บทที่ 3

การเข้าและถอดรหัสข้อความไทยในเอกสารพีดีเอฟ

จากปัญหาของการค้นข้อความไทยในเอกสารพีดีเอฟ ที่ไม่สามารถค้นข้อความได้อย่างถูกต้อง เนื่องมาจากเครื่องมือการค้นในโปรแกรมแสดงเอกสารพีดีเอฟไม่เข้าใจการเข้ารหัสที่แบบอักษรไทยใช้ ในการแก้ไขปัญหา ผู้วิจัยทำการสร้างเครื่องมือการค้นข้อความในเอกสารพีดีเอฟที่เข้าใจการเข้ารหัสของชุดแบบไทยที่ใช้ในเอกสารพีดีเอฟ โดยผู้วิจัยทำการศึกษาว่าข้อความไทยในเอกสารพีดีเอฟมีการเข้ารหัสอย่างไรและควรใช้วิธีการถอดรหัสอย่างไร ในบทนี้จะกล่าวถึงการเข้ารหัสตัวอักษรไทยในเอกสารพีดีเอฟและวิธีในการที่จะถอดรหัสตัวอักษรเหล่านั้น

3.1 การเข้ารหัสตัวอักษรของแบบอักษรไทยในเอกสารพีดีเอฟ

เอกสารทั่วไปเมื่อทำการแปลงมาอยู่ในรูปแบบเอกสารพีดีเอฟ แบบอักษรที่ใช้ในเอกสารเดิมอาจจะถูกเปลี่ยนชนิดหรือเปลี่ยนการเข้ารหัสต่างไปจากข้อมูลเดิม ขึ้นอยู่กับข้อกำหนดและสภาพแวดล้อมที่ใช้ในการสร้างเอกสารพีดีเอฟ เอกสารพีดีเอฟภาษาไทยก็เช่นกัน แบบอักษรไทยที่ใช้ในเอกสารทั่วไปจะเข้ารหัสตามข้อกำหนดมาตรฐาน มอก.620 เมื่อแปลงมาอยู่ในรูปแบบเอกสารพีดีเอฟ พบว่า แบบอักษรไทยในเอกสารพีดีเอฟมีการเข้ารหัสตัวอักษรอยู่ 3 แบบด้วยกัน คือ

1. เข้ารหัสตามข้อกำหนดมาตรฐาน (Standard)
2. เข้ารหัสแบบผู้ใช้กำหนดการเข้ารหัสตัวอักษรเอง (Custom)
3. ใช้ข้อมูลการเข้ารหัสที่มีอยู่ในแฟ้มแบบอักษร (Built-in)

ข้อความไทยในเอกสารพีดีเอฟที่ใช้แบบอักษรไทย ที่มีการเข้ารหัสตัวอักษรแบบผู้ใช้กำหนดการเข้ารหัสตัวอักษรเอง และแบบใช้ข้อกำหนดการเข้ารหัสในแฟ้มแบบอักษร จะเป็นปัญหาในการที่เครื่องมือการค้นข้อความไม่สามารถค้นข้อความได้ เนื่องจากไม่เข้าใจการเข้ารหัสที่แบบอักษรไทยใช้ จึงทำให้ถอดรหัสข้อความออกมาไม่ถูกต้อง

3.1.1 การเข้ารหัสของแบบอักษรไทยแบบผู้ใช้กำหนดการเข้ารหัสตัวอักษรเอง

จากการศึกษาพบว่า แบบอักษรไทยที่มีการเข้ารหัสโดยผู้ใช้กำหนดการเข้ารหัสตัวอักษรเอง เมื่อพิจารณาตามรหัสตัวอักษรและชื่อของตัวอักษรที่ใช้ในการเข้ารหัสแล้ว จะมีอยู่ด้วยกัน 4 ลักษณะ คือ

ลักษณะที่ 1 การเข้ารหัสของแบบอักษร มีรหัสตัวอักษรตรงกับมาตรฐาน มอก.620

การเข้ารหัสในลักษณะนี้ รหัสของตัวอักษรแต่ละตัวอักษรในเอกสารพีดีเอฟมีค่าคงที่ และรหัสตัวอักษรตรงตามข้อกำหนดมาตรฐาน มอก.620 แต่ชื่อตัวอักษรในแบบอักษรเดิมเปลี่ยนไป

ตัวอย่างเช่น ข้อความ "ABCDกขคงงคขกABCD" ใช้แบบอักษร AngsanaUPC ขนาด 14,18 ตามลำดับ ภายในเอกสารพีดีเอฟ จะใช้รหัสตัวอักษรและการเข้ารหัสตัวอักษร ดังนี้

F1(ABCD\241\242\244\247)TjTf

Encoding << 1/G01 /G02 /G03 .../GFE /GFF >>

F2(\241\242\244\247ABCD)TjTf

Encoding << 1/G01 /G02 /G03 .../GFE /GFF >>

จากตัวอย่างจะเห็นว่า ข้อความในเอกสาร ที่เป็นข้อความภาษาอังกฤษจะยังคงตัวอักษรเดิมที่ใช้ ส่วนข้อความไทยจะใช้เป็นรหัสตัวอักษรฐานแปด โดยรหัสตัวอักษร 'ก' คือ 241_8 รหัสตัวอักษร 'ข' คือ 242_8 รหัสตัวอักษร 'ค' คือ 244_8 และรหัสตัวอักษร 'ง' คือ 247_8

จะได้ว่า การเข้ารหัสในลักษณะนี้ รหัสตัวอักษรตรงตามข้อกำหนดมาตรฐาน มอก.620 (ดู ตารางที่ 6 ประกอบ)

ตารางที่ 6 ตารางการเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ
ที่รหัสตัวอักษรตรงตามมาตรฐาน มอก.620

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	o200	o201	o202	o203	o204	o205	o206	o207	o210	o211	o212	o213	o214	o215	o216	o217
	ฐ	ฑ	ฒ	ณ	ด	...	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ
9	o220	o221	o222	o223	o224	o225	o226	o227	o230	o231	o232	o233	o234	o235	o236	o237
	ฟ	ภ	ม	ย	ร	ฤ	ล	ฦ	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ
A	o240	o241	o242	o243	o244	o245	o246	o247	o250	o251	o252	o253	o254	o255	o256	o257
		ก	ข	ฃ	ค	ฅ	ฆ	ง	จ	ฉ	ช	ซ	ฌ	ญ	ฎ	ฏ
B	o260	o261	o262	o263	o264	o265	o266	o267	o270	o271	o272	o273	o274	o275	o276	o277
	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ	ฟ
C	o300	o301	o302	o303	o304	o305	o306	o307	o310	o311	o312	o313	o314	o315	o316	o317
	ภ	ม	ย	ร	ฤ	ล	ฦ	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ	ฯ
D	o320	o321	o322	o323	o324	o325	o326	o327	o330	o331	o332	o333	o334	o335	o336	o337
	ะ	ั	า	ำ	ิ	ี	ึ	ื	ุ	ู	ฺ		฿			
E	o340	o341	o342	o343	o344	o345	o346	o347	o350	o351	o352	o353	o354	o355	o356	o357
	เ	แ	โ	ใ	ไ	ๅ	ๆ	็	่	้	๊	๋	์	๎	๏	๐
F	o360	o361	o362	o363	o364	o365	o366	o367	o370	o371	o372	o373	o374	o375	o376	o377
	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	๚	๛				

