Fady John | fje3683

# Assignment 3: Word Count Wizard

**Description:**

The system is able to count the number of unique words in the entire text provided by the user, return the frequency of any word in that given text, among other operations. The wizard can take the input text and organize it in a hash table.

## Hash Functions:

### Hash Function 1:

**Pseudo Code:**

```
Algorithm: hashFun2
Input: string key
Output: hashVal
hashVal ← 0
i ← 0
while i < key.length() do
    hashVal ← hashVal + key[i] * 33 ^ (length of key - i - 1)
    i ← i+1
return hashVal
```

**C++ implementation:**

```
long hashFun1(string key)
{
    long hashVal = 0;
    for (int i = 0; i < key.length(); i++) {
        hashVal += key[i] * pow(33, key.length() - i - 1);
    }
    return hashVal;
}
```

### Hash Function 2:

**Pseudo Code:**

```
Algorithm: hashFun2
Input: string key
Output: hashVal
hashVal ← 0
```

```
i ← 0
while i < key.length() do
    hashVal ← shift hashVal 5 binary positions to the left or shift hash 27 binary
    positions to the right
    hashVal ← hashVal + key[i]
    i ←  i+1
return hashVal
```

**C++ implementation:**

```
long hashFun2(string key)
{
        long hashVal = 0;
        for (int i = 0; i < key.length(); i++) {
                hashVal = ((hashVal << 5) | (hashVal >> 27));
                hashVal += (unsigned long)key[i];
   }
   return hashVal;
}
```

## Hash Function 3:

**Pseudo Code:**

```
Algorithm: hashFun3
Input: string key
Output: hashVal
hashVal ← 0
i ← 0
while i < key.length() do
    hashVal ← hashVal + key[i]
    i ←  i+1
return hashVal
```

**C++ implementation:**

```
long hashFun3(string key)
{
        long hashVal = 0;
        for (int i = 0; i < key.length(); i++) {
                hashVal += key[i];
   }
   return hashVal;
}
```

**Hash Function 4:**

**Pseudo Code:**

```
Algorithm: hashFun4
Input: string key
Output: hashVal
hashVal ← 0
i ← 1
while i < key.length() do
        hashVal ← hashVal + key[i]*key[i-1]
        i ←  i+1
return hashVal
```

**C++ implementation:**

```
long hashFun4(string key)
{
        long hashVal = 0;
        for (int i = 1; i < key.length(); i++) {
                hashVal += key[i] * key[i-1];
   }
   return hashVal;
}
```

**Hash Function 5:**

**Pseudo Code:**

```
Algorithm: hashFun5
Input: string key
Output: hashVal
hashVal ← 0
i ← 0
while i < key.length() do
        hashVal ← key[i] + (shift hashVal 6 binary positions to the left) + (shift hashVal
        16 binary positions to the right) - hashVal
        i ←  i+1
return hashVal
```

**C++ implementation:**

```
long hashFun5(string key)
```

```
{
        long hashVal = 0;
        for (int i = 0; i < key.length(); i++) {
                hashVal = key[i] + (hashVal << 6) + (hashVal << 16) - hashVal;
   }
   return hashVal;
}
```

**Hash Function 6:**

**Pseudo Code:**

```
Algorithm: hashFun6
Input: string key
Output: hashVal
hashVal ← 0
i ← 0
while i < key.length() do
        hashVal ← hashVal + key[i]^3
        i ←  i+1
return hashVal
```

**C++ implementation:**

```
long hashFun6(string key)
{
        long hashVal = 0;
        for (int i = 0; i < key.length(); i++) {
                hashVal += pow(key[i], 3);
   }
   return hashVal;
}
```

