

บทที่ 8

ผลการทดลองเปรียบเทียบประสิทธิภาพ

ในบทนี้ได้ทำการรวบรวม วิเคราะห์และเปรียบเทียบข้อมูลสถิติที่เกิดจากการทดลองวัดประสิทธิภาพของโปรแกรมตัดคำภาษาไทยที่ได้ทำไปในบทที่แล้วโดยได้ทำการทดลองกับวิธีการตัดคำ 6 วิธีดังนี้

- ก. แบบย้อนกลับคือเป็นแบบ Backtracking ของสัมพันธ์ (ดูบทที่ 6)
- ข. แบบเทียบคำยาวสุดคือแบบ Longest word pattern matching [10]
- ค. แบบเทียบคำสั้นสุดคือแบบ Shortest word pattern matching [10]
- ง. แบบใช้ความถี่ของคำคือแบบ Word usage frequency [8]
- จ. แบบใช้พจนานุกรมลดคำกำกวมคือแบบ Ambiguity dictionary (ดูบทที่ 6)
- ฉ. แบบเลือกคำน้อยที่สุด คือแบบ Maximal matching (ดูบทที่ 6 และ [14])

8.1 จำนวนคำที่ตัดออกมาถูกต้องและจำนวนคำที่ตัดออกมา

จากการทดลองทำให้ได้ผลจำนวนคำภาษาไทยที่โปรแกรมตัดคำตัดออกมาและถูกต้องกับจำนวนคำภาษาไทยทั้งหมดที่โปรแกรมตัดคำนั้น ๆ ตัดออกมาแสดงไว้ในตารางที่ 8.1 จากผลการทดลองที่ได้นี้จะนำไปทำการวิเคราะห์เปรียบเทียบให้เห็นคุณลักษณะที่ชัดเจนต่อไป

โปรแกรมตัดคำ / จำนวนบรรทัดในชุด	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000	6500
แบบย้อนกลับ	649/ 721	791/ 877	899/ 1014	941/ 1066	1017/ 1158	1042/ 1192	1079/ 1242	1142/ 1325	1173/ 1373	1197/ 1406	1238/ 1468	1268/ 1506	1279/ 1522
เทียบคำยาวสุด	609/ 879	840/ 1166	964/ 1417	1013/ 1550	1090/ 1716	1121/ 1820	1166/ 1945	1238/ 2114	1276/ 2242	1312/ 2360	1368/ 2499	1402/ 2586	1414/ 2668
เทียบคำสั้นสุด	434/ 1585	514/ 2265	578/ 2885	596/ 3247	631/ 3679	646/ 3969	665/ 4278	703/ 4641	719/ 4950	726/ 5201	751/ 5505	767/ 5754	773/ 5947
ใช้ความถี่ของคำ	572/ 906	687/ 1247	775/ 1540	806/ 1693	861/ 1900	884/ 2035	908/ 2173	954/ 2395	974/ 2558	992/ 2701	1025/ 2857	1048/ 2975	1056/ 3083
ใช้พจนานุกรมคำกำกวม	612/ 781	741/ 1070	837/ 1322	869/ 1436	931/ 1609	953/ 1697	983/ 1798	1034/ 1968	1051/ 2097	1066/ 2208	1100/ 2334	1122/ 2412	1130/ 2479
เลือกคำน้อยสุด	594/ 793	735/ 1032	833/ 1239	878/ 1344	956/ 1486	981/ 1563	1018/ 1653	1086/ 1801	1121/ 1894	1148/ 1978	1194/ 2088	1222/ 2145	1236/ 2227

ตารางที่ 8.1 จำนวนคำที่ตัดได้ถูกต้องและจำนวนคำที่ตัดออกมาทั้งหมด

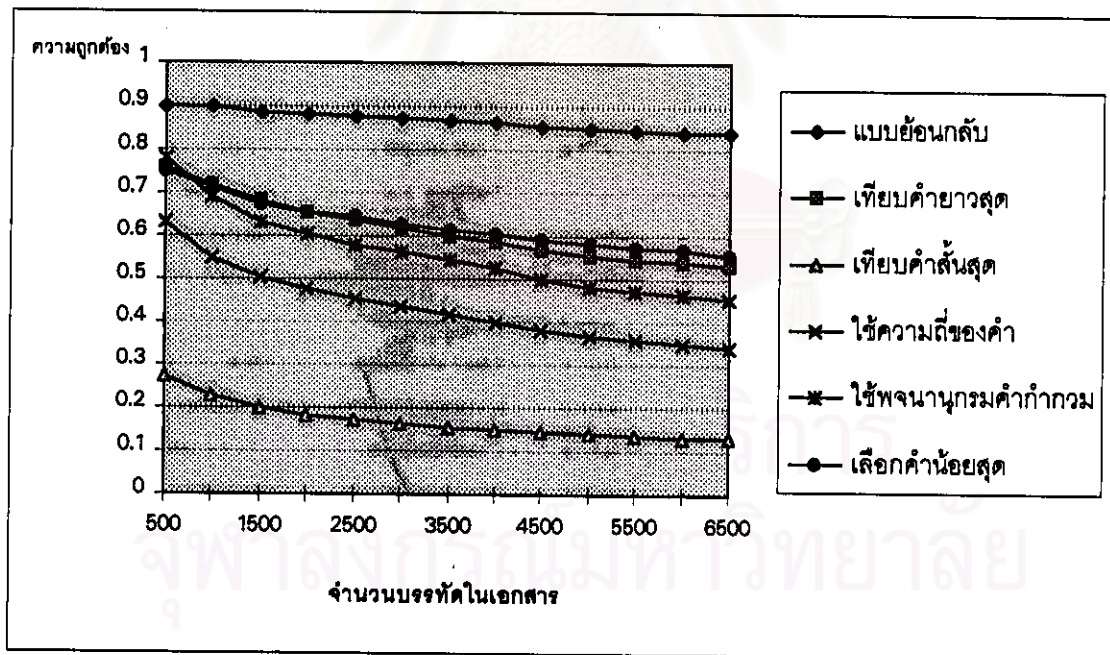
8.2 สัดส่วนความถูกต้องต่อจำนวนคำที่ตัดออกมา

สัดส่วนความถูกต้องของคำที่ตัดออกมาได้เปรียบเทียบกับจำนวนคำภาษาไทยทั้งหมดที่โปรแกรมตัดคำนั้น ๆ ตัดออกมามีค่าได้จาก นำจำนวนคำที่ตัดออกมาได้ถูกต้องหารด้วยจำนวนคำทั้งหมดที่โปรแกรมนั้น ๆ ตัดออกมาจะแสดงไว้ดังตารางที่ 8.2

เพื่อให้สังเกตเห็นความแตกต่างของผลการตัดคำได้ชัดเจนได้นำข้อมูลจากตาราง 8.2 ไปทำเป็นกราฟได้ดังรูปที่ 8.2 ซึ่งพบว่าโปรแกรมตัดคำแบบย้อนกลับ (Backtracking) ให้อัตราความถูกต้องนี้สูงที่สุดคืออยู่ในช่วง 0.84 ถึง 0.9