ลักษณะที่ 2 การเข้ารหัสของแบบอักษรมีรหัสตัวอักษรลดลงจากมาตรฐาน มอก.620 ไปด้วยค่า 35_8

การเข้ารหัสในลักษณะนี้ รหัสตัวอักษรมีค่าคงที่ แต่เมื่อเปรียบเทียบกับมาตรฐาน มอก.620 จะลดลงไปด้วยค่า 35_8 และชื่อของตัวอักษรในแบบอักษรเดิมเปลี่ยนไป

ตัวอย่างเช่น ข้อความ "ABCDกขคงงคขกABCD" ใช้แบบอักษร AngsanaUPC ขนาด14,18 ตามลำดับ ภายในเอกสารพีดีเอฟ จะใช้รหัสตัวอักษรและการเข้ารหัสตัวอักษร ดังนี้

F1(\44\45\46\47\204\205\207\212)TjTf

Encoding << 1/G36 /G37 /G38 /G39 /G132 /G133 /G135 /G138 >>

F2(\212\207\205\204\44\45\46\47)TjTf

Encoding << 1/G36 /G37 /G38 /G39 /G132 /G133 /G135 /G138 >>

จะเห็นว่า รหัสตัวอักษรของตัวอักษร 'A' ตามมาตรฐานมอก.620 คือ 101_8

การเข้ารหัสในลักษณะนี้ ตัวอักษร 'A' ใช้รหัสตัวอักษร 44_8

จะได้ผลต่างในการเข้ารหัสใหม่คือ ($101_8 - 44_8$) = ลดลงจากเดิม 35_8

รหัสตัวอักษรของตัวอักษร 'ก' ตามมาตรฐานมอก.620 คือ 241_8

การเข้ารหัสในลักษณะนี้ ตัวอักษร 'ก' ใช้รหัสตัวอักษร 204_8

จะได้ผลต่างในการเข้ารหัสใหม่คือ ($241_8 - 204_8$)= ลดลงจากเดิม 35_8

การเข้ารหัสแบบนี้ รหัสตัวอักษรจะลดลงไปด้วยค่า 35_8 (ดูตารางที่ 7 ประกอบ)

ตารางที่ 7 ตารางการเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ
ที่รหัสตัวอักษรลดลงจากมาตรฐาน มอก.620

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
6	o140	o141	o142	o143	o144	o145	o146	o147	o150	o151	o152	o153	o154	o155	o156	o157
				ฃ	ฅ	ฆ	ง	จ	...	'	้	๊	๋	์	'	้
7	o160	o161	o162	o163	o164	o165	o166	o167	o170	o171	o172	o173	o174	o175	o176	o177
	๊	๋	์	ฌ	'	'	"	"	◆	—	——	็	๋	์	'	้
8	o200	o201	o202	o203	o204	o205	o206	o207	o210	o211	o212	o213	o214	o215	o216	o217
	็	๋	์		ก	ข	ฃ	ค	ฅ	ฆ	ง	จ	ฉ	ช	ซ	ฌ
9	o220	o221	o222	o223	o224	o225	o226	o227	o230	o231	o232	o233	o234	o235	o236	o237
	ญ	ฎ	ฏ	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ
A	o240	o241	o242	o243	o244	o245	o246	o247	o250	o251	o252	o253	o254	o255	o256	o257
	ฝ	พ	ฟ	ภ	ม	ย	ร	ฤ	ล	ฦ	ว	ศ	ษ	ส	ห	ฬ
B	o260	o261	o262	o263	o264	o265	o266	o267	o270	o271	o272	o273	o274	o275	o276	o277
	อ	ฮ	ฯ	ะ	ั	า	ำ	ิ	ี	ึ	ื	ุ	ู	ฺ	฻	฼
C	o300	o301	o302	o303	o304	o305	o306	o307	o310	o311	o312	o313	o314	o315	o316	o317
			฿	เ	แ	โ	ใ	ไ	ๅ	ๆ	็	่	้	๊	๋	์
D	o320	o321	o322	o323	o324	o325	o326	o327	o330	o331	o332	o333	o334	o335	o336	o337
	ํ	๎	๏	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	๚	๛	๜
E	o340	o341	o342	o343	o344	o345	o346	o347	o350	o351	o352	o353	o354	o355	o356	o357
	๝	๞														

<u>ลักษณะที่ 3 การเข้ารหัสแบบอักษรมีการเข้ารหัสตัวอักษรแต่ละครั้งไม่คงที่ แต่ชื่อตัวอักษรคงที่</u>

การเข้ารหัสในลักษณะนี้ รหัสตัวอักษรในเอกสารพีดีเอฟไม่คงที่ ขึ้นอยู่กับลำดับของตัวอักษรในเอกสารของแบบอักษรนั้น แต่ชื่อตัวอักษรในการเข้ารหัสคงที่

ตัวอย่างเช่น ข้อความ "ABCDกขคงงคขกABCD" ใช้แบบอักษร AngsanaUPC ขนาด14,18 ตามลำดับ ภายในเอกสารพีดีเอฟ จะใช้รหัสตัวอักษรและการเข้ารหัสตัวอักษร ดังนี้

F1(ABCD \001\002\003\004)TjTf

Encoding << 1/afii59681 /afii59682; /afii59684; /afii59687; >>

F2/\001\002\003\004 ABCD)TjTf

Encoding << 1/afii59687; /afii59684; /afii59682; /afii59681; >>

จากข้อมูลการเข้ารหัสข้างต้น จะได้ว่า

รหัสตัวอักษรไทยจะขึ้นอยู่กับลำดับตัวอักษรที่ปรากฏในเอกสารนั้นๆ พิจารณาที่ ตัวอักษร 'ก' เป็นอักษรตัวแรกของแบบอักษรขนาด 14 ในเอกสารพีดีเอฟ ตัวอักษร 'ก' จะใช้รหัสตัวอักษร 001 เข้ารหัสตัวอักษร "afii59681" ในแบบอักษรขนาด 18 ตัวอักษร 'ก' เป็นตัวอักษรลำดับที่ 4 ในเอกสารพีดีเอฟ ตัวอักษร 'ก' จะใช้รหัสตัวอักษร 004 เข้ารหัสตัวอักษร "afii59681 "

จะเห็นว่า การเข้ารหัสแต่ละครั้ง รหัสตัวอักษรของ 'ก' ไม่คงที่ แต่ชื่อของตัวอักษรที่ใช้ในการเข้ารหัสคงที่ (ดูตารางที่ 8 ประกอบ)

