บทที่ 5

การทดสอบความสัมพันธ์ของขนาดของคีย์กับขนาดแฟ้มดรรชนี

จากการทดสอบหาความสัมพันธ์ของขนาดแฟ้มดรรชนี (index file length) กับขนาดของคีย์ (key length) ต่อข้อมูลภาษาไทย โดยการเปลี่ยนแปลงค่า ขนาดของคีย์จาก 1 จนถึง 30 แล้วนำข้อมูลที่ได้ไปเขียนเป็นกราฟ ดังนี้

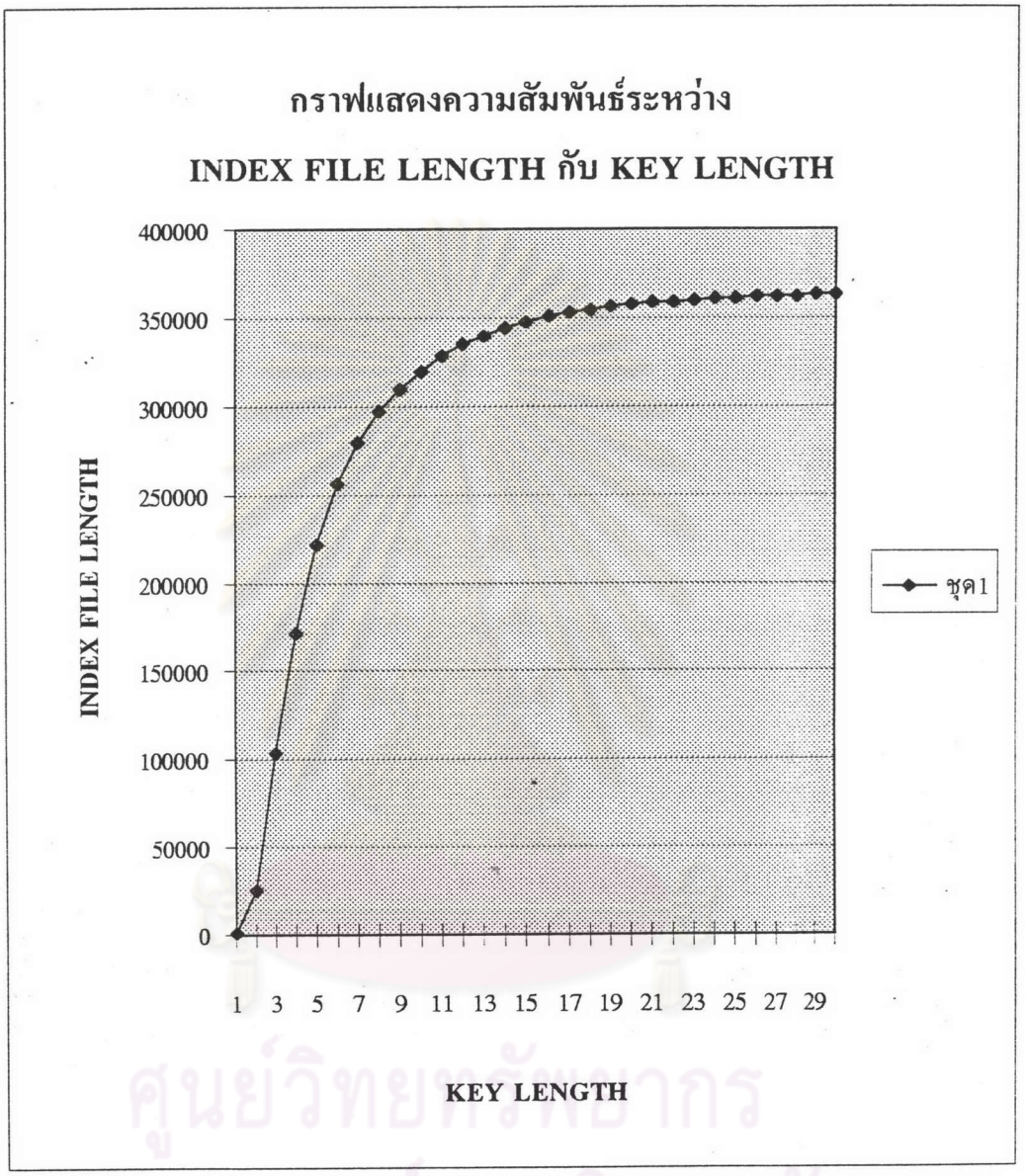
5.1 ตารางแสดงข้อมูลที่ใช้ในการทดสอบ

Key Lengths (Bytes)	Index File Lengths (Bytes)	Key Lengths (Bytes)	Index File Lengths (Bytes)
1	1224	2	25248
3	103438	4	171674
5	221500	6	255828
7	280076	8	297184
9	310148	10	320312
11	328418	12	334858
13	339968	14	344168
15	347682	16	350636
17	352862	18	354710
19	356110	20	357342
21	358308	22	359190
23	359974	24	360604
25	361150	26	361654
27	362116	28	362452
29	362732	30	362998

ตารางที่ 5.1 แสดงข้อมูลที่ใช้ในการทดสอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5.2 กราฟแสดงความสัมพันธ์



รูปที่ 5.1 แสดงความสัมพันธ์ระหว่างขนาดแฟ้มดรรชนีกับขนาดของคีย์

จากการทดสอบพบว่า ขนาดแฟ้มดรรชนีแปรผันตามขนาดของคีย์แต่อัตราไม่คงที่ โดยมีอัตราการเพิ่มขนาดแฟ้มดรรชนีต่อขนาดของคีย์น้อยกว่า 1% หลังจากขนาดของคีย์มีขนาด 15 ไบต์เป็นต้นไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5.3 แหล่งข้อมูลภาษาไทย

ใช้ข้อมูลประเภทข่าวสารจากศูนย์อิเล็กทรอนิกส์คอมพิวเตอร์และเทคโนโลยีแห่งชาติ (NECTEC) ณ วันที่ 01/04/96 (ขนาดแฟ้มข้อมูล 26052 ไบต์)

5.4 การคำนวณหาขนาดแฟ้มดรรชนี

หาได้จากสูตร

index file length = (file length - (key length - 1) - duplicate)*14 + 6

โดย : - file length คือ ขนาดของแฟ้มข้อมูลที่ใช้ทดสอบ
- duplicate คือ จำนวนครั้งที่ซ้ำของการเพิ่มข้อมูล
- 14 คือ ขนาดของโหนดในต้นไม้ ดังนี้

```
typedef struct {
        short skipb;
        long left, right, displace;
} node;
```

- 6 คือ จำนวนโหนดในต้นไม้รวมกับขนาดของคีย์ ดังนี้

```
short n; long key_length;
```

หมายเหตุ : จำนวนของซิสตริง คือ ขนาดของแฟ้มข้อมูลที่ใช้ทดสอบ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย