



บทที่ 2

พระไตรปิฎกฉบับคอมพิวเตอร์ (BUDSIR)

2.1 ประวัติความเป็นมา

ท่านเจ้าคุณพระเทพเวที (พระเทพเวที 2530: 76-77) ให้ความหมายของ พระไตรปิฎก ว่าหมายถึง คัมภีร์ที่บรรจุพุทธพจน์ 3 ชุด หรือ ประมวลแห่งคัมภีร์ที่รวบรวมพระธรรมวินัย 3 หมวด กล่าวคือ พระวินัยปิฎก พระสุตตันตปิฎก และพระอภิธรรมปิฎก

พระวินัยปิฎก ประมวลไว้ซึ่งพุทธบัญญัติเกี่ยวกับ ความประพฤติ ความเป็นอยู่ ขนบธรรมเนียม และการดำเนินกิจการต่าง ๆ ของภิกษุสงฆ์และภิกษุณีสงฆ์ ส่วนพระสุตตันตปิฎก ได้ประมวลไว้ซึ่งพระธรรมเทศนา คำบรรยายธรรมต่าง ๆ ที่ตรัสยกย่องให้เหมาะแก่บุคคลและโอกาส ตลอดจนบทประพันธ์ เรื่องเล่า และเรื่องราวทั้งหลายที่เป็นชั้นเดิมในพระพุทธศาสนา สุดท้ายคือ พระอภิธรรมปิฎก ได้ประมวลหลักธรรมและคำอธิบายที่เป็นหลักวิชาล้วน ๆ ไม่เกี่ยวข้องกับบุคคลหรือเหตุการณ์

การศึกษาพระไตรปิฎก เป็นศาสนกิจหลักของพระสงฆ์ที่ต้องการปฏิบัติตามพระธรรมวินัย และเผยแผ่พระพุทธศาสนา ในขณะที่เดียวกันพุทธศาสนิกชนจำนวนมากก็ใฝ่ใจศึกษาพระไตรปิฎก เพื่อนำหลักธรรมมาใช้เป็นหลักแนวทางในการดำรงชีวิต และนักวิชาการหลายสาขาก็สามารถค้นคว้าหาหลักฐานข้อมูลเกี่ยวกับเรื่องราวในสาขาวิชาของตนจากพระไตรปิฎก เนื่องจากคำสอนและเรื่องราวต่าง ๆ ที่จารึกไว้ในพระไตรปิฎกนั้น เกี่ยวข้องกับแขนงวิชาต่าง ๆ แทบทุกสาขา ได้แก่ ภาษาศาสตร์ ภูมิศาสตร์ ประวัติศาสตร์ มานุษยวิทยา โบราณคดี ปรัชญา จิตวิทยา รัฐศาสตร์ เศรษฐศาสตร์ และนิติศาสตร์

ในประเทศไทยพระไตรปิฎกได้มีการตรวจชำระ และตีพิมพ์เป็นเล่มหนังสือด้วยอักษรไทย ครั้งแรกในสมัยรัชกาลที่ 5 เมื่อปี พ.ศ. 2436 เรียกว่า พระไตรปิฎกฉบับสยามรัฐ ครั้งล่าสุดจัดพิมพ์เมื่อปี พ.ศ. 2523 มีจำนวน 45 เล่ม แต่ละเล่มมีความหนา (เฉพาะเนื้อหา) 260 - 750 หน้า ดังนั้น ในการศึกษาพระไตรปิฎก นอกจากจะพบอุปสรรคในเรื่องของภาษาบาลี ซึ่งไม่ค่อยจะคุ้นเคยเหมือนภาษาไทยของเราเองแล้ว ยังพบกับปริมาณข้อมูลจำนวนมาก การค้นคว้าติดตามหาข้อสรุปในเรื่องราว เหตุการณ์ หรือข้อความเชิงวิชาการที่น่าสนใจ จากข้อมูลที่กระจัดกระจายอยู่ในหลาย ๆ หัวข้อ หลาย ๆ เล่ม ให้ได้อย่างรวดเร็ว ถูกต้อง และสมบูรณ์ครบถ้วนนั้น จึงทำได้ยากมาก

จากข้อปัญหาและอุปสรรคข้างต้น จึงได้มีการศึกษาการนำไมโครคอมพิวเตอร์มาจัดเก็บพระไตรปิฎกเพื่อช่วยการศึกษาค้นคว้าในเชิงวิชาการ การดำเนินงานที่แล้วเสร็จ โดยสำนักคอมพิวเตอร์ มหาวิทยาลัยมหิดล (คุภชัย ตั้งวงศ์คานต์, ดำรัส วงศ์สว่าง และ นันทิกา เบญจเทพานันท์ 2531: 4,

6-9, 13-15, 17-27) เมื่อเดือน พฤษภาคม พ.ศ. 2531 มีดังนี้

1. สร้าง text editor ไทย-บาลี 25 บรรทัด บนเครื่องประเภทไอบีเอ็มพีซี และได้บันทึกข้อมูลของพระไตรปิฎก จำนวน 45 เล่มพิมพ์ รวม 24.23 ล้านตัวอักษรเข้าคอมพิวเตอร์
2. สร้าง BUDSIR เพื่อการสืบค้นสารสนเทศพระไตรปิฎก ด้วยศัพท์ในหลายรูปแบบ ทั้งคำเดี่ยว วลีหรือพหูพจน์ รวมทั้งส่วนของศัพท์ที่เกิดจากการสมมูล
3. ปรับรตต้นฉบับพระไตรปิฎกคอมพิวเตอร์ภาษาบาลีอักษรไทย เป็นภาษาบาลีอักษรโรมัน โดยคงข้อความไว้ตามฉบับของอักษรไทยทั้งหมด พระไตรปิฎกคอมพิวเตอร์อักษรโรมันนี้ มีปริมาณของข้อมูลรวม 30.13 ล้านตัวอักษร
4. พัฒนาโปรแกรมคอมพิวเตอร์ชื่อ บุคเซอร์ทู่ (BUDSIR-II) เพื่อการสืบค้นสารสนเทศพระไตรปิฎกฉบับภาษาบาลีอักษรโรมัน โดยใช้หลักการเดียวกันกับ บุคเซอร์