ตารางที่ 8 ตารางชื่อตัวอักษรที่แบบอักษรไทยใช้ในเอกสารพีดีเอฟ

	0	1	2	3	4	5	6	7
128	f700 ฐ	f701;	f702;	f703;	f704,	ellipsis; …	f705;	f706;
136	f707;	f708;	f709;	f70a;	f70b;	f70c;	f70d;	f70e;
144	f70f; ญ							
152	f710:	f711:	f712:		f714:	f715:	f716:	f717:
160		afii59681 ก	afii59682; ข	afii59683; ฃ	afii59684; ค	afii59685; ฅ	afii59686; ฆ	afii59687; ง
168	afii59688; จ	afii59689; ฉ	afii59690; ช	afii59691; ซ	afii59692; ฌ	afii59693; ญ	afii59694; ฎ	afii59695; ฏ
176	afii59696; ฐ	afii59697, ฑ	afii59698; ฒ	afii59699; ณ	afii59700; ด	afii59701; ต	afii59702; ถ	afii59703; ท
184	afii59704; ธ	afii59705; น	afii59706, บ	afii59707; ป	afii59708; ผ	afii59709; ฝ	afii59710; พ	afii59711; ฟ
192	afii59712, ภ	afii59713, ม	afii59714, ย	afii59715; ร	afii59716, ฤ	afii59717; ล	afii59718; ฦ	afii59719; ว
200	afii59720; ศ	afii59721; ษ	afii59722; ส	afii59723; ห	afii59724; ฬ	afii59725; อ	afii59726; ฮ	afii59727, ฯ
208	afii59729; ะ	afii59728; ั	afii59730; า	afii59731; ำ	afii59732; ิ	afii59733; ี	afii59734; ึ	afii59735; ื
216	afii59736; ุ	afii59737; ู	afii59738; ฺ	afii59739;	afii59740;	afii59741;	afii59742;	afii59743; ฿
224	afii59744; เ	afii59745; แ	afii59746; โ	afii59747; ใ	afii59748; ไ	afii59749; ๅ	afii59750; ๆ	afii59751; ็
232	afii59752; ่	afii59753, ้	afii59754, ๊	afii59755; ๋	afii59756; ์	afii59757; ํ	afii59758; ๎	afii59759; ๏
240	afii59760, ๐	afii59761, ๑	afii59762; ๒	afii59763; ๓	afii59764; ๔	afii59765; ๕	afii59766; ๖	afii59767; ๗
248	afii59768; ๘	afii59769; ๙	afii59770; ๚	afii59771; ๛	f718;	f719;	f71a;	

ลักษณะที่ 4 การเข้ารหัสของแบบอักษรมีการเข้ารหัสตัวอักษรแต่ละครั้ง รหัสตัวอักษรและชื่อตัวอักษรไม่คงที่

การเข้ารหัสในลักษณะนี้ รหัสตัวอักษรในเอกสารพีดีเอฟไม่คงที่ขึ้นอยู่กับลำดับของตัวอักษรในเอกสารของแบบอักษรนั้น และชื่อตัวอักษรในการเข้ารหัสถูกเปลี่ยนไปจากชื่อที่ใช้ในแบบอักษรเดิมและชื่อที่ถูกเปลี่ยนไปนั้นไม่คงที่

ตัวอย่างเช่น ข้อความ "ABCDกขคงงคขกABCD" ใช้แบบอักษร AngsanaUPC ขนาด14,18 ตามลำดับ

ภายในเอกสารพีดีเอฟ จะใช้รหัสตัวอักษรและการเข้ารหัสตัวอักษร ดังนี้

F1(\001\002\003\004\005\006\007\010)TjTf

Encoding << 1/G01 /G02 /G03 /G04 /G05 /G06 /G07 /G08 >>

F2(\001\002\003\004\005\006\007\010)TjTf

Encoding << 1/G01 /G02 /G03 /G04 /G05 /G06 /G07 /G08 >>

จากตัวอย่างข้างต้น จะได้ว่า

ทั้งข้อความภาษาไทยและภาษาอังกฤษ รหัสตัวอักษรจะถูกเปลี่ยนไปขึ้นอยู่กับลำดับตัวอักษรที่ปรากฏในเอกสารนั้นๆ เช่น ตัวอักษร 'A' เป็นอักษรตัวแรกของแบบอักษรขนาด 14 เมื่อแปลงมาอยู่ในรูปเอกสารพีดีเอฟ ตัวอักษร 'A' จะใช้รหัสตัวอักษร 001 เข้ารหัสตัวอักษร /G01 ในแบบอักษรขนาด 18 ตัวอักษร 'A' เป็นตัวอักษรลำดับที่ 5 ในแบบอักษร ตัวอักษร 'A' ในแบบอักษรนี้ จะใช้รหัสตัวอักษร 005 เข้ารหัสตัวอักษร /G05 ตัวอักษรภาษาไทยก็เช่นกัน ตัวอักษร 'ก' เป็นอักษรลำดับที่ 5 ของแบบอักษรขนาด 14 เมื่อแปลงมาอยู่ในรูปเอกสารพีดีเอฟ ตัวอักษร 'ก' จะใช้รหัสตัวอักษร 005 เข้ารหัสตัวอักษร /G05 ในแบบอักษรขนาด 18 ตัวอักษร 'ก' เป็นตัวอักษรลำดับที่ 4 ในแบบอักษร ตัวอักษร 'ก' ในแบบอักษรนี้ จะใช้รหัสตัวอักษร 004 เข้ารหัสตัวอักษร /G04 จะเห็นว่า การเข้ารหัสแต่ละครั้งรหัสตัวอักษรและชื่อของตัวอักษรไม่คงที่

3.1.2 การเข้ารหัสแบบอักษรไทยที่ใช้ข้อกำหนดการเข้ารหัสในแฟ้มแบบอักษร

โดยปกติเอกสารพีดีเอฟ จะต้องมีข้อมูลการเข้ารหัสระบุไว้ในพจนานุกรมแบบอักษร ถ้าไม่มีข้อมูลการเข้ารหัสในพจนานุกรมแบบอักษร ให้ใช้การเข้ารหัสที่อยู่ในแฟ้มแบบอักษรนั้น แบบอักษรไทยที่มีการเข้ารหัสแบบนี้ จะใช้การเข้ารหัสตามข้อกำหนดมาตรฐานยูนิโค้ด (ดูตารางที่ 9 ประกอบ)