## Number of collisions in each file corresponding to the hash function used:

| File Name | HashFun 1 | HashFun 2 | HashFun 3 | HashFun 4 | HashFun 5 | HashFun 6 |
|---|---|---|---|---|---|---|
| 10947-8.txt | 5155 | 4818 | 13784 | 4924 | 4739 | 4929 |
| 2550-0.txt | 3382 | 3168 | 8623 | 3260 | 3208 | 3301 |
| 34313-8.txt | 2617 | 2458 | 6370 | 2486 | 2424 | 2461 |
| 56870-8.txt | 3045 | 2825 | 7894 | 3009 | 2930 | 3016 |
| 13799.txt | 3928 | 3725 | 10409 | 3822 | 3659 | 3885 |
| 26772.txt | 957 | 877 | 1982 | 932 | 928 | 899 |
| 34766-0.txt | 4988 | 4676 | 13277 | 4664 | 4582 | 4903 |
| 57006-0.txt | 2457 | 2348 | 6144 | 2383 | 2635 | 2425 |
| 14744-8.txt | 2765 | 2669 | 7242 | 2728 | 2604 | 2765 |
| 2781-0.txt | 1505 | 1393 | 3259 | 1414 | 1424 | 1455 |
| 373-0.txt | 4126 | 3932 | 11072 | 3955 | 3892 | 4030 |
| 57040-0.txt | 4218 | 3768 | 10763 | 3811 | 3773 | 3938 |
| 15717-8.txt | 3338 | 3035 | 8411 | 3183 | 3080 | 3228 |
| 28062.txt | 621 | 583 | 1159 | 603 | 577 | 606 |
| 38172-8.txt | 4619 | 4296 | 12129 | 4372 | 4268 | 4428 |
| 5737-0.txt | 3849 | 3545 | 9794 | 3564 | 3577 | 3696 |
| 1626-0.txt | 4426 | 4031 | 11051 | 4033 | 3970 | 4131 |
| 28650.txt | 506 | 487 | 866 | 503 | 482 | 477 |
| 38531-8.txt | 5209 | 4823 | 13877 | 4969 | 4833 | 5009 |
| 58341-0.txt | 3931 | 3683 | 10241 | 3680 | 3718 | 3844 |
| 17669-8.txt | 5020 | 4820 | 13879 | 4899 | 4760 | 4948 |
| 28698.txt | 785 | 756 | 1639 | 776 | 788 | 791 |
| 39706.txt | 1015 | 993 | 2347 | 1029 | 1001 | 1068 |
| 58735.txt | 634 | 563 | 1056 | 580 | 569 | 590 |
| 18776-8.txt | 3386 | 3249 | 8918 | 3363 | 3237 | 3503 |
| 28726-8.txt | 4762 | 4388 | 12830 | 4594 | 4453 | 4636 |
| 40745-8.txt | 3027 | 2838 | 7877 | 3000 | 2882 | 2987 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 58743.txt | 609 | 600 | 1208 | 610 | 622 | 608 |
| 1944-0.txt | 2982 | 2766 | 7413 | 2839 | 2742 | 2860 |
| 29503.txt | 570 | 538 | 1011 | 531 | 558 | 569 |
| 41562.txt | 707 | 696 | 1405 | 723 | 698 | 692 |
| 58991.txt | 766 | 753 | 1526 | 718 | 738 | 720 |
| 1982-0.txt | 397 | 342 | 382 | 326 | 277 | 305 |
| 29618.txt | 515 | 514 | 891 | 508 | 526 | 526 |
| 42664.txt | 568 | 531 | 1066 | 563 | 568 | 556 |
| 58995-8.txt | 530 | 518 | 968 | 520 | 534 | 517 |
| 21782.txt | 749 | 759 | 1489 | 737 | 745 | 717 |
| 29750.txt | 562 | 537 | 952 | 559 | 542 | 539 |
| 49598-8.txt | 2507 | 2413 | 6321 | 2380 | 2383 | 2471 |
| 59255.txt | 813 | 816 | 1684 | 804 | 786 | 811 |
| 22426-8.txt | 1418 | 1358 | 3355 | 1388 | 1382 | 1420 |
| 30029-8.txt | 605 | 582 | 1071 | 598 | 599 | 600 |
| 50877.txt | 667 | 628 | 1231 | 673 | 631 | 641 |
| 59368.txt | 533 | 538 | 968 | 526 | 525 | 551 |
| 22522-8.txt | 1803 | 1717 | 4410 | 1781 | 1690 | 1793 |
| 30044.txt | 470 | 446 | 730 | 452 | 444 | 442 |
| 51008.txt | 477 | 467 | 845 | 497 | 487 | 470 |
| 6040.txt | 2497 | 2452 | 6514 | 2467 | 2437 | 2544 |
| 22662-8.txt | 745 | 666 | 1422 | 736 | 707 | 703 |
| 31217-8.txt | 5464 | 4972 | 14386 | 5110 | 4947 | 5265 |
| 51129.txt | 705 | 677 | 1360 | 667 | 659 | 696 |
| 6073-0.txt | 3008 | 2944 | 7947 | 3110 | 2916 | 2988 |
| 22897-8.txt | 949 | 870 | 1861 | 881 | 885 | 880 |
| 3181-0.txt | 847 | 815 | 1669 | 854 | 822 | 780 |
| 51193.txt | 820 | 770 | 1629 | 795 | 766 | 821 |
| 6120-0.txt | 4413 | 4045 | 11443 | 4166 | 4114 | 4290 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2305-0.txt | 4230 | 3621 | 9699 | 3626 | 3527 | 3751 |
| 31840.txt | 622 | 596 | 1114 | 612 | 595 | 613 |
| 51268.txt | 843 | 803 | 1789 | 848 | 822 | 828 |
| 6168.txt | 1432 | 1402 | 3546 | 1450 | 1410 | 1437 |
| 23099.txt | 436 | 430 | 656 | 426 | 433 | 424 |
| 32040.txt | 765 | 732 | 1547 | 782 | 765 | 767 |
| 51296.txt | 627 | 622 | 1209 | 623 | 597 | 656 |
| 6696-8.txt | 5437 | 4846 | 13725 | 4801 | 4797 | 5037 |
| 23210-0.txt | 846 | 785 | 1613 | 799 | 750 | 803 |
| 32046-8.txt | 5798 | 5453 | 15885 | 5478 | 5436 | 5715 |
| 51493.txt | 624 | 567 | 1097 | 582 | 581 | 577 |
| 8129-8.txt | 1654 | 1625 | 4203 | 1661 | 1646 | 1720 |
| 2327-8.txt | 2262 | 2082 | 5405 | 2113 | 2100 | 2114 |
| 32067.txt | 800 | 758 | 1611 | 800 | 760 | 770 |
| 51498.txt | 538 | 498 | 904 | 507 | 505 | 514 |
| 877-0.txt | 746 | 723 | 1510 | 709 | 738 | 748 |
| 2334-0.txt | 9914 | 8860 | 26028 | 9158 | 8638 | 9303 |
| 32077.txt | 664 | 650 | 1252 | 660 | 602 | 650 |
| 51603.txt | 564 | 512 | 919 | 535 | 514 | 515 |
| 8933-0.txt | 3962 | 3760 | 10726 | 3939 | 3792 | 3950 |
| 23942-8.txt | 631 | 638 | 1218 | 641 | 608 | 629 |
| 32078.txt | 755 | 724 | 1534 | 760 | 750 | 764 |
| 51687.txt | 757 | 712 | 1448 | 737 | 707 | 731 |
| 9205.txt | 551 | 537 | 1030 | 545 | 541 | 550 |
| 2429-0.txt | 2046 | 2007 | 5092 | 1991 | 1988 | 2083 |
| 32104.txt | 683 | 684 | 1365 | 731 | 681 | 680 |
| 51699.txt | 688 | 671 | 1344 | 683 | 660 | 658 |
| 9629-8.txt | 2251 | 2219 | 5770 | 2257 | 2204 | 2276 |
| 24313-8.txt | 3504 | 3233 | 9203 | 3360 | 3270 | 3384 |