โปรแกรมตัดคำ / จำนวนบรรทัดอินพุต	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000	6500
แบบย้อนกลับ	0.9001	0.9019	0.8966	0.8927	0.8782	0.8742	0.8688	0.8619	0.8543	0.8514	0.844	0.842	0.8403
เทียบคำยาวสุด	0.7611	0.7204	0.6903	0.6595	0.6353	0.6159	0.5995	0.5856	0.5691	0.5559	0.5474	0.5422	0.53
เทียบคำสั้นสุด	0.2736	0.2269	0.2003	0.1836	0.1715	0.1628	0.1554	0.1515	0.1453	0.1396	0.1364	0.1333	0.13
ใช้ความถี่ของคำ	0.6313	0.5509	0.5032	0.4778	0.4532	0.4344	0.4179	0.3983	0.3808	0.3673	0.3588	0.3523	0.3425
ใช้พจนานุกรมคำกำกวม	0.7836	0.6925	0.6331	0.6043	0.5786	0.5616	0.5467	0.5254	0.5012	0.4888	0.4713	0.4652	0.4558
เลือกคำน้อยสุด	0.7491	0.7122	0.6723	0.6533	0.6433	0.6276	0.6158	0.603	0.5919	0.5804	0.5718	0.5667	0.555

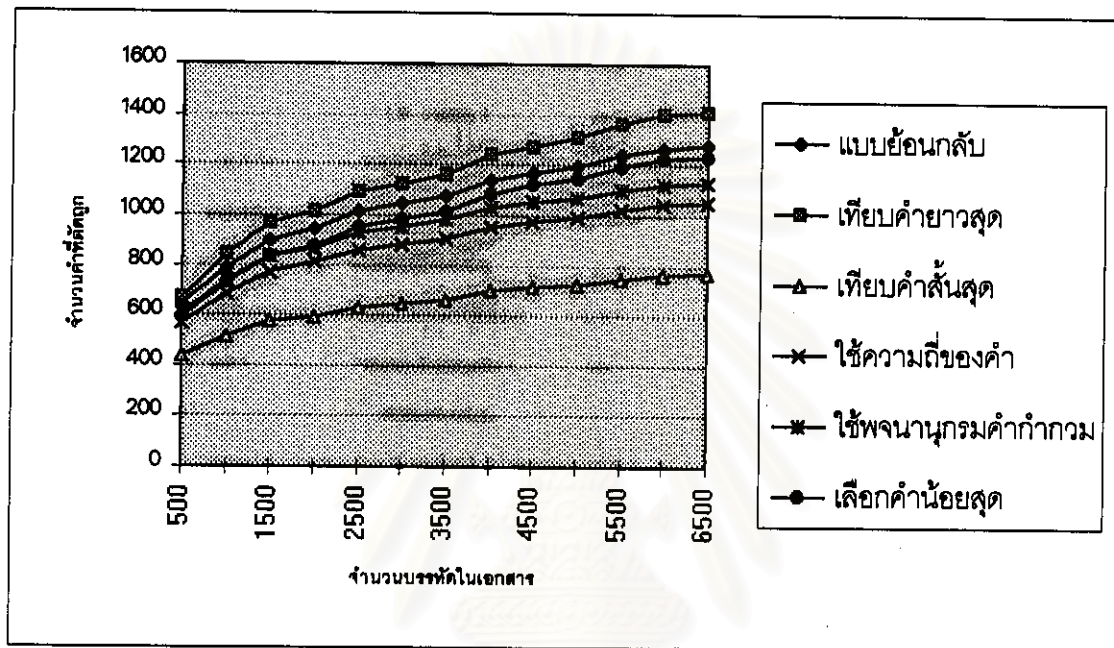
ตาราง 8.2 สัดส่วนความถูกต้องต่อจำนวนคำที่ตัดออกมา



รูป 8.2 สัดส่วนความถูกต้องต่อจำนวนคำที่ตัดออกมา

8.3 เปรียบเทียบจำนวนคำภาษาไทยที่ตัดออกมาได้ถูกต้อง

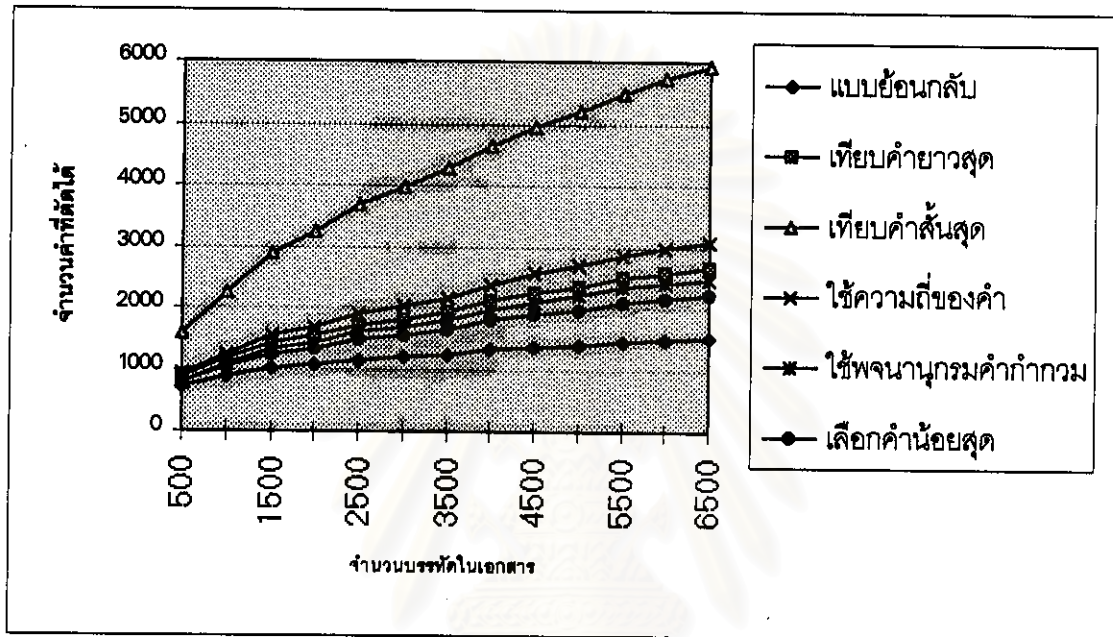
เมื่อนำข้อมูลค่าที่ตัดออกมาได้ถูกต้องในตารางที่ 8.1 มาทำเป็นกราฟดังในรูปที่ 8.3 จะเห็นว่าโปรแกรมตัดคำภาษาไทยแบบเทียบคำที่ยาวที่สุดได้จำนวนคำที่ตัดออกมาได้ถูกต้องมากที่สุด คืออยู่ระหว่าง 669 คำที่จำนวนข้อมูลอินพุท 500 บรรทัด ถึง 1,414 คำที่จำนวนข้อมูลอินพุท 6,500 บรรทัด



รูป 8.3 เปรียบเทียบจำนวนคำภาษาไทยที่ตัดออกมาได้ถูกต้อง

8.4 เปรียบเทียบจำนวนคำภาษาไทยที่ตัดออกมาได้ทั้งหมด

เมื่อนำข้อมูลค่าที่ตัดออกมาได้ทั้งหมดในตารางที่ 8.1 มาทำเป็นกราฟดังในรูปที่ 8.4 จะเห็นว่าโปรแกรมตัดคำภาษาไทยแบบเทียบค่าที่สั้นที่สุดได้จำนวนคำที่ตัดออกมาได้ทั้งหมดมากที่สุดคืออยู่ระหว่าง 1,585 คำที่จำนวนข้อมูลอินพุต 500 บรรทัด ถึง 5,947 คำที่จำนวนข้อมูลอินพุต 6,500 บรรทัด



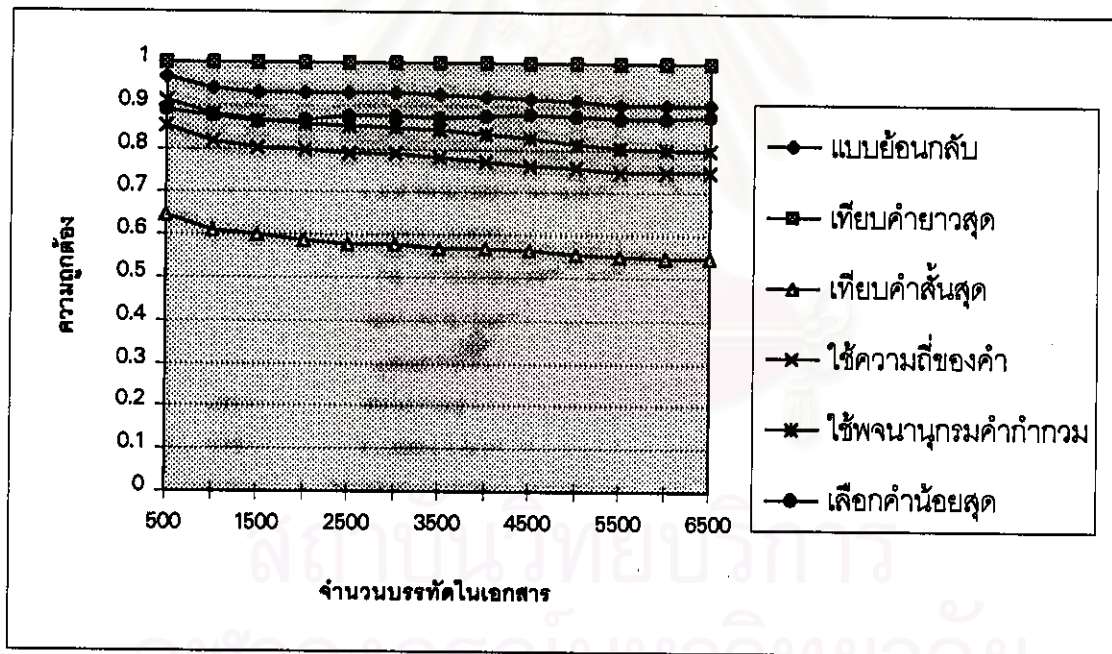
รูป 8.4 เปรียบเทียบจำนวนคำภาษาไทยที่ตัดออกมาได้ทั้งหมด

8.5 ความถูกต้องโดยใช้ผลของโปรแกรมแบบเทียบค่าที่ยาวที่สุดเป็นฐานอ้างอิง

สัดส่วนความถูกต้องของค่าที่ตัดออกมาได้ โดยใช้ผลของโปรแกรมแบบเทียบค่าที่ยาวที่สุด (Longest Pattern Matching) เป็นฐานอ้างอิงจะแสดงไว้ดังตารางที่ 8.3 และให้กราฟดังรูปที่ 8.5 จะเห็นได้ว่าโปรแกรมตัดค่าแบบย้อนกลับให้สัดส่วนความถูกต้องดีที่สุดใน นั่นคืออยู่ในช่วง 0.97 ที่ 500 บรรทัดจนถึง 0.90 ที่ 6,500 บรรทัดตามลำดับ

โปรแกรมตัดค่า / จำนวนบรรทัดในชุด	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000	6500
แบบย้อนกลับ	0.9701	0.9417	0.9285	0.9289	0.933	0.9289	0.9264	0.9225	0.9193	0.9183	0.9057	0.9044	0.9045
เทียบค่ายาวสุด	1	1	1	1	1	1	1	1	1	1	1	1	1
เทียบค่าสั้นสุด	0.6487	0.6119	0.5995	0.5864	0.5789	0.5753	0.5703	0.5679	0.5655	0.5654	0.549	0.5471	0.5407
ใช้ความถี่ของค่า	0.855	0.8179	0.8039	0.7969	0.7959	0.7868	0.7767	0.7706	0.7633	0.7561	0.7499	0.7476	0.7408
ใช้พจนานุกรมค่ากำกวม	0.9148	0.8931	0.8885	0.8878	0.8841	0.8801	0.8741	0.8689	0.8637	0.8585	0.8541	0.8503	0.8493
เลือกค่าน้อยสุด	0.8879	0.876	0.8641	0.8567	0.8771	0.8751	0.8781	0.8779	0.8789	0.876	0.8728	0.8719	0.8741

ตาราง 8.3 ความถูกต้องโดยใช้ผลของโปรแกรมแบบเทียบค่าที่ยาวที่สุดเป็นฐานอ้างอิง



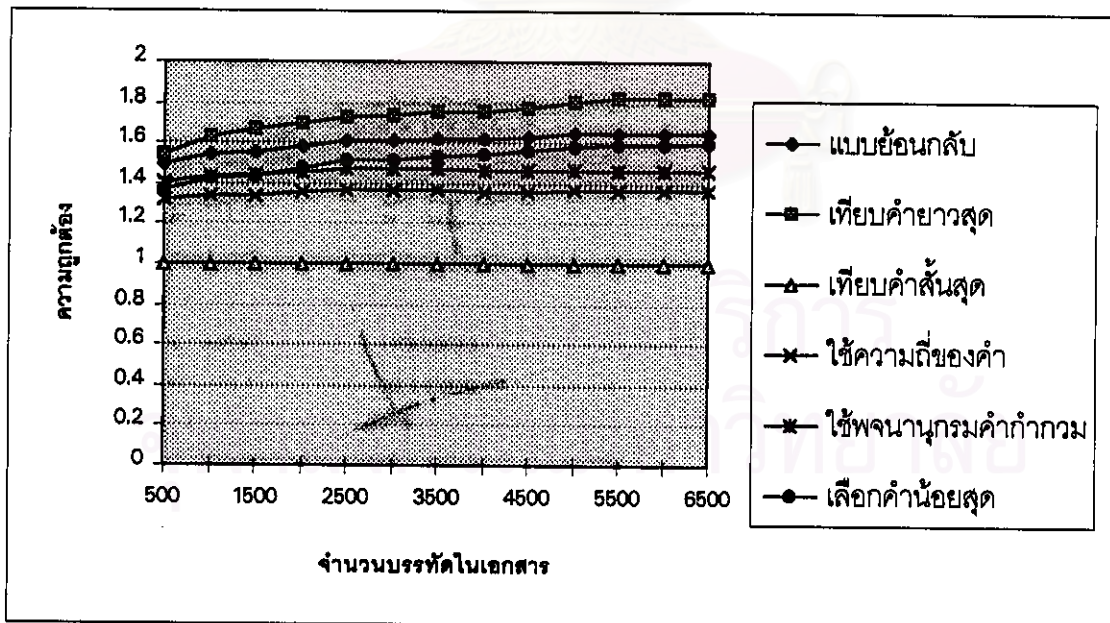
รูป 8.5 ความถูกต้องโดยใช้ผลของโปรแกรมแบบเทียบค่าที่ยาวที่สุดเป็นฐานอ้างอิง

8.6 ความถูกต้องโดยใช้ผลของโปรแกรมแบบเทียบค่าที่สั้นที่สุดเป็นฐานอ้างอิง

อัตราความถูกต้องของค่าที่ตัดออกมาได้โดยใช้ผลของโปรแกรมแบบเทียบค่าที่สั้นที่สุด (Shortest Pattern Matching) เป็นฐานอ้างอิงจะแสดงไว้ดังตารางที่ 8.4 และกราฟจะดังรูปที่ 8.6 ผลที่ได้พบว่าจำนวนค่าที่ตัดออกมาได้และถูกต้องมากที่สุดได้จากโปรแกรมตัดค่าแบบเทียบค่าที่ยาวที่สุดคืออยู่ในช่วง 1.54 ถึง 1.82 ที่ขนาดข้อมูล 500 และ 6,500 บรรทัดตามลำดับ เป็นที่น่าสังเกตว่าถ้าเปรียบเทียบกับวิธีการตัดค่าแบบเทียบค่าสั้นสุดแล้ว ความถูกต้องของการตัดค่าของโปรแกรมตัดค่าทุกแบบมีแนวโน้มที่ได้ความถูกต้องมากขึ้นเรื่อย ๆ เมื่อข้อมูลเพิ่มมากขึ้น ซึ่งชี้ให้เห็นว่าการตัดค่าแบบเทียบค่าที่สั้นที่สุดจะมีประสิทธิภาพที่ด้อยกว่าการตัดค่าแบบอื่น ๆ ทั้งหมด

