่ยวกับโครงสร้างไวยากรณ์ที่จำเป็น ข้อมูลเหล่านี้อาจนำมาใช้เฉพาะงาน เช่น ในระบบพาร์เซอร์ ระบบคอมพิวเตอร์ที่ใช้ในการตรวจสอบเพื่อแยกแยะความกำกวม
13. กรอบการก เป็นข้อมูลเกี่ยวกับคำกริยา ที่จะให้รายละเอียดว่ากริยานี้เกิดกับการกระทำอะไรได้บ้าง
14. ข้อมูลลักษณะประจำ แสดงรายละเอียดของคำ เช่น tense, aspect, mood ฯลฯ
15. ตัวอย่างการใช้งาน อาจแสดงในรูปของประโยคตัวอย่าง
16. ส่วนวน เป็นกลุ่มคำที่เมื่อนำมารวมกันแล้วทำให้มีความหมายเฉพาะ
17. ข้อมูลอรรถกถา เป็นข้อมูลที่แบ่งตามมโนทัศน์ แสดงคำที่มีความหมายเหมือนกันหรือคล้ายคลึงกัน
18. ข้อมูลตัวอย่างการใช้งานและการผันแปรของความหมายกับเวลา คำบางคำอาจใช้ในความหมายต่างกันเมื่อต่างเวลาหรือต่างสถานการณ์
19. คำแปล กรณีนี้ใช้กับพจนานุกรมที่มีหลายภาษา โดยให้คำในภาษาอื่นๆที่มีความหมายเหมือนหรือใกล้เคียงกัน
20. การกระจายคำ เป็นการแตกคำจากรากไปยังคำอื่นที่เป็นไปได้ เช่น การเติมพรีฟิกซ์ หรือ โพลฟิกซ์
21. แหล่งของข้อมูลข่าวสารที่ได้มา เป็นการเก็บแหล่งอ้างอิงของข้อมูลที่จะทำให้สามารถค้นคว้าต่อได้
22. วันเวลาในการปรับปรุงแก้ไขข้อมูล เก็บรายละเอียด วัน เวลา ที่ปรับปรุงรายละเอียดข้อมูล