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
| 32133.txt | 628 | 602 | 1186 | 609 | 585 | 625 |
| 51752.txt | 753 | 745 | 1483 | 736 | 744 | 712 |
| 9790-8.txt | 5358 | 4967 | 14157 | 4974 | 4874 | 5000 |
| 24558.txt | 923 | 885 | 1880 | 904 | 861 | 882 |
| 32347.txt | 419 | 412 | 641 | 423 | 400 | 411 |
| 54183-0.txt | 2969 | 2826 | 7590 | 2887 | 2831 | 2968 |
| pg4081.txt | 3478 | 3252 | 9310 | 3438 | 3303 | 3435 |
| 24878-8.txt | 3836 | 3684 | 10176 | 3647 | 3582 | 3724 |
| 3254.txt | 25366 | 22162 | 67092 | 22653 | 21546 | 22857 |
| 55514-0.txt | 4845 | 4471 | 12861 | 4569 | 4532 | 4646 |
| 25035.txt | 766 | 758 | 1607 | 756 | 758 | 768 |
| 32735.txt | 773 | 771 | 1630 | 805 | 780 | 818 |
| 55865-0.txt | 1676 | 1633 | 4043 | 1680 | 1620 | 1667 |
| 2518.txt | 2308 | 2216 | 6046 | 2348 | 2286 | 2325 |
| 32845-8.txt | 5289 | 4980 | 14289 | 5099 | 4914 | 5133 |
| 5592.txt | 3564 | 3363 | 9385 | 3399 | 3299 | 3482 |
|  |  |  |  |  |  |  |
| **Average:** | **2371** | **2207** | **5939** | **2255** | **2198** | **2285** |

**Findings:**

Based on the previous experimental results exploring the performance of the six hash functions when applied to the 101 text files, **Hash Function 5** resulted in the least average number of collisions. Hence, hash function 5 will be the default hash function of the program.