โปรแกรมตัดค่า / จำนวนบรรทัดในชุด	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000	6500
แบบย้อนกลับ	1.4954	1.6389	1.6856	1.5789	1.6117	1.613	1.6285	1.6245	1.6314	1.6468	1.6498	1.6532	1.6645
เทียบค่ายาวสุด	1.6432	1.6342	1.6878	1.6897	1.7874	1.7503	1.7534	1.751	1.7747	1.8072	1.8216	1.8279	1.8292
เทียบค่าสั้นสุด	1	1	1	1	1	1	1	1	1	1	1	1	1
ใช้ความถี่ของค่า	1.318	1.3365	1.3408	1.3974	1.3645	1.3684	1.3684	1.357	1.3647	1.3664	1.3646	1.3656	1.3661
ใช้พจนานุกรมค่ากำกวม	1.6101	1.6418	1.6481	1.6881	1.6784	1.6782	1.6782	1.6708	1.6818	1.6883	1.6847	1.6888	1.6816
เลือกค่าน้อยสุด	1.3687	1.42	1.4412	1.4792	1.5191	1.5168	1.5308	1.5445	1.5581	1.5813	1.5889	1.5922	1.599

ตาราง 8.4 ความถูกต้องโดยใช้ผลของโปรแกรมแบบเทียบค่าที่สั้นที่สุดเป็นฐานอ้างอิง



รูป 8.6 ความถูกต้องโดยใช้ผลของโปรแกรมแบบเทียบค่าที่สั้นที่สุดเป็นฐานอ้างอิง

8.7 สัดส่วนความถูกต้องต่อจำนวนคำในพจนานุกรมที่ใช้งาน

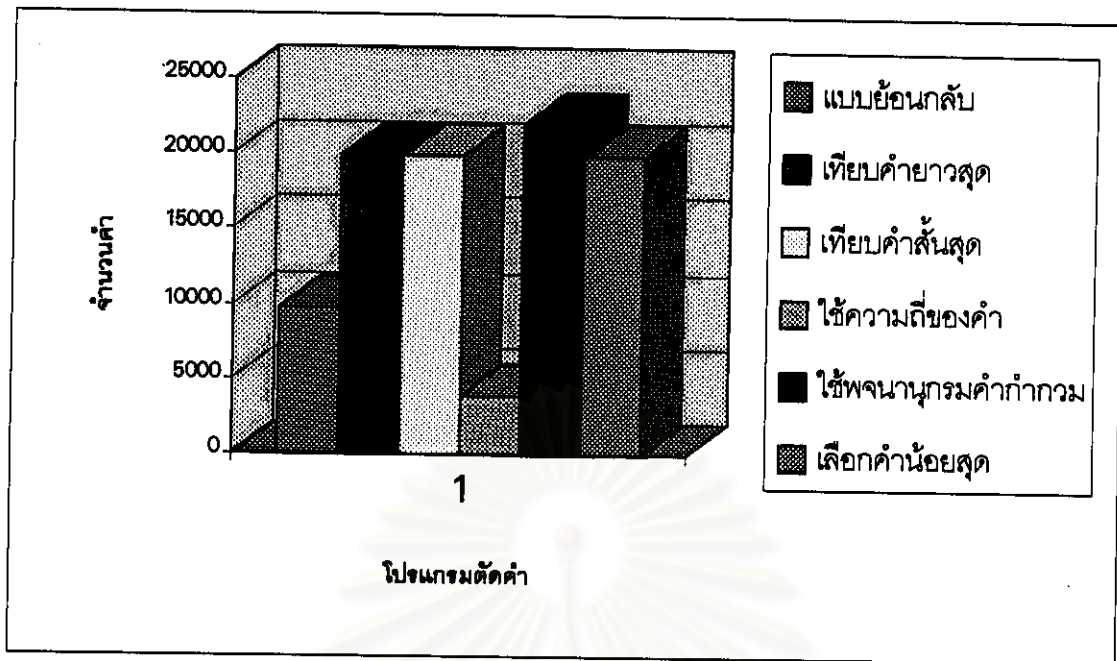
โปรแกรมตัดคำแต่ละตัวมีจำนวนคำที่เลือกใช้สำหรับเป็นพจนานุกรมไม่เท่ากัน การวัดความถูกต้องในข้อนี้คำนวณจากจำนวนคำที่ตัดออกมาได้ถูกต้องหารด้วยจำนวนคำที่ใช้เป็นพจนานุกรมสำหรับแต่ละโปรแกรม ความถูกต้องนี้จะชี้ให้เห็นประสิทธิภาพในการเลือกใช้คำที่จะมาเป็นพจนานุกรม ตาราง 8.5 และรูป 8.7 ต่อไปนี้เปรียบเทียบจำนวนคำที่ใช้เป็นพจนานุกรมของโปรแกรมตัดคำแบบต่าง ๆ จากผลการทดลองเมื่อคำนวณสัดส่วนความถูกต้องนี้แล้วได้ตั้งตารางที่ 8.6

ถ้านำข้อมูลจากตาราง 8.6 มาทำเป็นกราฟจะได้ดังรูป 8.7 ซึ่งพบว่าโปรแกรมตัดคำแบบใช้ความถี่ของคำจะให้ผลออกมาดีที่สุดคืออยู่ในช่วง 0.14 ถึง 0.27 เมื่อข้อมูลอินพุตอยู่ที่ 500 ถึง 6,500 บรรทัดตามลำดับ

โปรแกรมตัดคำ	จำนวนคำในพจนานุกรม
แบบย้อนกลับ	9,851
เทียบคำยาวสุด	17,859
เทียบคำสั้นสุด	17,859
ใช้ความถี่ของคำ	3,955
ใช้พจนานุกรมคำกำกวม	21,990
เลือกค่าน้อยสุด	17,859

ตาราง 8.5 เปรียบเทียบจำนวนคำในพจนานุกรม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