2.2 โครงสร้างของภาษาบาลีและการจัดเก็บเข้าคอมพิวเตอร์

2.2.1 โครงสร้างของภาษาบาลี

ภาษาบาลี มีชื่อเรียกอีกชื่อหนึ่งว่า ภาษามคธ จัดอยู่ในตระกูลภาษาอินเดีย-ยุโรป เป็นภาษาที่มีกตติปัจจัยแตกต่างจากภาษาไทย ซึ่งจัดอยู่ในภาษาคำโดด ภาษาบาลีแท้จริงแล้วไม่มีตัวอักษรเป็นของตนเองมีเฉพาะเสียง ไม่มีรูปตัวอักษร ชาติใดที่เรียนภาษาบาลี ก็จะใช้ตัวอักษรของตนเองในการถ่ายทอดเสียงภาษาบาลี โดยจัดระบบเสียงทั้งสระและพยัญชนะ สำหรับตัวอักษรภาษาไทยที่ใช้ถ่ายทอดเสียงภาษาบาลี มีพยัญชนะ 33 ตัว และ สระ 8 ตัว ดังนี้

พยัญชนะ 33 ตัว คือ

ก ข ค ฉ ง	จ ฉ ช ฌ ญ	
ฎ จ ฏ ฒ ณ	ต ถ ท ธ น	
ป ผ ฝ ภ ม	ย ร ล ว ส	ห ฬ ° (อँग หรือ นิคหิต)

สระ 8 ตัว คือ

อ (อ่านว่า อะ) -า -ิ -ึ -ุ -เ -โ-

เครื่องหมายวรรคตอน ได้แก่ ๔ (เครื่องหมายจบประโยค) . (จุดนิท)

นอกจากนี้ ยังได้ใช้เครื่องหมายวรรคตอน ซึ่งไม่มีอยู่ในภาษาบาลี เพื่อความสะดวก

ในการใช้อักษรภาษาไทยบันทึกหรือตีพิมพ์ภาษาบาลี ได้แก่ () . และชุดตัวเลขไทย

ข้อแตกต่างของภาษาบาลีกับภาษาไทยในแง่ของการป้อนข้อมูลเข้า (data entry) และการแสดงข้อความออกทางจอภาพ (display) มีดังนี้

1. ตัวอักษรภาษาไทยที่ใช้ถ่ายทอดเสียงภาษาบาลีนั้น จะมี 2 ตัวอักษรที่ต่างกัน คือ

ภาษาไทย

ภาษาบาลี

ญ , ฐ

ณ , จ

2. ภาษาบาลีใช้ . (จุดนิท) ใส่ไว้ข้างล่างตัวอักษรที่เป็นตัวสะกด หรือตัวควบกล้ำ เช่น สตุตทานิ, อวิชชา, พุรหม, ตตฺร
3. ภาษาบาลีจะเว้นช่องว่างระหว่างคำทุกคำเช่นเดียวกับภาษาอังกฤษ จึงไม่ต้องมีปัญหारेื่องการตัดคำ เช่นที่พบในภาษาไทย ตัวอย่างเช่น "เอวมุเม สตุ อภิ สมย ภควา"
4. ภาษาบาลีไม่มีวรรณยุกต์ จึงทำให้การจัดระดับของภาษาบาลีมีเพียง 3 ระดับคือ
 - ระดับที่ 1 ° ° ° (อँग)
 - ระดับที่ 2 พยัญชนะทั้งหมด (ยกเว้น °) อ า เ โ
 - และ ระดับที่ 3 , ุ . (จุดนิท)

2.2.2 การจัดเก็บข้อมูลภาษาบาลีเข้าคอมพิวเตอร์

การเก็บข้อมูลพระไตรปิฎกภาษาบาลีอักษรไทยเข้าคอมพิวเตอร์นั้น ได้ใช้รหัสคอมพิวเตอร์แทนอักษรไทย ที่ได้มีการกำหนดไว้ก่อนแล้ว สำหรับบันทึกในส่วนของอักษรไทย ยกเว้นตัว ณ จ และ . (จุดนิท) ที่แตกต่างออกไปนั้น ได้กำหนดรหัสในตาราง ASCII ในตำแหน่งที่ "ว่าง" ซึ่งเหลือจากการแทนอักษรไทยและอังกฤษ รหัสที่ใช้ (ระบุด้วยเลขฐาน 16) ได้แก่

ณ ใช้รหัสเป็น "9A" , จ ใช้รหัสเป็น "CD" และ . (จุดนิท) ใช้รหัสเป็น "E5"

นอกจากนี้ ในการแสดงผลของภาษาบาลีบนจอภาพให้ได้ 25 บรรทัด ตัวอักษร (นิคหิต) ที่ใช้พยัญชนะต้นร่วมกับสระ , หรือสระ ° เช่น สกฺธิ ปจฺจสฺสโล เป็นต้น จะถูกเปลี่ยนมาใช้รหัสของสระร่วมแทน ดังนี้

สระร่วม ° ใช้รหัสเป็น 89 (เลขฐาน 16)

และ ° ใช้รหัสเป็น DB (เลขฐาน 16)

รูปที่ 2.1 แสดงรหัสคอมพิวเตอร์สำหรับตัวอักษรไทย-บาลี-อังกฤษ ในตาราง ASCII และในรูปที่ 2.2 จะแสดงให้เห็นข้อความพระไตรปิฎกภาษาบาลีอักษรไทย ซึ่งถูกเก็บเข้าคอมพิวเตอร์