ตารางที่ 9 ตารางรหัสยูนิโค้ดที่แบบอักษรไทยใช้ในเอกสารพีดีเอฟ

	0	1	2	3	4	5	6	7
128	f700	f701;	f702;	f703;	f704;		f705;	f706;
136	f707;	f708;	f709;	f70a;	f70b;	f70c;	f70d;	f70e;
144	f70f;							
152	f710;	f711;	f712;	f713;	f714;	f715;	f716;	f717;
160		0E01 ก	0E02 ข	0E03 ฃ	0E04 ค	0E05 ฅ	0E06 ฆ	0E07 ง
168	0E08 จ	0E09 ฉ	0E0A ช	0E0B ซ	0E0C ฌ	0E0D ญ	0E0E ฎ	0E0F ฏ
176	0E10 ฐ	0E11 ฑ	0E12 ฒ	0E13 ณ	0E14 ด	0E15 ต	0E16 ถ	0E17 ท
184	0E18 ธ	0E19 น	0E1A บ	0E1B ป	0E1C ผ	0E1D ฝ	0E1E พ	0E1F ฟ
192	0E20 ภ	0E21 ม	0E22 ย	0E23 ร	0E24 ฤ	0E25 ล	0E26 ฦ	0E27 ว
200	0E28 ศ	0E29 ษ	0E2A ส	0E2B ห	0E2C ฬ	0E2D อ	0E2E ฮ	0E2F ฯ
208	0E30 ะ	0E31 ั	0E32 า	0E33 ำ	0E34 ิ	0E35 ี	0E36 ึ	0E37 ื
216	0E38 ุ	0E39 ู	0E3A ฺ	0E3B	0E3C	0E3D	0E3E	0E3F ฿
224	0E40 เ	0E41 แ	0E42 โ	0E43 ใ	0E44 ไ	0E45 ๅ	0E46 ๆ	0E47 ็
232	0E48 ่	0E49 ้	0E4A ๊	0E4B ๋	0E4C ์	0E4D ํ	0E4E ๎	0E4F ๏
240	0E50 ๐	0E51 ๑	0E52 ๒	0E53 ๓	0E54 ๔	0E55 ๕	0E56 ๖	0E57 ๗
248	0E58 ๘	0E59 ๙	0E5A ๚	0E5B ๛	f718;	f719;	f71a;	

3.2 การถอดรหัสข้อความไทยในเอกสารพีดีเอฟ

จากหัวข้อที่ผ่านมาได้นำเสนอลักษณะการเข้ารหัสของแบบอักษรไทย ที่เป็นปัญหาในการค้นข้อความ ทำอย่างไรจึงจะถอดรหัสข้อความไทยที่ไม่เป็นไปตามข้อกำหนดมาตรฐานเหล่านั้นให้ได้ข้อความไทยที่ถูกต้องตรงตามข้อกำหนดมาตรฐาน มอก.620

จึงแบ่งปัญหาที่จะต้องแก้ไขในกระบวนการถอดรหัสข้อความไทยในเอกสารพีดีเอฟออกเป็นขั้นตอนดังนี้

- ทำอย่างไรจึงจะทราบว่า แบบอักษรไทยมีการเข้ารหัสลักษณะใดตามที่ได้กล่าวมา
- ทำการถอดรหัสข้อความอย่างไร จึงจะได้รหัสอักขระไทยที่ถูกต้องตามมาตรฐาน มอก.620

จากปัญหาข้างต้น ออกแบบกระบวนการถอดรหัสข้อความไทยในเอกสารพีดีเอฟ โดยให้มีขั้นตอนดังนี้

1. วิเคราะห์ลักษณะการเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ
2. คัดแยกข้อความในเอกสารพีดีเอฟแล้วแปลงให้เป็นรหัสที่ตรงตามมาตรฐาน มอก.620

3.2.1 การวิเคราะห์ลักษณะการเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ

เนื่องจาก การเข้ารหัสของแบบอักษรไทยตามที่ได้กล่าวมาในหัวข้อการเข้ารหัสไม่มีระบุไว้ในแฟ้มเอกสารพีดีเอฟ ทำอย่างไรจึงจะทราบว่าเอกสารพีดีเอฟนี้มีการเข้ารหัสตามลักษณะใด จากการศึกษาเอกสารพีดีเอฟภาษาไทยที่ใช้กันทั่วไปในระบบเครือข่ายอินเตอร์เน็ต ได้แก่ วารสารตัวอย่างเพื่อใช้ในการโฆษณาของอ.ส.ม.ท (www.mcot.or.th/v_bookworld1) จำนวน 79 เอกสาร ตัวอย่างนิตยสารขวัญเรือน (www.thaimag.com/kwanruen) จำนวน 15 เอกสาร รายงานของนิสิตภาควิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเกษตรศาสตร์ ประกอบวิชาสัมมนา (www.cpc.ku.ac.th/~semina) จำนวน 35 เอกสาร จดหมายข่าวของศูนย์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย (www.eng.chula.ac.th) จำนวน 8 เอกสาร เอกสารของสมาคมวิศวกรรมยานยนต์ประเทศไทย (www.eng.chula.ac.th/~mech/tsae) จำนวน 10 เอกสาร บทความทางการแพทย์ของ นายสุรชัย อัญเชิญ ภาควิชาเภสัชวิทยา คณะเภสัชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย (www.pharm.chula.ac.th/surachai) จำนวน 2 เอกสาร บทความสอนการใช้งานคอมพิวเตอร์ของ นายสรวิชย์ พลสิทธ์ ชุมนุมคอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าลาดกระบัง (www.kmitl.ac.th/~oom) จำนวน 2 เอกสาร เอกสารคู่มือการใช้งานเครือข่ายอินเตอร์เน็ตของศูนย์คอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย (www.it.chula.ac.th) จำนวน 1 เอกสาร และจากผู้วิจัยดำเนินการทดลองสร้าง จะได้ว่า รายละเอียดในข้อมูลแบบอักษรและข้อมูลสภาพแวดล้อมบางส่วนในเอกสาร สามารถใช้ระบุลักษณะการเข้ารหัสของแบบอักษรไทยได้ แต่วิธีการนี้ไม่สามารถนำมาวิเคราะห์การเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟได้ทุกเอกสาร ในบางเอกสารก็มีรายละเอียดข้อมูลไม่เพียงพอที่จะใช้วิธีนี้ทำการวิเคราะห์ จึงแบ่งกระบวนการการวิเคราะห์ลักษณะการเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ ออกเป็น 2 ขั้นตอน คือ

ขั้นตอนที่1 กระบวนการวิเคราะห์การเข้ารหัสแบบอักษรไทยในเอกสารที่มีข้อมูลเพียงพอในการวิเคราะห์การเข้ารหัส

ขั้นตอนนี้อาศัย ข้อมูลชื่อแบบอักษร ประเภทแบบอักษร พจนานุกรมแบบอักษร เครื่องมือในการสร้างเอกสารพีดีเอฟ มาวิเคราะห์การเข้ารหัสแบบอักษร ได้ดังนี้ (ดูตารางที่ 10 ประกอบ)

ตารางที่ 10 ตารางข้อมูลที่ใช้ในการวิเคราะห์การเข้ารหัสชุดแบบอักษรไทยในเอกสารพีดีเอฟ

ข้อมูลที่ใช้ในการวิเคราะห์การเข้ารหัสชุดแบบอักษรไทยในเอกสารพีดีเอฟ				ลักษณะการเข้ารหัส				
ชื่อชุดแบบอักษร	ประเภทชุดแบบอักษร	สภาพแวดล้อมในการสร้าง	องค์ประกอบอื่นๆ	มอก.620	ยูนิโค้ด	ลดลงด้วยค่าคงที่	ชื่ออักษรคงที่	รูปแบบไม่คงที่
ลงท้ายด้วย "UPC"	ทรูไทป์	ดิสทิลเลอร์รุ่นที่ 3		X				
ลงท้ายด้วย "UPC"	ทรูไทป์	ดิสทิลเลอร์รุ่นที่ 4			X			
		ดิสทิลเลอร์รุ่นที่ 3	ที่ข้อมูลผู้ผลิต(Producer) ระบุว่าสร้างโดย ผลิตภัณฑ์ของอะโดบี เช่น Adobe PageMaker	X				
			ในตารางแสดงความกว้างลำดับตัวอักษรตัวแรกเท่ากับ 32 และตัวสุดท้ายเท่ากับ 255	X				
ขึ้นต้นด้วย "MSTT"	ทรูไทป์	ดิสทิลเลอร์รุ่นที่ 3				X		
ลงท้ายด้วย "UPC"	ประเภทที่ 1	ดิสทิลเลอร์รุ่นที่ 3				X		
			ชื่ออักษรในการเข้ารหัสเริ่มต้นด้วย "afii59"				x	
ลงท้ายด้วย "UPC"	ประเภทที่ 1		รหัสตัวอักษรเริ่มต้นที่ 001					X