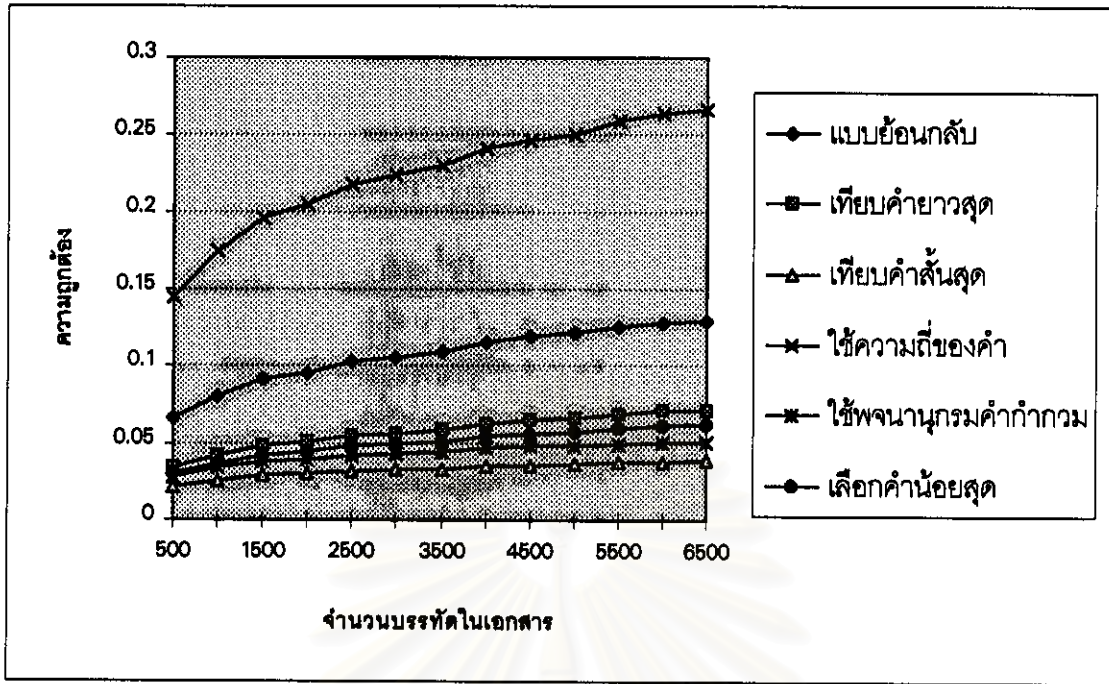


รูป 8.7 เปรียบเทียบจำนวนค่าในพจนานุกรม

โปรแกรมตัดค่า / จำนวนบรรทัด	500	1000	1500	2000	2500	3000	3500	4000	4500	5000	5500	6000	6500
แบบย้อนกลับ	0.0859	0.0803	0.0913	0.0955	0.1032	0.1058	0.1095	0.1159	0.1191	0.1215	0.1258	0.1287	0.1298
เทียบค่ายาวสุด	0.0337	0.0423	0.0485	0.051	0.0549	0.0584	0.0587	0.0623	0.0643	0.0681	0.0689	0.0706	0.0712
เทียบค่าสั้นสุด	0.0219	0.0259	0.0291	0.03	0.0318	0.0325	0.0335	0.0354	0.0362	0.0366	0.0378	0.0388	0.0389
ใช้ความถี่ของค่า	0.1446	0.1737	0.196	0.2046	0.2177	0.2235	0.2296	0.2412	0.2463	0.2508	0.2592	0.265	0.267
ใช้พจนานุกรมค่ากำกับ	0.0278	0.0337	0.0381	0.0395	0.0423	0.0433	0.0447	0.047	0.0478	0.0485	0.05	0.051	0.0514
เลือกค่าน้อยสุด	0.0299	0.037	0.0419	0.0442	0.0481	0.0494	0.0513	0.0547	0.0584	0.0578	0.0601	0.0615	0.0622

ตาราง 8.6 สัดส่วนความถูกต้องต่อจำนวนค่าในพจนานุกรมที่ใช้งาน

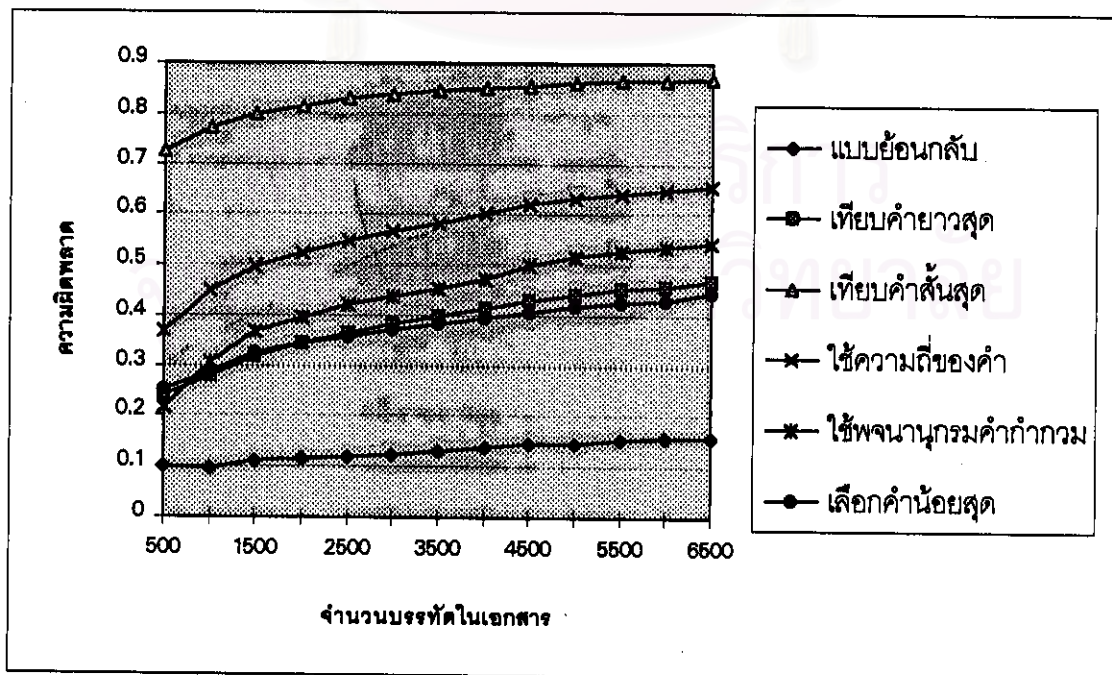
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูป 8.8 สัดส่วนความถูกต้องต่อจำนวนค่าในพจนานุกรมที่ใช้งาน

8.8 เปรียบเทียบความผิดพลาดที่เกิดขึ้นของแต่ละโปรแกรม

สัดส่วนความผิดพลาดของการตัดค่าได้มาจากนำจำนวนค่าที่ตัดออกมาไม่ถูกต้องหารด้วยจำนวนค่าที่ตัดออกมาได้ทั้งหมดได้ดังรูปที่ 8.9 ซึ่งพบว่ากรตัดค่าแบบย้อนกลับให้ค่าที่ผิดพลาดน้อยที่สุด ขณะที่แบบเทียบค่าสั้นสุดมีความผิดพลาดในการตัดค่ามากที่สุด



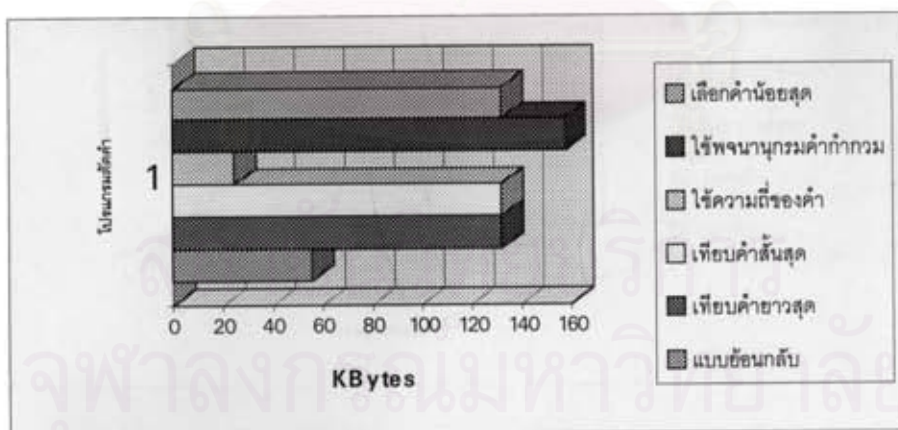
รูป 8.9 เปรียบเทียบความผิดพลาดที่เกิดขึ้นจากโปรแกรมตัดค่าต่าง ๆ