4 bit แรก	4 bit ท้าย															
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1																
2	sp	!	"	#	*	%	&		()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8										°	'	~	.	:	-	
9	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	ณ					
A		ก	ข	ค	ฌ	ง	จ	ฉ	ช	ฌ	ณ	ญ	ฎ	ฏ	ฐ	ฑ
B	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ	ฟ	ภ	ม
C	ย	ร	ฤ	ล	ว	ศ	ษ	ส	ห	ฬ	อ	ธ	ช	ฉ	า	ำ
D	เ	แ	ไ	ใ	เ	อ	๑			๐	๐	๐	๐	๐	๐	๐
E	.	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑
F	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑	๑

รูปที่ 2.1 ตาราง ASCII สำหรับตัวอักษรไทย-บาลี-อังกฤษ

(๕) วิปัสสี ภิกขเว ภควา อรหํ สมมาสัมพุทโธ ปาฏลिया
 มุเล อภิสัมพุทโธ ๑ ลีชี ภิกขเว ภควา อรหํ สมมาสัมพุทโธ
 ปุณฑริกสฺส มุเล อภิสัมพุทโธ ๑ เวสฺสภู ภิกขเว ภควา อรหํ
 สมมาสัมพุทโธ สาลสฺส มุเล อภิสัมพุทโธ ๑ กกุสนฺโธ ภิกขเว
 ภควา อรหํ สมมาสัมพุทโธ ลีรีสฺส มุเล อภิสัมพุทโธ ๑ โจนาคมน
 ภิกขเว ภควา อรหํ สมมาสัมพุทโธ อุตฺมพรสฺส มุเล
 อภิสัมพุทโธ ๑ กสฺสโป ภิกขเว ภควา อรหํ สมมาสัมพุทโธ
 นิโครธสฺส มุเล อภิสัมพุทโธ ๑ อหํ ภิกขเว เอตฺรหิ อรหํ
 สมมาสัมพุทโธ อสฺสตุถสฺส มุเล อภิสัมพุทโธ ๑

(๖) วิปัสสีสฺส ภิกขเว ภควโต อรหโต สมมาสัมพุทฺธสฺส
 ขณฺฑติสฺสํ นาม สวากฺขํ อโหสิ อคฺคํ ภาทฺทํ ๑ ลีชีสฺส
 ภิกขเว ภควโต อรหโต สมมาสัมพุทฺธสฺส อภิสุมฺภวํ นาม
 สวากฺขํ อโหสิ อคฺคํ ภาทฺทํ ๑ เวสฺสภุสฺส ภิกขเว ภควโต
 อรหโต สมมาสัมพุทฺธสฺส โสณฺฑตฺรํ (๑) นาม สวากฺขํ อโหสิ อคฺคํ
 ภาทฺทํ ๑ กกุสนฺธสฺส ภิกขเว ภควโต อรหโต สมมาสัมพุทฺธสฺส
 วิธูรสนฺธิวํ นาม สวากฺขํ อโหสิ อคฺคํ ภาทฺทํ ๑ โจนาคมนสฺส
 ภิกขเว ภควโต อรหโต สมมาสัมพุทฺธสฺส ภิกฺขโยสฺสตุถฺรํ นาม
 สวากฺขํ อโหสิ อคฺคํ ภาทฺทํ ๑ กสฺสโปสฺส ภิกขเว ภควโต
 อรหโต สมมาสัมพุทฺธสฺส ติสฺสภารทฺวาชํ นาม สวากฺขํ อโหสิ
 อคฺคํ ภาทฺทํ ๑ มยฺหํ ภิกขเว เอตฺรหิ สารีปุตฺตโมคฺคฺคฺคฺคฺคฺคฺคฺคฺค
 นาม สวากฺขํ อโหสิ อคฺคํ ภาทฺทํ ๑

(๗) วิปัสสีสฺส ภิกขเว ภควโต อรหโต สมมาสัมพุทฺธสฺส
 ตโย สวากานํ สนฺนิปาตา อเหสุ เอโก สวากานํ สนฺนิปาโต

รูปที่ 2.2 ตัวอย่างข้อความภาษาบาลีอักษรไทยที่นำเก็บเข้าแฟ้มข้อมูลพระไตรปิฎก

2.3 บุตเซอ์ (BUDSIR หรือ BUDdhist Scripture Information Retrieval)

2.3.1 โครงสร้างของ BUDSIR

BUDSIR (BUDdhist Scripture Information Retrieval) คือชุดซอฟต์แวร์ที่สร้างบนไมโครคอมพิวเตอร์ประเภท IBM/PC เพื่อการจัดเก็บข้อมูลและการค้นหาข้อมูลในพระไตรปิฎก ซึ่งจะแยกกล่าวได้เป็น 2 ส่วน คือ กลุ่มโปรแกรมของบุตเซอ์ และฐานข้อมูลพระไตรปิฎก

กลุ่มโปรแกรมของ BUDSIR

BUDSIR ประกอบด้วย 3 กลุ่มโปรแกรม เพื่อทำหน้าที่ต่าง ๆ กัน ดังนี้

1. กลุ่มโปรแกรมเพื่อการจัดสร้างและบำรุงรักษาฐานข้อมูลพระไตรปิฎก
2. กลุ่มโปรแกรมเพื่อการค้นหาและดึงข้อมูล
3. กลุ่มโปรแกรมเพื่อบริการผู้ใช้งาน ที่สำคัญ ได้แก่
 - โปรแกรมแปลงข้อความเป็นอักษรโรมัน
 - โปรแกรม print screen ของบาลีและโรมัน