จาก ตารางที่ 10 จะได้ว่า

- เอกสารที่สร้างโดยผลิตภัณฑ์ ของบริษัท Adobe Inc เช่น PageMaker เข้ารหัสแบบผู้ใช้กำหนดเอง จะได้การเข้ารหัสชุดแบบไทยที่มีรหัสตัวอักษรตรงมาตรฐาน มอก.620
- ตารางแสดงความกว้างในแบบอักษร ลำดับตัวอักษรตัวแรกเริ่มต้นที่ ลำดับตัวที่ 32 และลำดับตัวอักษรตัวสุดท้ายเท่ากับ 255 จะได้ การเข้ารหัสชุดแบบไทยที่มีรหัสตัวอักษรตรงมาตรฐาน มอก.620
- ชื่อแบบอักษรคงชื่อเดิม แบบอักษรเป็นประเภททรูไทป์ สร้างโดย Distiller รุ่นที่ 4 ไม่พบการเข้ารหัสในเอกสารพีดีเอฟ จะได้ การเข้ารหัสชุดแบบไทยที่ใช้ข้อกำหนดการเข้ารหัสตามมาตรฐานยูนิโค้ด
- ชื่อแบบอักษรคงชื่อเดิม แบบอักษรเป็นประเภทที่ 1 เข้ารหัสแบบผู้ใช้กำหนดเอง สร้างโดย Distiller รุ่นที่ 3 จะได้ การเข้ารหัสชุดแบบไทยที่มีรหัสตัวอักษรลดลงจากข้อกำหนดในมาตรฐาน มอก.620 ไปด้วยค่า 35_8
- ชื่อแบบอักษรขึ้นต้นด้วย MSTT แบบอักษรเป็นประเภททรูไทป์ สร้างโดย Distiller รุ่นที่ 3 จะได้ การเข้ารหัสชุดแบบไทยที่มีรหัสตัวอักษรลดลงจากข้อกำหนดในมาตรฐาน มอก.620 ไปด้วยค่า 35_8
- ชื่อตัวอักษรในข้อมูลการเข้ารหัส ใช้ชื่อตัวอักษรที่เริ่มต้นด้วย "afii59" จะได้ การเข้ารหัสชุดแบบไทยที่รหัสตัวอักษรไม่คงที่ แต่ชื่อตัวอักษรในข้อมูลการเข้ารหัสคงที่
- รหัสตัวอักษรตัวแรกเริ่มต้นด้วย รหัสตัวอักษร 001 และใช้ชื่ออักขระ "G01" จะได้ การเข้ารหัสชุดแบบไทยที่รหัสตัวอักษรและชื่อตัวอักษรในข้อมูลการเข้ารหัสไม่คงที่

ขั้นตอนที่2 กระบวนการวิเคราะห์การเข้ารหัสแบบอักษรไทยในเอกสารที่มีข้อมูลไม่เพียงพอในการวิเคราะห์การเข้ารหัส

จะทำการวิเคราะห์โดยให้ผู้ใช้ช่วยทำการวิเคราะห์ โดยจะนำข้อความในเอกสารพีดีเอฟมาแสดงเป็นข้อความ 2 ข้อความ ข้อความแรกจะใช้รหัสตัวอักษรตามที่ใช้ในเอกสารพีดีเอฟ ข้อความที่สองจะบวกรหัสตัวอักษรด้วยค่า 35_8 ให้ผู้ใช้เลือกว่าสามารถอ่านข้อความใดได้ เพื่อตรวจสอบดูว่าเอกสารพีดีเอฟนี้ใช้การเข้ารหัสที่ตรงมาตรฐาน มอก.620 หรือลดลงจากมาตรฐาน มอก. 620

ตัวอย่างเช่น เอกสารพีดีเอฟแสดงข้อความ "สมรวย" เอกสารเข้ารหัสแบบอักษรในลักษณะที่รหัสตัวอักษรลดลงจากมาตรฐาน มอก.620 ด้วยค่า 35_8

ภายในเอกสารพีดีเอฟ จะใช้รหัสตัวอักษรและการเข้ารหัสตัวอักษร ดังนี้

F1(\255\244\246\252\245)TjTf

Encoding << 1/G164 /G167 /G168 /G170 /G173 >>

เมื่อวิเคราะห์การเข้ารหัสโดยใช้สภาพแวดล้อมในเอกสารพีดีเอฟแล้วพบว่าข้อมูลในเอกสารพีดีเอฟไม่เพียงพอในการวิเคราห์การเข้ารหัส จะนำรหัสตัวอักษรในเอกสารพีดีเอฟมาแสดงใน 2 ลักษณะ คือ