8.9 เปรียบเทียบการใช้งานทรัพยากร

การใช้งานของทรัพยากรของโปรแกรมตัดคำแบบต่าง ๆ มีผลมาจากขนาดของพจนานุกรมที่ใช้งาน เนื่องมาจากโปรแกรมตัดคำทำการโหลดพจนานุกรมไว้ในหน่วยความจำเพื่อทำการเปรียบเทียบ โดยที่โครงสร้างข้อมูลของพจนานุกรมก็มีผลต่อขนาดเช่นกัน ขนาดเป็นไบต์ของพจนานุกรมของโปรแกรมตัดคำแบบต่าง ๆ แสดงไว้ดังตาราง 8.7 และกราฟเปรียบเทียบขนาดแสดงไว้ดังรูปที่ 8.10 พบว่าโปรแกรมตัดคำแบบใช้ความถี่ของคำใช้ทรัพยากรน้อยที่สุด

โปรแกรมตัดคำ	หน่วยความจำ (ไบต์)
แบบย้อนกลับ	56,828
เทียบคำยาวสุด	134,782
เทียบคำสั้นสุด	134,782
ใช้ความถี่ของคำ	24,902
ใช้พจนานุกรมค่ากำกับ	161,028
เลือกค่าน้อยสุด	134,782

ตาราง 8.7 เปรียบเทียบการใช้งานทรัพยากร



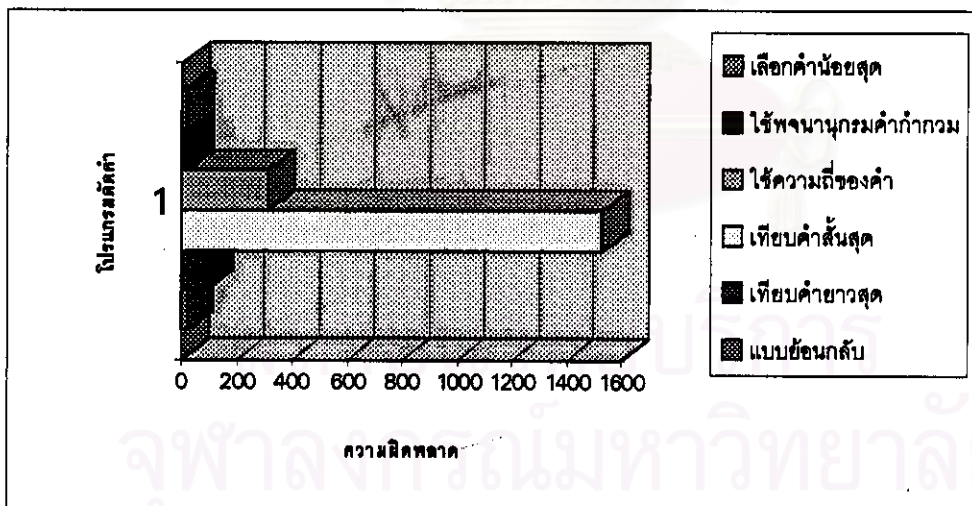
รูป 8.10 เปรียบเทียบการใช้งานทรัพยากร

8.10 เปรียบเทียบความผิดพลาดจากความกำกวมของค่า

เมื่อทำการวัดความสามารถในการจัดการกับค่ากำกวมตั้งวิธีวัดที่ได้กล่าวมาแล้วในบทที่ 7 ได้ผลจำนวนค่าที่ตัดผิดเนื่องจากความกำกวมของค่านี้แสดงไว้ในตารางที่ 8.8 และกราฟเปรียบเทียบแสดงไว้ในรูป 8.11 พบว่า โปรแกรมตัดค่าแบบใช้พจนานุกรมลดค่ากำกวมให้ค่าผิดนี้น้อยที่สุด

โปรแกรมตัดค่า	ความผิดพลาด
แบบย้อนกลับ	20
เทียบค่ายาวสุด	93
เทียบค่าสั้นสุด	1525
ใช้ความถี่ของค่า	310
ใช้พจนานุกรมค่ากำกวม	15
เลือกค่าน้อยสุด	16

ตาราง 8.8 ความผิดพลาดจากความกำกวมของค่า



รูป 8.11 เปรียบเทียบความผิดพลาดจากความกำกวมของค่า

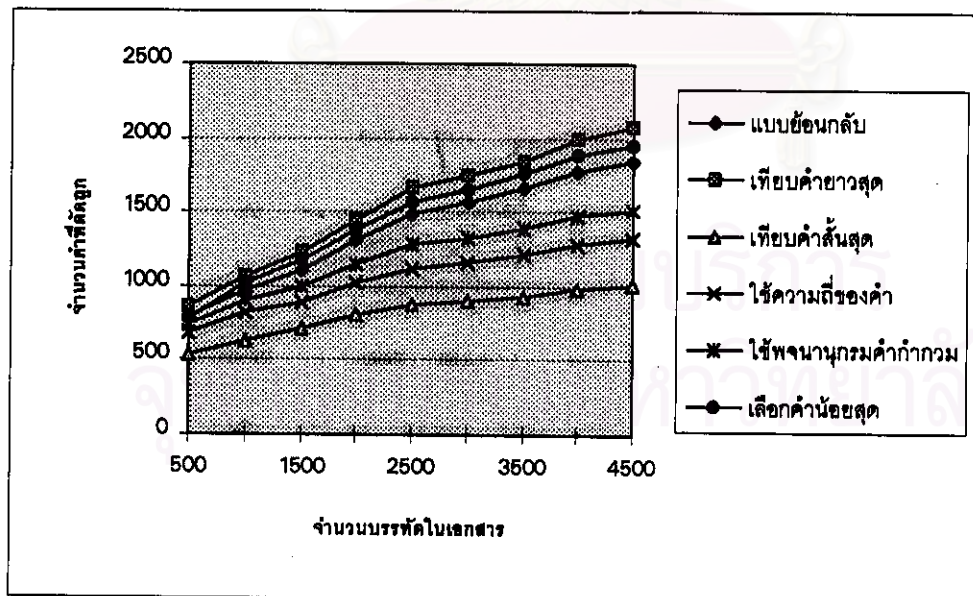
8.11 ผลการทดลองที่ได้จากฐานข้อมูลภาษาไทยด้าแบงค์

เพื่อให้การทดสอบประสิทธิภาพน่าเชื่อถือยิ่งขึ้น ได้ทำการทดลองวัดประสิทธิภาพกับฐานข้อมูลอีกชุดหนึ่งคือฐานข้อมูลภาษาไทยด้าแบงค์[15] เพื่อเปรียบเทียบความแตกต่างของผลการทดลองที่ได้ โดยพิจารณาในแง่ของ

- ก. จำนวนคำที่ตัดออกมาได้ถูกต้อง ดังแสดงไว้โดยตาราง 8.9 และรูปที่ 8.12
- ข. จำนวนคำที่ตัดออกมาได้ทั้งหมดดังตาราง 8.10 และรูป 8.13
- ค. ความถูกต้องของการตัดคำดังตารางที่ 8.11 และรูปที่ 8.14 และ
- ง. ความถูกต้องต่อจำนวนคำในพจนานุกรมดังตารางที่ 8.12 และรูปที่ 8.15 ตามลำดับ

โปรแกรมตัดคำ/ จำนวนบรรทัดในเอกสาร	500	1000	1500	2000	2500	3000	3500	4000	4500
แบบย้อนกลับ	789	970	1110	1307	1492	1570	1860	1776	1848
เทียบคำยาวสุด	861	1070	1224	1448	1661	1747	1848	1994	2077
เทียบคำสั้นสุด	533	632	714	809	879	903	933	983	1017
ใช้ความถี่ของคำ	688	814	891	1020	1127	1161	1212	1289	1326
ใช้พจนานุกรม คำก่าวม	743	897	995	1150	1279	1327	1389	1476	1520
เลือกคำน้อยสุด	780	1028	1172	1380	1571	1656	1766	1890	1958