ฐานข้อมูลพระไตรปิฎก

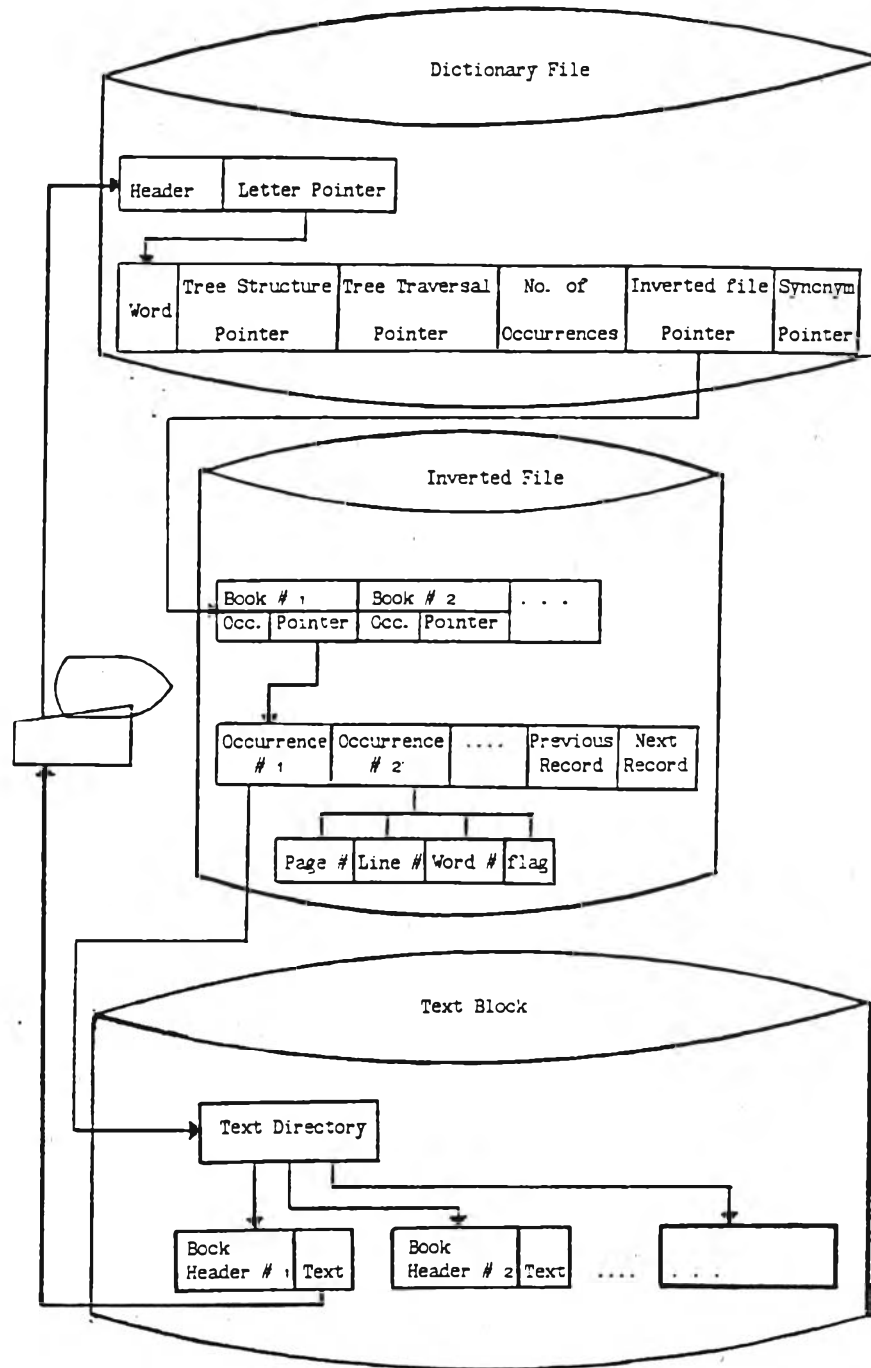
ฐานข้อมูลพระไตรปิฎกมีโครงสร้างเป็นแบบ **แฟ้มข้อมูลย้อนกลับ** (inverted files) ดังแสดงในรูปที่ 2.3 ซึ่งประกอบด้วยแฟ้มข้อมูล 3 ประเภท ดังนี้

1. แฟ้มข้อมูลพระไตรปิฎก (Text Files)

เป็นแฟ้มข้อมูลซึ่งเก็บพระไตรปิฎกรวม 45 เล่ม ข้อมูลพระไตรปิฎกถูกจัดเก็บในแฟ้มข้อมูลเป็นลักษณะ บล็อกข้อความ (text block) และจัดระเบียบอยู่ในรูปแบบ (format) ซึ่งพร้อมที่จะแสดงผลบนจอภาพ หรือออกทางเครื่องพิมพ์ได้ทันที นอกจากนี้ยังมีแฟ้มข้อมูล **ไดเรกทอรี** (Text Directory) ซึ่งจะใช้สำหรับควบคุมการเข้าถึงข้อมูลให้สามารถถึงสารสนเทศพระไตรปิฎกหน้าใด ๆ หรือหัวข้อใด ๆ ในเล่มที่ต้องการได้ถูกต้อง แฟ้มข้อมูลพระไตรปิฎก (Text Files) มีขนาดรวม 35 MB

2. แฟ้มข้อมูลพจนานุกรม (Dictionary Files)