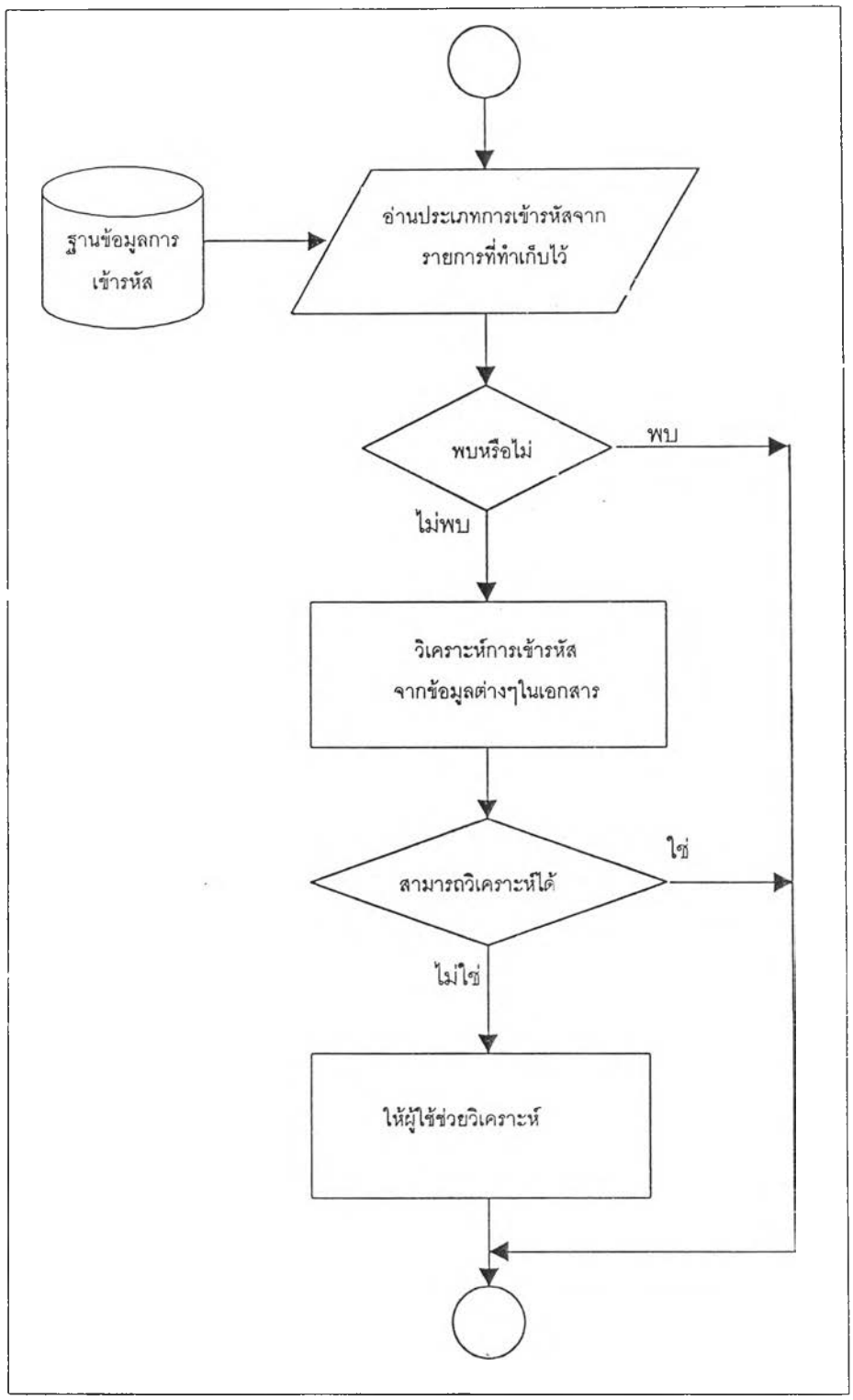
ข้อความที่ใช้รหัสตัวอักษรที่ใช้ในเอกสารพีดีเอฟ คือ " ญคจชง "

และข้อความที่บวกรหัสตัวอักษรด้วยค่า 35_8 คือ " สมรวย "

จะเห็นว่าข้อความที่สองสามารถอ่านได้ เอกสารนี้มีการเข้ารหัสแบบที่รหัสตัวอักษรลดลงจากมาตรฐาน มอก.620 ด้วยค่า 35_8

สาเหตุที่ใช้ข้อความเพียง 2 ข้อความ ทั้งนี้เนื่องจาก การเข้ารหัสแบบอักษรที่รหัสตัวอักษรไม่คงที่แต่ชื่อตัวอักษรในการเข้ารหัสคงที่ ชื่อตัวอักษรสามารถใช้ระบุการเข้ารหัสแบบนี้ได้ ถ้าไม่พบชื่อตัวอักษรที่ใช้ตรวจสอบก็จะไม่ใช่การเข้ารหัสแบบนี้ เอกสารที่เข้ารหัสโดยใช้การเข้ารหัสที่มีอยู่ในแฟ้มแบบอักษร ตรวจสอบก็จะไม่พบข้อมูลการเข้ารหัสในเอกสารพีดีเอฟ ทำให้การเข้ารหัสที่ยังไม่ทราบแน่ชัดจะเป็น การเข้ารหัสที่รหัสตัวอักษรตรงมาตรฐาน มอก.620 การเข้ารหัสที่รหัสตัวอักษรลดลงมาตรฐาน มอก.620 และการเข้ารหัสที่รหัสตัวอักษรและชื่อตัวอักษรไม่คงที่ จึงนำข้อความในเอกสารพีดีเอฟ มาแสดงโดยการเข้ารหัสใน 2 ลักษณะ คือ ใช้รหัสตัวอักษรที่ใช้ในเอกสารพีดีเอฟ ถ้าสามารถอ่านได้จะเป็นการเข้ารหัสที่รหัสตัวอักษรตรงมาตรฐาน มอก.620 แต่ถ้าข้อความที่บวกรหัสตัวอักษรด้วยค่า 35_8 สามารถอ่านได้จะเป็นการเข้ารหัสที่รหัสตัวอักษรลดลงจากมาตรฐาน มอก.620 ด้วยค่า 35_8

กระบวนการวิเคราะห์การเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ มีการทำงานดังนี้



รูปที่ 19 ผังงานการวิเคราะห์การเข้ารหัสแบบอักษรไทยในเอกสารพีดีเอฟ

3.2.2 การคัดแยกข้อความไทย แล้วแปลงรหัสอักขระให้ตรงตามข้อกำหนด มอก.620

เนื่องจากในเอกสารพีดีเอฟ จะประกอบด้วยข้อมูลทั้งที่เป็น ข้อความ รูปภาพ เสียง ภาพเคลื่อนไหว และอื่นๆ การถอดรหัสข้อความไทยในเอกสารพีดีเอฟ จะทำการแยกวัตถุข้อความออกมาจากเอกสารพีดีเอฟ เมื่อคัดแยกข้อความออกมาแล้ว จะทำการแปลงรหัสตัวอักษรของข้อความที่ได้ให้เป็นรหัสตัวอักษรที่ตรงตามข้อกำหนด มอก.620 โดยจะแบ่งกระบวนการถอดรหัสได้ดังนี้

การถอดรหัสข้อความไทยในเอกสารพีดีเอฟที่เข้ารหัสตัวอักษรตรงตามมาตรฐาน มอก.620

เอกสารพีดีเอฟที่มีการเข้ารหัสในลักษณะนี้ รหัสตัวอักษรจะตรงตามมาตรฐาน มอก.620 การถอดรหัสจะทำการเพียง ท่องเข้าไปในแฟ้มเอกสารพีดีเอฟเพื่อแยกข้อความออกมาจากวัตถุอื่นๆในเอกสาร แล้วนำรหัสตัวอักษรของข้อความเหล่านั้นออกมาใช้

การถอดรหัสข้อความไทยในเอกสารพีดีเอฟที่เข้ารหัสตัวอักษรลดลงจากมาตรฐาน มอก.620

การถอดรหัสเอกสารพีดีเอฟที่มีการเข้ารหัสลดลงด้วยค่าที่คงที่ เมื่อทำการคัดแยกข้อความออกมาจากเอกสารพีดีเอฟแล้ว จะนำค่ารหัสตัวอักษรเหล่านั้น ใช้เป็นตำแหน่งไปเปรียบเทียบในตารางเพื่อหารหัสตัวอักษร ที่ถูกต้อง (ดูตารางที่ 11 ประกอบ)

ตารางที่ 11 ตารางถอดรหัสข้อความไทยในเอกสารพีดีเอฟที่มี
การเข้ารหัสตัวอักษรลดลง จากมาตรฐาน มอก.620