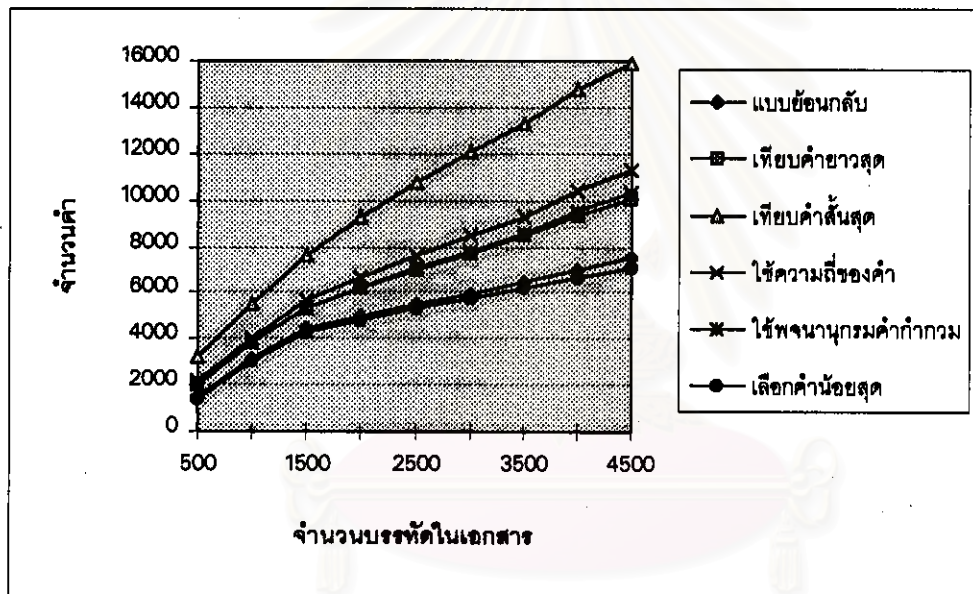
ตาราง 8.9 จำนวนคำที่ตัดได้ถูกต้อง



รูป 8.12 จำนวนคำที่ตัดได้ถูกต้อง

โปรแกรมตัดค่า/ จำนวนบรรทัดในเอกสาร	500	1000	1500	2000	2500	3000	3500	4000	4500
แบบย้อนกลับ	1498	3151	4464	5019	5466	5963	6434	6961	7501
เทียบค่ายาวสุด	1975	3626	5355	6202	6995	7722	8453	9327	10096
เทียบค่าสั้นสุด	3240	5540	7635	9253	10762	12029	13285	14763	15930
ใช้ความถี่ของค่า	2217	4027	5662	6666	7626	8483	9305	10395	11293
ใช้พหุนามกรม ค่าค่ากรรม	2015	3806	5376	6251	7049	7819	8568	9546	10347
เลือกค่าน้อยสุด	1425	3101	4312	4840	5316	5754	6176	6666	7122

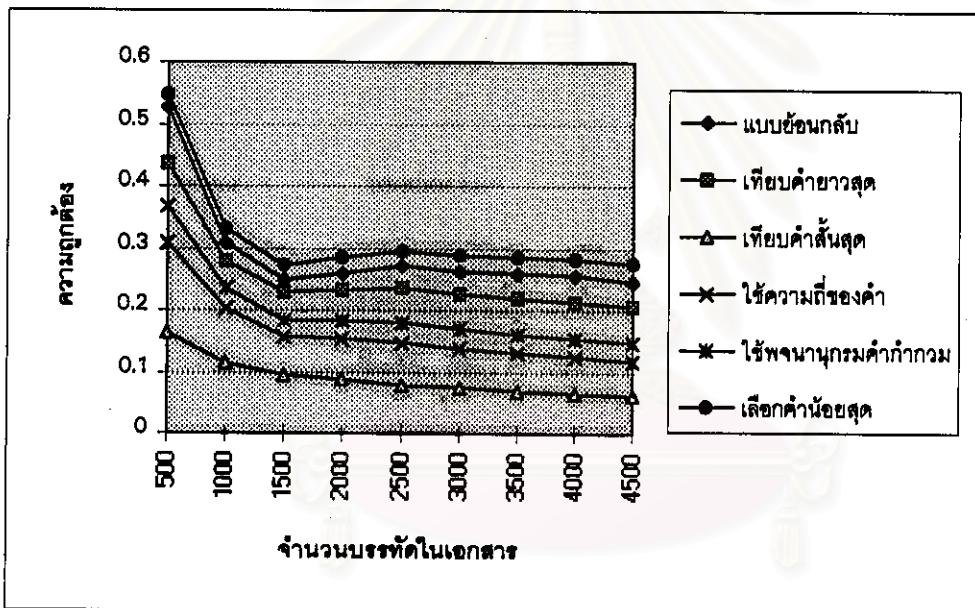
ตาราง 8.10 จำนวนค่าที่ตัดได้ทั้งหมด



รูป 8.13 จำนวนค่าที่ตัดได้ทั้งหมด

โปรแกรมตัดค่า จำนวนบรรทัด	500	1000	1500	2000	2500	3000	3500	4000	4500
แบบย้อนกลับ	0.5267	0.3078	0.2487	0.2604	0.273	0.2633	0.258	0.2544	0.2484
เทียบค่ายาวสุด	0.4359	0.2797	0.2286	0.2332	0.2375	0.2262	0.2186	0.2136	0.2057
เทียบค่าสั้นสุด	0.1645	0.1141	0.0935	0.0874	0.0802	0.0751	0.0702	0.0666	0.0638
ใช้ความถี่ของค่า	0.309	0.2021	0.1574	0.1526	0.1478	0.1389	0.1303	0.124	0.1174
ใช้พจนานุกรม ค่ากำหนด	0.3687	0.2357	0.1851	0.184	0.1814	0.1697	0.1621	0.1546	0.1469
เลือกค่าน้อยสุด	0.5474	0.3309	0.2718	0.2851	0.2955	0.2878	0.2859	0.2835	0.2749