เป็นแฟ้มข้อมูลซึ่งเก็บคำศัพท์ (คำที่ไม่ซ้ำกัน) ทุกคำในพระไตรปิฎก คำศัพท์ทุกคำมี



รูปที่ 2.3 แสดงโครงสร้างฐานข้อมูลพระไตรปิฎก

ตัวชี้ (pointer) โยงไปหา รายการตำแหน่งคำ (Occurrence List) ใน แฟ้มข้อมูลย้อนกลับ (Inverted File) การจัดเรียงลำดับของคำศัพท์จะอยู่ในรูปของ โครงสร้างแบบทรี (tree structure) โดยมีตัวชี้โยงไปยังลำดับต่าง ๆ ในทรี (tree) นอกจากนี้ ยังมี แฟ้มดิกแพค (Dict-pack) ซึ่งเก็บรวบรวมคำศัพท์ทุกคำเข้าด้วยกันเป็นแฟ้มลำดับ (sequential file) แฟ้มข้อมูลพจนานุกรม มีขนาดรวม 9 MB

3. แฟ้มข้อมูลย้อนกลับ (Inverted files)

เป็นแฟ้มข้อมูลเก็บรายการตำแหน่งคำ ได้แก่ ตำแหน่งของคำศัพท์ที่ปรากฏในแฟ้มข้อมูลบล็อกข้อความ ทั้งหมด โดยมีตัวชี้โยงจากคำศัพท์แต่ละคำในแฟ้มข้อมูลพจนานุกรมมายังรายการตำแหน่งของคำนั้น ๆ ตำแหน่ง (Occurrence Code) ใช้เป็นรหัสตัวเลขระบุ เล่ม/หน้า/บรรทัด/คำที่ และแฟล็ก (flag) ซึ่งเป็นเครื่องหมายแสดงว่าเป็นคำสุดท้ายของบรรทัดหรือคำสุดท้ายของหน้า เพื่อสะดวกแก่การประมวลผลในตอนค้นหา วลี นั้นเอง แฟ้มข้อมูลย้อนกลับ มีขนาดรวม 20 MB

2.3.2 ระบบสืบค้นสารสนเทศของ BUDSIR

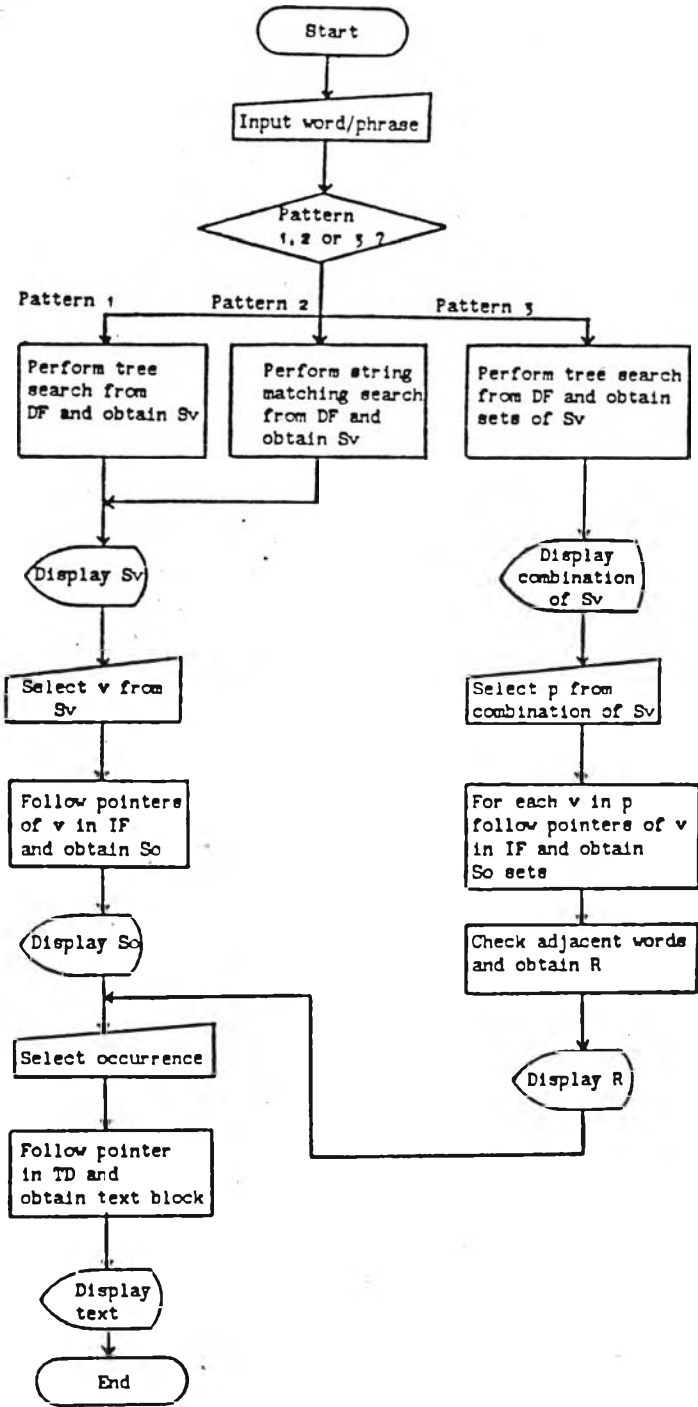
บุคเซอร์ได้ถูกออกแบบในส่วนของ การค้น ให้ใช้งานได้ในลักษณะที่เรียกว่า ใช้เมนูเป็นหลัก (Menu-driven) ซึ่งมีขั้นตอนการแสดงผลเมนูให้เลือกไปตามลำดับ การค้นสารสนเทศพระไตรปิฎกโดยบุคเซอร์ สามารถจะทำได้ 2 วิธี ดังนี้