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
6	o140	o141	o142	o143	o144	o145	o146	o147	o150	o151	o152	o153	o154	o155	o156	o157
				176	212	213	214	215	133	232	233	234	235	236	232	233
7	o160	o161	o162	o163	o164	o165	o166	o167	o170	o171	o172	o173	o174	o175	o176	o177
	234	235	236	173	145	146	147	148	149	150	151	209	153	231	232	233
8	o200	o201	o202	o203	o204	o205	o206	o207	o210	o211	o212	o213	o214	o215	o216	o217
	234	235	236		161	162	163	164	165	166	167	168	169	170	171	172
9	o220	o221	o222	o223	o224	o225	o226	o227	o230	o231	o232	o233	o234	o235	o236	o237
	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188
A	o240	o241	o242	o243	o244	o245	o246	o247	o250	o251	o252	o253	o254	o255	o256	o257
	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204
B	o260	o261	o262	o263	o264	o265	o266	o267	o270	o271	o272	o273	o274	o275	o276	o277
	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220
C	o300	o301	o302	o303	o304	o305	o306	o307	o310	o311	o312	o313	o314	o315	o316	o317
	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236
D	o320	o321	o322	o323	o324	o325	o326	o327	o330	o331	o332	o333	o334	o335	o336	o337
	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	216
E	o340	o341	o342	o343	o344	o345	o346	o347	o350	o351	o352	o353	o354	o355	o356	o357
	217	254														

การถอดรหัสข้อความไทยในเอกสารพีดีเอฟ ที่เข้ารหัสตัวอักษรในลักษณะที่รหัสตัวอักษรไม่คงที่ แต่ชื่อตัวอักษรในการเข้ารหัสคงที่

เอกสารที่มีการเข้ารหัสในลักษณะนี้ จะมีชื่อตัวอักษรคงที่ตลอดการเข้ารหัส การถอดรหัสจะใช้ชื่อตัวอักษรในการแปลงค่ารหัสตัวอักษรที่คัดแยกออกมาจากเอกสารพีดีเอฟ ไปใช้ค่ารหัสตัวอักษรที่ตรงตามมาตรฐาน มอก.620 เมื่อพิจารณาการเข้ารหัสด้วยชื่อตัวอักษรตามที่แสดงในตารางที่ 8 จะเห็นว่า ชื่อตัวอักษรที่ใช้ในแบบอักษรไทยจะแบ่งออกได้เป็น 2 กลุ่มด้วยกัน คือ กลุ่มที่ชื่อตัวอักษรขึ้นต้นด้วย "afii" และกลุ่มที่ชื่อตัวอักษรขึ้นต้นด้วย "f7"

กลุ่มที่ชื่อตัวอักษรขึ้นต้นด้วย "afii" จะเริ่มต้นที่รหัสตัวอักษร 161 ใช้ชื่อตัวอักษร "afii59681" เป็นลำดับ ไปจนถึงรหัสตัวอักษร 251 ใช้ชื่อตัวอักษร "afii59771" การถอดรหัสตัวอักษรของกลุ่มที่ชื่อตัวอักษรขึ้นต้นด้วย "afii" สามารถทำได้ดังนี้ นำค่าตัวเลข 3 ตัวหลังของชื่อ ลบด้วย 520 จะได้รหัสตัวอักษรที่ตรงตามมาตรฐาน มอก.620

กลุ่มที่ชื่อตัวอักษรขึ้นต้นด้วย "f7" จะนำค่าตัวเลขที่ต่อท้าย "f7" สร้างเป็นตารางอ้างอิงเพื่อใช้สำหรับในการแปลงค่ารหัสตัวอักษร โดยค่าตัวเลขที่ต่อท้าย "f7" จะเป็นตำแหน่งที่เก็บค่ารหัสตัวอักษร และในตารางจะเก็บค่ารหัสตัวอักษรที่ตรงตามมาตรฐาน มอก.620 เมื่อทำการคัดแยกข้อความออกมาจากเอกสารพีดีเอฟ จะดูว่ารหัสตัวอักษรนี้เข้ารหัสด้วยชื่ออักษรว่าอย่างไร ถ้าเป็นชื่อที่เริ่มต้นด้วย "f7" จะนำค่า 2 ตัวเลขที่อยู่ด้านหลัง "f7" เป็นค่าตำแหน่งของตารางอ้างอิงไปหาค่ารหัสตัวอักษรในตาราง (ดูตารางที่ 12 ประกอบ)

ตารางที่ 12 ตารางถอดข้อความไทยในเอกสารพีดีเอฟ
ที่แบบอักษรเข้ารหัสด้วยชื่อตัวอักษร "f7"

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
	176	212	213	214	215	232	233	234	235	236	232	233	234	235	236	173
1	16	17	18	19	20	21	22	23	24	25	26	27				
	209	254	231	232	233	234	235	236	216	217	254	219				

ตัวอย่างเช่น เอกสารพีดีเอฟแสดงข้อความ "กินดีมีเงิน" เข้ารหัสโดยชื่อตัวอักษรคงที่ ภายในเอกสารพีดีเอฟ จะใช้รหัสตัวอักษรและการเข้ารหัสตัวอักษร ดังนี้

F1(\001\002\003\004\005\006\005\007\010\002\003)TjTf

Encoding << 1/afii59681 /f701; /afii59705: /afii59700; /f702; /afii59713; /afii59744; /afii59687; >>

จะทำการถอดรหัสดังนี้ นำค่ารหัสตัวอักษร 001 ไปหาชื่อตัวอักษรในข้อมูลการเข้ารหัสของพจนานุกรมแบบอักษร จะได้ว่า รหัสตัวอักษรเข้ารหัสด้วยชื่อ afii59681 นำค่าตัวเลข 681 ลบด้วย 520 จะได้รหัสตัวอักษร 161 ('ก') ตัวอักษรถัดไป 002 เข้ารหัสด้วยชื่อ f701 เจอว่าชื่อตัวอักษรคือ f7 นำค่าตัวเลขที่ตามหลัง f7 คือ 01 ไปหาค่ารหัสตัวอักษรในตาราง จะได้รหัสตัวอักษร 212 (สระอิ)ทำการถอดรหัสอย่างนี้ต่อไปเรื่อยๆจนหมดข้อความ

จากที่ทราบแล้วว่าการเข้ารหัสตัวอักษรในลักษณะนี้จะทำการเข้ารหัสตัวอักษรเริ่มต้นที่ 001 002 เป็นลำดับไปเรื่อย ทำให้มีตัวอักษรบางตัวใช้ค่ารหัสอักขระ 13 หรือ 32 ซึ่งในความหมายของการเข้ารหัสตามมาตรฐานทั่วไป จะเป็นรหัสตัวอักษรควบคุม ค่ารหัสอักขระ 13 หมายถึง การขึ้นบรรทัดใหม่และค่ารหัสอักขระ 32 หมายถึง การเว้นวรรค ทำให้เมื่อท่องเข้าไปในเอกสารพีดีเอฟทำการคัดแยกข้อความออกมาจากเอกสารพีดีเอฟ ตัวอักษรที่ใช้ค่ารหัสอักขระ 13 และ 32 ซึ่งมีความหมายว่า การเว้นวรรค หรือการขึ้นบรรทัดใหม่ จะไม่ถูกทำการคัดแยกออกมาด้วย จึงทำให้การถอดรหัสข้อความในเอกสารที่มีการเข้ารหัสในลักษณะนี้ จะมีข้อจำกัดที่ไม่สามารถคัดตัวอักษรตัวที่ 13 และ ตัวที่ 32 ออกมาได้