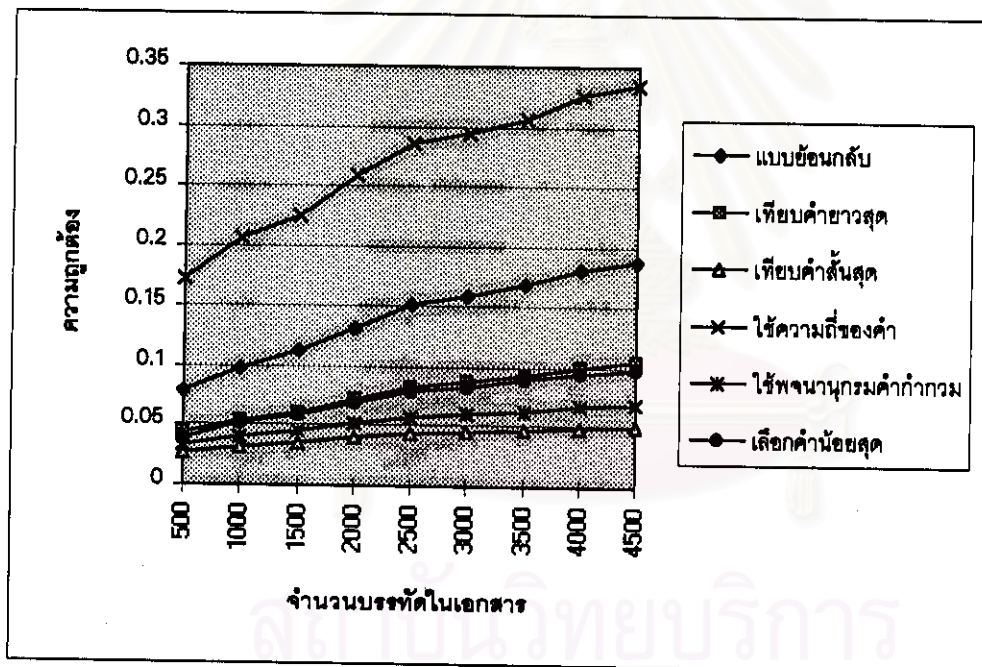
ตาราง 8.11 ความถูกต้องของการตัดค่า



รูป 8.14 ความถูกต้องของการตัดค่า

โปรแกรมตัดค่า /จำนวนบรรทัด	500	1000	1500	2000	2500	3000	3500	4000	4500
แบบย้อนกลับ	0.0801	0.0985	0.1127	0.1327	0.1515	0.1594	0.1685	0.1803	0.1874
เทียบค่ายาวสุด	0.0434	0.0539	0.0616	0.0728	0.0836	0.088	0.0931	0.1004	0.1046
เทียบค่าสั้นสุด	0.0268	0.0318	0.036	0.0407	0.0443	0.0455	0.047	0.0495	0.0512
ใช้ความถี่ของค่า	0.1732	0.2058	0.2253	0.2579	0.285	0.2936	0.3064	0.3259	0.3353
ใช้พจนานุกรม ค่ากำหนด	0.0338	0.0408	0.0452	0.0523	0.0582	0.0603	0.0632	0.0671	0.0691
เลือกค่าน้อยสุด	0.0393	0.0517	0.059	0.0685	0.0791	0.0834	0.0869	0.0952	0.0986

ตาราง 8.12 ความถูกต้องต่อจำนวนค่าในพจนานุกรม



รูป 8.15 ความถูกต้องต่อจำนวนค่าในพจนานุกรม

8.12 การวิเคราะห์ผลการทดลอง

จากผลการวัดประสิทธิภาพในบทนี้พบว่าคุณภาพของข้อมูลภาษาไทยที่นำมาใช้วัดประสิทธิภาพมีผลน้อยมากต่อผลการวัดประสิทธิภาพที่ได้ ซึ่งในงานวิจัยนี้ได้เลือกข้อมูลมาใช้วัดจากสองแหล่งที่มีคุณภาพต่างกันดังที่ได้กล่าวในรายละเอียดมาแล้วในบทที่ 6 พบว่าผลของการวัดทำให้ได้คุณลักษณะดีหรือข้อดีของโปรแกรมตัดคำในแง่มุมต่าง ๆ ที่เหมือนกัน

จากผลการวิจัยสามารถวิเคราะห์ได้ประเด็นที่สำคัญต่าง ๆ โดยแยกตามข้อดีและข้อด้อยในแง่มุมต่าง ๆ ของโปรแกรมตัดคำภาษาไทยดังต่อไปนี้

8.12.1 ข้อดี

- ก. โปรแกรมตัดคำแบบเทียบคำสั้นสุดให้จำนวนคำที่ตัดออกมาได้ทั้งหมดมากที่สุด
- ข. โปรแกรมตัดคำแบบเทียบคำยาวสุดให้จำนวนคำที่ตัดออกมาได้ถูกต้องมากที่สุด
- ค. โปรแกรมตัดคำแบบย้อนกลับให้จำนวนคำที่ตัดผิดออกมาน้อยที่สุด
- ง. โปรแกรมตัดคำแบบย้อนกลับให้สัดส่วนความถูกต้องต่อจำนวนคำที่ตัดออกมา สูงสุด
- จ. โปรแกรมตัดคำแบบใช้ความถี่ของคำให้สัดส่วนความถูกต้องต่อจำนวนคำใน

พจนานุกรมสูงสุด

- ฉ. โปรแกรมตัดคำแบบใช้ความถี่ของคำใช้ทรัพยากรน้อยที่สุด
- ช. โปรแกรมตัดคำแบบใช้พจนานุกรมลดคำกำกวมให้ความผิดพลาดจากความกำกวมของคำน้อยที่สุด

8.12.2 ข้อด้อย

- ก. โปรแกรมตัดคำแบบย้อนกลับให้จำนวนคำที่ตัดออกมาได้ทั้งหมดน้อยที่สุด
- ข. โปรแกรมตัดคำแบบเทียบคำสั้นสุดให้จำนวนคำที่ตัดออกมาได้ถูกต้องน้อยที่สุด
- ค. โปรแกรมตัดคำแบบเทียบคำสั้นสุดให้จำนวนคำที่ตัดผิดออกมามากที่สุด
- ง. โปรแกรมตัดคำแบบเทียบคำสั้นสุดให้สัดส่วนความถูกต้องต่อจำนวนคำที่ตัดออกม
ต่ำสุด
- จ. โปรแกรมตัดคำแบบเทียบคำสั้นสุดให้สัดส่วนความถูกต้องต่อจำนวนคำใน
พจนานุกรมต่ำสุด
- ฉ. โปรแกรมตัดคำแบบใช้พจนานุกรมลดคำกำกวมใช้ทรัพยากรมากที่สุด
- ช. โปรแกรมตัดคำแบบเทียบคำสั้นสุดให้ความผิดพลาดจากความกำกวมของคำมากที่สุด

8.13 ความเหมาะสมของการประยุกต์ใช้โปรแกรมตัดคำ

จากผลการวิจัยนี้สามารถทำให้เห็นแนวทางเบื้องต้นในการพิจารณานำโปรแกรมตัดคำภาษาไทยไปใช้กับงานประเภทต่าง ๆ ให้เหมาะสม อย่างไรก็ตามน่าจะมีการศึกษาวิจัยเพิ่มเติมให้ลึกซึ้งถึงคุณลักษณะและความต้องการที่จำเป็นสำหรับงานประเภทต่าง ๆ ที่ต้องการการตัดคำภาษาไทยสำหรับการพิจารณาในเบื้องต้นนี้จะเป็นดังนี้

ก. งานทางด้านจัดการรูปแบบเอกสาร เป็นงานที่ไม่ต้องการผลการตัดคำที่ละเอียดมากนัก การตัดคำแบบเทียบคำที่ยาวที่สุดหรือแบบเลือกค่าน้อยสุดให้ผลการวัดประสิทธิภาพที่เพียงพอสำหรับงานประเภทนี้ได้

ข. งานตรวจสอบคำสะกด ต้องการโปรแกรมที่ตัดคำได้คำผิดน้อยที่สุด โปรแกรมตัดคำแบบย้อนกลับจึงเหมาะสำหรับงานประเภทนี้

ค. งานแปลภาษาด้วยเครื่องหรืองานสังเคราะห์เสียงพูดจากประโยคภาษาไทย ควรเลือกโปรแกรมตัดคำที่ให้ความถูกต้องทางไวยากรณ์และความหมายสูง นั่นคือมีความผิดพลาดจากความกำกวมของค่าน้อยที่สุด โปรแกรมตัดคำแบบใช้พจนานุกรมลดคำกำกวมจึงเหมาะสมสำหรับงานประเภทนี้

สำหรับโปรแกรมตัดคำแบบใช้ความถี่ของคำมีแนวโน้มที่จะสามารถพัฒนาเพิ่มเติมเพื่อที่จะมีความสามารถนำไปใช้งานได้ แต่สำหรับโปรแกรมตัดคำแบบเทียบคำสั้นที่สุดจะมีความผิดพลาดมากที่สุดจึงไม่เหมาะที่จะนำไปประยุกต์ใช้งานได้เลย

ในบทนี้ได้รวบรวมผลการวัดและเปรียบเทียบประสิทธิภาพของโปรแกรมตัดคำ ในแง่มุมที่ต่าง ๆ กันรวมทั้งการวิเคราะห์ผลการวัดประสิทธิภาพที่ได้ ในบทต่อไปจะได้สรุปผลการวิจัยปัญหาและข้อเสนอแนะต่าง ๆ