ก. การค้นสารสนเทศตามเล่ม/ข้อ/หน้า

การค้นโดยวิธีนี้ มีลักษณะเช่นเดียวกับการเปิดอ่านหนังสือ โดยเลือกเล่มที่ต้องการอ่านก่อน เมื่อได้เล่มแล้ว ให้ป้อนเลขข้อหรือเลขหน้า ที่ต้องการเข้าไป จากนั้น สารสนเทศพระไตรปิฎกที่ข้อ/หน้านั้น ๆ จะถูกดึงออกมาแสดงบนจอภาพคอมพิวเตอร์ จำนวน 23 บรรทัด และจะมีปุ่มอำนวยความสะดวกแก่ผู้ใช้ให้สามารถเลื่อนดูข้อความในหน้าถัดไป หรือหน้าที่แล้ว สามารถแปลงสารสนเทศที่แสดงบนจอภาพให้เป็นภาษาบาลีอักษรโรมัน สามารถสั่งพิมพ์ข้อความบนจอภาพออกทางเครื่องพิมพ์ และอื่น ๆ

ข. การสืบค้นสารสนเทศด้วยคำ ส่วนของคำ หรือวลี

คำ หมายถึง คำศัพท์ภาษาบาลี เช่น นิพพาน ทุกข์ สุข อนิจจา
วลี หมายถึง กลุ่มคำภาษาบาลีที่เรียงต่อกันไป เช่น "อตุตา ทิ อตุตโน นาโก"



Remarks : Symbols and meaning

- | | | | |
|----|-----------------|----|------------------------------|
| DF | Dictionary File | So | Occurrence Set |
| IF | Inverted File | R | Resultant of occurrence sets |
| TD | Text Directory | v | Vocabulary |
| Sv | Vocabulary Set | p | Phrase |

รูปที่ 2.4 แสดงผังงานของขบวนการสืบค้นสารสนเทศของ BUDSIR

ส่วนของคำ หมายถึง ส่วนของคำศัพท์ภาษาบาลี ซึ่งประกอบด้วยกลุ่มอักษรบาลี และมีเครื่องหมาย "*" , "?" (wild card characters) เติมอยู่ด้วย

โดยที่ * แทน อักษรหรือกลุ่มอักษรใด ๆ ไม่จำกัดความยาว
 ? แทน อักษรใด ๆ 1 อักษร

ตัวอย่างเช่น นิพพาน* ใช้แทน คำศัพท์ใด ๆ ที่ขึ้นต้นด้วย "นิพพาน"
 ลุข? ใช้แทน คำศัพท์ใด ๆ ที่ขึ้นต้นด้วย "ลุข" ตามด้วยอักษรใด ๆ 1 ตัว
 กุมาร ใช้แทน คำศัพท์ใด ๆ ที่มี "กุมาร" เป็นส่วนหนึ่งของคำ
 "อตุต* อุปม* กตฺวา" ใช้แทน วลีใด ๆ ที่มีคำแรกขึ้นต้นด้วย "อตุต" คำที่สองขึ้นต้นด้วย "อุปม" และคำที่สามคือ "กตฺวา"

ดังนั้นการค้นในลักษณะนี้ของ BUDSIR จะป้อนคำค้นได้ 3 รูปแบบดังนี้

(i) <string>

(ii) *<string>

และ (iii) <string> <string> <string>

โดยที่ <string> หมายถึง คำบาลี หรือกลุ่มอักษรที่มี * หรือ ? ต่อท้ายได้

ขบวนการสืบค้นสารสนเทศด้วยคำ ส่วนของคำ หรือวลี ของ BUDSIR จะมีขั้นตอนการดำเนินงาน ดังแสดงในผังงาน รูปที่ 2.4 ซึ่งจะแสดงให้เห็นขั้นตอนการค้นในทั้ง 3 รูปแบบ เมื่อพบสารสนเทศพระไตรปิฎกที่ต้องการ จะนำออกแสดงทางจอภาพ พร้อมทั้งแสดงเป็นแถบสีสว่าง (highlight) บนคำหรือวลีที่ใช้ค้น และจะมีปุ่มอำนวยความสะดวกแก่ผู้ใช้ ให้สามารถดึงสารสนเทศที่ปรากฏคำ/วลีนั้น ๆ ในตำแหน่งแรกสุด ในตำแหน่งสุดท้าย ในตำแหน่งถัดไป หรือในตำแหน่งก่อนหน้า นอกจากนั้น ยังสามารถแปลงสารสนเทศที่แสดงบนจอภาพให้เป็นภาษาบาลีอักษรโรมัน สามารถสั่งพิมพ์ข้อความบนจอภาพออกทางเครื่องพิมพ์ และอื่น ๆ

2.4 ปัญหาและขีดจำกัดในการค้นของ BUDSIR

บุตรเซอร์ นับเป็นงานวิจัยเริ่มต้นด้านการนำคอมพิวเตอร์มาช่วยในการศึกษาค้นคว้าพระไตรปิฎก โดยใช้คอมพิวเตอร์เก็บบันทึกพระไตรปิฎก และช่วยในการค้น ซึ่งทำความพอใจให้แก่ผู้ศึกษาพระไตรปิฎกเป็นอย่างมาก อย่างไรก็ตาม บุตรเซอร์ก็มีขีดความสามารถในการค้นเพียงระดับหนึ่งเท่านั้น โดยที่บุตรเซอร์ยังไม่สามารถช่วยสืบค้นสารสนเทศพระไตรปิฎกในลักษณะที่หลากหลายซับซ้อน ซึ่งพบว่า จะเป็นประโยชน์แก่ผู้ศึกษาค้นคว้าพระไตรปิฎกมากยิ่งขึ้น ดังจะสรุปให้เห็น ดังต่อไปนี้