<u>การถอดรหัสข้อความไทยในเอกสารพีดีเอฟที่ไม่มีข้อมูลการเข้ารหัสในพจนานุกรมแบบอักษร</u>

การถอดรหัสข้อความไทยในเอกสารพีดีเอฟที่มีลักษณะเช่นนี้ จะทำการถอดรหัสโดยคัดแยกข้อความในเอกสารพีดีเอฟออกมาเป็นรหัสยูนิโค้ด รหัสยูนิโค้ดจะใช้รหัสขนาด 2 ไบต์ เช่น 0E01, 0E02 เป็นต้น การถอดรหัสข้อความจะใช้ค่ารหัสไบท์หลังของรหัสยูนิโค้ดในการแปลงเป็นรหัส มอก.620 วิธีการถอดรหัสยูนิโค้ดที่เริ่มต้นด้วย E0 จะนำค่ารหัสไบต์หลังของรหัสยูนิโค้ดที่พบแปลงจากตัวเลขฐานสิบหกเป็นตัวเลขฐานสิบ บวกด้วย 160 จะได้รหัสตัวอักษรที่ตรงตามมาตรฐาน มอก.620 วิธีการถอดรหัสยูนิโค้ดที่เริ่มต้นด้วย f7 จะทำเช่นเดียวกับการถอดรหัสด้วยชื่อตัวอักษรที่ขึ้นต้นด้วย "f7" คือ จะนำค่ารหัสไบต์หลังของยูนิโค้ดที่เริ่มต้นด้วย f7 ไปหาค่ารหัสตัวอักษรในตาราง

ตัวอย่าง เอกสารพีดีเอฟแสดงข้อความ "กินดีมีเงิน"

ไม่มีข้อมูลการเข้ารหัสในพจนานุกรมแบบอักษร ทำการคัดแยกข้อความจะได้รหัสตัวอักษร

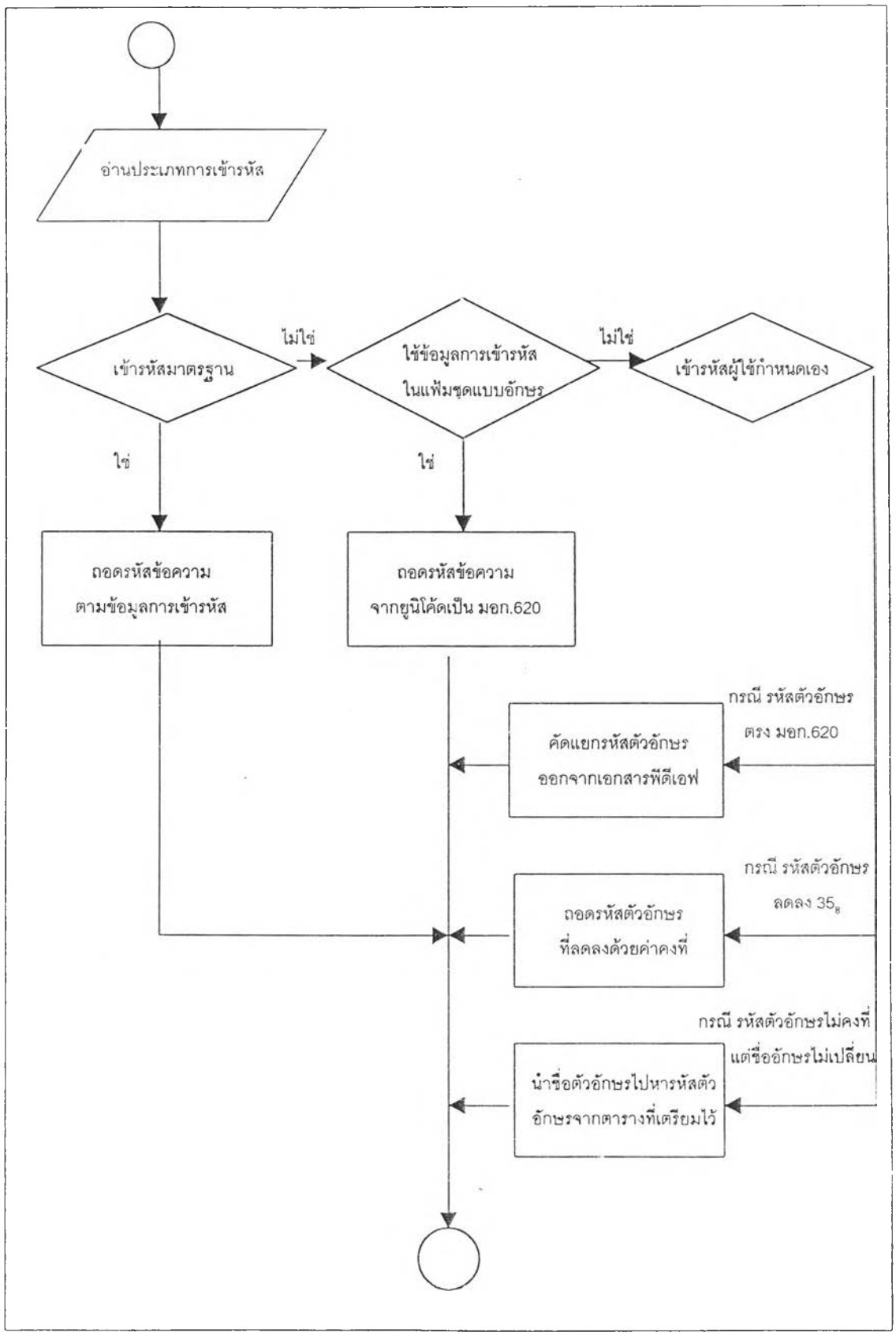
[0E01f701E0190E14f7020E21f7020E400E07f7010E19]

รหัสตัวอักษรตัวแรกคือ 0E ใช้รหัสตัวอักษรไบต์ถัดไป คือ 01 บวกด้วย 160 จะได้ 161 ('ก')

รหัสตัวอักษรตัวที่ 3 คือ f7 ใช้รหัสตัวอักษรไบต์ถัดไป คือ 01 หารหัสตัวอักษรจากตารางจะได้รหัสตัวอักษร 212 (สระอิ)

รหัสตัวอักษรตัวที่ 5 คือ 0E ใช้รหัสตัวอักษรไบต์ถัดไป คือ 25(19$_{16}$) บวกด้วย 160 จะได้ 185 ('น') ทำการถอดรหัสอย่างนี้ต่อไปเรื่อยๆจนหมดข้อความ

กระบวนการคัดแยกข้อความแล้วแปลงให้เป็นรหัสอักขระที่ตรงตามข้อกำหนด มอก.620



รูปที่ 20 ผังงานการคัดแยกข้อความแล้วแปลงให้เป็นรหัสอักขระที่ตรงตามข้อกำหนด มอก.620