2.4.1 ความซับซ้อนในการค้น เนื่องจากลักษณะของสารสนเทศที่เป็นภาษาธรรมชาติ

ภาษาบาลี นับเป็นภาษาธรรมชาติภาษาหนึ่ง และโดยลักษณะทั่วไปของข้อความที่เป็นภาษาธรรมชาติไม่ว่าภาษาใด การอ้างถึงเรื่องราวหนึ่ง มักต้องอ้างอิงด้วยคำ เกินกว่า 1 คำ เล่มอ จึงจะสื่อความหมายถึงเรื่องราวหนึ่ง ๆ ได้ตรงจุด ดังนั้น ในการสืบค้นสารสนเทศที่เป็นภาษาธรรมชาติด้วยคำเดียว โดยเฉพาะ ถ้าบ่อนคำค้นที่มีตำแหน่งการปรากฏมาก ๆ เป็นคำค้น ผู้ใช้ก็จะพบกับสารสนเทศผลลัพธ์จำนวนมาก ทั้ง ๆ ที่ผู้ใช้มีความต้องการเพียงบางส่วนของสารสนเทศเท่านั้น ซึ่งผู้ใช้ก็ต้องมาคัดเลือก และแยกแยะเอาเฉพาะส่วนที่ต้องการออกมาอีกทีหนึ่ง ฉะนั้น หากให้ผู้ใช้สามารถค้นหาสารสนเทศด้วยการอ้างอิงคำค้นได้เกินกว่า 1 คำ ก็จะช่วยให้ผู้ใช้สามารถค้นหาสารสนเทศเป้าหมายได้สะดวกและรวดเร็วยิ่งขึ้น

การสืบค้นสารสนเทศที่ต้องอ้างอิงคำค้นมากกว่า 1 คำ อาจมีได้หลายลักษณะ ซึ่งในลักษณะหนึ่ง ได้แก่ การค้นหาสารสนเทศที่ต้องปรากฏคำค้นที่ระบุตั้งแต่ 2 คำ ขึ้นไป ในสารสนเทศเดียวกัน เช่น การค้นหาสารสนเทศที่เป็นเรื่องราวของพระอานนท์ในเมืองสาวัตถี ซึ่งอาจจะระบุคำค้นเป็น "อานนท" และ "สาวัตถี" เป็นต้น

การสืบค้นสารสนเทศในอีกลักษณะหนึ่งที่ต้องระบุคำมากกว่า 1 คำ ได้แก่ การค้นหาสารสนเทศที่ต้องปรากฏคำค้นคำแรก แต่ต้องไม่ปรากฏคำค้นที่ระบุอีกคำหนึ่ง อยู่พร้อมกันในสารสนเทศเดียวกันนั้น เช่น การค้นหาสารสนเทศที่เป็นเรื่องราวของพระสารีบุตรในตอนที่ไม่ มีพระอานนท์ เข้ามาร่วมด้วย ซึ่งอาจจะระบุคำค้นแรกเป็น "สารีบุตร" และระบุคำที่สองเป็น "อานนท" เป็นต้น

จะเห็นได้ว่า ความต้องการค้นในลักษณะนี้ บุคเซอร์ไม่อาจดำเนินการค้นได้

2.4.2 ความซับซ้อนในการค้น เนื่องจากลักษณะ เฉพาะของสารสนเทศที่เป็นภาษาบาลี

จากการที่ภาษาบาลีเป็นภาษาที่มีวิภัติปัจจัย คือคำศัพท์คำหนึ่งสามารถเปลี่ยนรูปไปได้หลายแบบขึ้นกับหน้าที่ของคำในประโยค เช่น อานนท (อ่านว่า อา-นัน-ทะ) อานนุโท (อ่านว่า อา-นัน-โท) อานนุทั (อ่านว่า อา-นัน-ทัง) ซึ่งหมายถึง พระอานนท์ทั้ง 3 คำ เป็นต้น ดังนั้น เมื่อต้องการสืบค้นสารสนเทศหัวข้อใด ๆ ในพระไตรปิฎก ที่เกี่ยวกับพระอานนท์ในทุกลักษณะให้ได้ครบถ้วนสมบูรณ์ ก็ต้องค้นด้วยคำศัพท์ทุกคำ ที่มีความหมายถึงพระอานนท์ ด้วยบุคเซอร์ผู้ใช้จะต้องค้นหลายคราว เท่ากับจำนวนคำศัพท์เหล่านั้น โดย ใช้คำศัพท์ทีละคำ ไม่สามารถจะระบุคำค้นทั้งกลุ่มเพื่อสืบค้นสารสนเทศได้ครบถ้วนในคราวเดียว และยิ่งกว่านั้น ผู้ใช้ต้องนำสารสนเทศผลลัพธ์ที่ได้จากการค้นทุกคราวมาเปรียบเทียบ และคัดเลือกเอาสารสนเทศที่เป็นหัวข้อซ้ำกันออกไปอีกทีหนึ่ง แน่แน่นอนว่าถ้ายิ่ง

จำนวนคำมาก และปริมาณสารสนเทศมาก ก็ยิ่งเสียเวลามากยิ่งขึ้น

2.4.3 ความซับซ้อนในการค้นเนื่องจากลักษณะ เฉพาะของ เนื้อหาสารสนเทศพระไตรปิฎก

จากลักษณะ เนื้อหาของสารสนเทศพระไตรปิฎก ผู้ใช้มักต้องการสืบค้นสารสนเทศด้วยการ บ้อนพุทธภษิต หรือพุทธพจน์ จึงจะตรงเป้าหมาย ซึ่งพุทธภษิต หรือพุทธพจน์ที่ใช้ค้น อาจ ประกอบด้วยคำศัพท์เกินกว่า 5 คำ ดังนั้น การสืบค้นสารสนเทศพระไตรปิฎกในลักษณะนี้ ไม่อาจทำได้บนชุดเซออร์ ซึ่งจำกัดจำนวนคำที่ประกอบเป็นวลีที่ใช้ค้นไว้ไม่เกิน 5 คำ

2.4.4 ความหลากหลายในรูปแบบของ ส่วนของคำ เนื่องจากลักษณะ เฉพาะของคำภาษาบาลี

จากการที่ภาษาบาลีเป็นภาษาที่มีการสมาสและสนธิ ซึ่งเป็นการผสมกันของคำศัพท์ 2 คำ แล้วได้เป็นคำศัพท์ใหม่ ทั้งนี้ คำศัพท์ใหม่อาจจะคงรูปดั้งเดิมเหมือนนำมาต่อกันเฉย ๆ หรือ มีการเปลี่ยนรูปอักษร เพิ่มหรือลดอักษรไปจากเดิมก็ได้ ตามแต่กฎเกณฑ์ของการสมาส-สนธิ และจากลักษณะการใช้ ส่วนของคำ เพื่อระบุถึงกลุ่มคำศัพท์ที่เกิดจากการสมาส-สนธิ (ซึ่งมี ส่วนของคำเหมือนกัน) ที่มีให้ใช้ในชุดเซออร์นั้น นับว่าอำนวยความสะดวกแก่ผู้ใช้เป็นอันมาก แต่รูปแบบการบ้อนส่วนของคำยังจำกัด ไม่สามารถบ้อนส่วนของคำในรูปแบบอื่น ๆ ที่จะอ้างถึงกลุ่มคำศัพท์ที่เกิดจากการสมาส-สนธิ ในอีกหลายรูปแบบ ซึ่งน่าจะได้ขยายขอบข่ายออกไป ได้แก่

- ส่วนของคำที่มี "*" นำหน้า และมี "?" เติมท้าย

เช่น "*กุมาร???" ใช้แทนกลุ่มคำศัพท์ จณฑาลกุมาริกา กุมาริโก

- ส่วนของคำประเภทที่มี "?" นำหน้า และอาจมี "*" หรือ "?" เติมท้ายหรือไม่ก็ได้

เช่น "???กุมาร" ใช้แทนกลุ่มคำศัพท์ ราชกุมาร กุลกุมาร

- ส่วนของคำประเภทที่มี "?" แทรกในระหว่างเทอม และอาจมี "*" หรือ "?" นำหน้า และ/หรือ เติมท้ายหรือไม่ก็ได้

เช่น "*โรวาท" ใช้แทนกลุ่มคำศัพท์ ราโรวาท อาจารย์โรวาท พุโรวาท

การจัดให้สามารถใช้ส่วนของคำ ในรูปแบบเหล่านี้ จะเป็นประโยชน์แก่ผู้ใช้อย่างแน่นอน

2.4.5 ปัญหาด้านความสะดวกในการค้น

แม้ว่าชุดเซออร์จะได้ออกแบบการใช้งาน ให้อำนวยความสะดวกแก่ผู้ใช้เป็นอันมากก็ตาม แต่ ก็พบว่า ยังมีบางกรณีที่ไม่อาจทำได้หรือไม่ได้จัดทำไว้ ซึ่งจะเป็นที่ต้องการของผู้ใช้ ได้แก่

- การเก็บผลลัพธ์การค้นแต่ละครั้งไว้ ในการใช้ชุดเซออร์เมื่อผู้ใช้ต้องการเรียกดูสารสนเทศที่ค้นผ่านไปแล้ว ผู้ใช้ต้องค้นใหม่

- การลิ่งพิมพ์สารสนเทศผลลัพธ์ บุคเซอร์จะให้พิมพ์ได้ทีละหน้า เฉพาะหน้าที่แสดงอยู่บนจอภาพ ถ้าต้องการพิมพ์ผลทั้งหมดต้องดึงสารสนเทศหน้าต่อไป แล้วลิ่งพิมพ์อีก ทำซ้ำ ๆ เช่นนี้ไปเรื่อย ๆ จึงนับว่าไม่สะดวกแก่ผู้ใช้ที่ต้องการผลพิมพ์ทั้งหมด
- การค้นสารสนเทศด้วยวลีบนบุคเซอร์ เมื่อป้อนวลีที่ค้น เป็นวลีที่ประกอบด้วยคำตัดปลาย (ส่วนของคำที่มี "*" เติมท้าย) บุคเซอร์จะแสดงรายการวลีที่เกิดจากการผสมของกลุ่มคำศัพท์ที่เป็นไปได้ทางจอภาพ แล้วให้ผู้ใช้เลือกวลีที่ต้องการ เพื่อค้นหาว่ามีในพระไตรปิฎกหรือไม่ แต่ วลีส่วนหนึ่งในรายการวลี จะไม่พบในพระไตรปิฎก ซึ่งจะทำให้ผู้ใช้รู้สึกเบื่อหน่ายในการเลือกและค้นซ้ำแล้วซ้ำอีกโดยเปล่าประโยชน์ จึงควรจัดทำฟังก์ชันช่วยค้นหาวลี และบอกจำนวนที่พบโดยอัตโนมัติ เพื่ออำนวยความสะดวกแก่ผู้ใช้

จากปัญหาดังกล่าวนี้อเอง จึงเกิดความคิดที่จะพัฒนา ระบบสืบค้นสารสนเทศพระไตรปิฎก ระบบใหม่ ที่มีขอบข่ายการค้นขยายออกไปอย่างกว้างขวาง โดยเฉพาะอย่างยิ่ง การค้นนั้นจะให้ผู้ใช้เป็นผู้กำหนดทิศทางการค้นด้วยตัวเอง อันจะเป็นประโยชน์แก่ผู้ศึกษาค้นคว้าพระไตรปิฎกอย่างยิ่ง

แนวความคิดต่าง ๆ ที่จะนำมาใช้ในการพัฒนาระบบสืบค้นสารสนเทศพระไตรปิฎกระบบใหม่นี้ จะได้บรรยายในบทที่ 3 ต่อไป