*Dumbino* ~ a console word (x-gram) checker

No time for many words, so 2 steps are given below 'dir' and 'Am-am.bat', first shows what this demo provides regarding files second shows automated script giving all unfamiliar words (and 2-grams) – see page 17. Well, one more: in order to check your text file against *googlebooks-eng-us-all-4gram-20090715* corpus (140,222,335 4-grams in 3,233,748,341 bytes) use simply 'Dumbino_26Clash_4-grams.BAT' - see page 18. Enjoy!

```
D:\Dumbino_r1>dir

03/29/2012  11:29 PM             5,457 Am-am.bat
03/29/2012  11:29 PM         4,024,155 english.dic_351116_wordlist.txt
03/29/2012  11:29 PM           460,867 Gibson, William - Cyberpunk 1 - Neuromancer.txt
03/29/2012  11:29 PM           460,486 Gibson, William - Neuromancer.txt
03/29/2012  11:29 PM             1,632 KAZE prompt.lnk
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_01_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_02_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_03_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_04_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_05_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_06_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_07_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_08_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_09_01p.c
03/29/2012  11:29 PM           315,203 Leprechaun_x-leton_10_01p.c
03/29/2012  11:29 PM           103,936 Leprechaun_x-leton_32bit_01_01p.exe
03/29/2012  11:29 PM           104,448 Leprechaun_x-leton_32bit_02_01p.exe
03/29/2012  11:29 PM           104,448 Leprechaun_x-leton_32bit_03_01p.exe
03/29/2012  11:29 PM           104,960 Leprechaun_x-leton_32bit_04_01p.exe
03/29/2012  11:29 PM           105,984 Leprechaun_x-leton_32bit_05_01p.exe
03/29/2012  11:29 PM           106,496 Leprechaun_x-leton_32bit_06_01p.exe
03/29/2012  11:29 PM           106,496 Leprechaun_x-leton_32bit_07_01p.exe
03/29/2012  11:29 PM           107,520 Leprechaun_x-leton_32bit_08_01p.exe
03/29/2012  11:29 PM           107,520 Leprechaun_x-leton_32bit_09_01p.exe
03/29/2012  11:29 PM           108,032 Leprechaun_x-leton_32bit_10_01p.exe
03/29/2012  11:29 PM            44,074 Overlapper-Blender_r1+.c
03/29/2012  11:29 PM            66,048 Overlapper-Blender_r1+1300MB.exe
03/29/2012  11:29 PM            70,656 QuickSortExternal_4+GB_32bit_ascending.exe
03/29/2012  11:29 PM            70,656 QuickSortExternal_4+GB_32bit_descending.exe
03/29/2012  11:29 PM           107,063 QuickSortExternal_4+GB_ascending.c
03/29/2012  11:29 PM           107,180 QuickSortExternal_4+GB_descending.c

D:\Dumbino_r1>Am-am.bat
First copy x-grammed...

D:\Dumbino_r1>dir "Gibson, William - Cyberpunk 1 - Neuromancer.txt"/b  1>"Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst"

D:\Dumbino_r1>Leprechaun_x-leton_32bit_01_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.01.wrd" 16000 y
Leprechaun_singleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
```

Purpose: Rips all distinct 1-grams (1-word phrases) with length 1..31 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
/; 00,081,971P/s; Phrase count: 81,971 of them 8,967 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 81,971P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,017,934P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 841KB
Total distinct phrases: 8,967
Total time: 1 second(s)
Total performance: 81,971P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_02_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.02.wrd" 16000 y
Leprechaun_doubleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 2-grams (2-word phrases) with length 5..41 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
-; 00,066,018P/s; Phrase count: 66,018 of them 37,067 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 66,018P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,074,134P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 4,195KB

```
Total distinct phrases: 37,067
Total time: 1 second(s)
Total performance: 66,018P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_03_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.03.wrd" 16000 y
Leprechaun_tripleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 3-grams (3-word phrases) with length 9..41 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
\; 00,053,679P/s; Phrase count: 53,679 of them 47,646 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 53,679P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,095,292P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 5,391KB
Total distinct phrases: 47,646
Total time: 1 second(s)
Total performance: 53,679P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_04_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.04.wrd" 16000 y
Leprechaun_quadrupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 4-grams (4-word phrases) with length 13..51 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
|; 00,043,272P/s; Phrase count: 43,272 of them 42,208 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
```

Phrases per second performance: 43,272P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,084,416P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.


Total memory needed for one pass: 5,597KB
Total distinct phrases: 42,208
Total time: 1 second(s)
Total performance: 43,272P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_05_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.05.wrd" 16000 y
Leprechaun_quintupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 5-grams (5-word phrases) with length 17..61 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
/; 00,034,597P/s; Phrase count: 34,597 of them 34,390 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 34,597P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,068,780P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.


Total memory needed for one pass: 5,234KB
Total distinct phrases: 34,390
Total time: 1 second(s)
Total performance: 34,597P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_06_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.06.wrd" 16000 y
Leprechaun_sextupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 6-grams (6-word phrases) with length 21..71 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.

```
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
-; 00,027,445P/s; Phrase count: 27,445 of them 27,382 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 27,445P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,054,764P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 4,703KB
Total distinct phrases: 27,382
Total time: 1 second(s)
Total performance: 27,445P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_07_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.07.wrd" 16000 y
Leprechaun_septupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 7-grams (7-word phrases) with length 25..81 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
\; 00,021,621P/s; Phrase count: 21,621 of them 21,597 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 21,621P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,043,194P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 4,132KB
Total distinct phrases: 21,597
Total time: 1 second(s)
Total performance: 21,621P/s i.e. phrases per second
Leprechaun: Done.
```

```
D:\Dumbino_r1>Leprechaun_x-leton_32bit_08_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.08.wrd" 16000 y
Leprechaun_octupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 8-grams (8-word phrases) with length 29..91 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
|; 00,016,984P/s; Phrase count: 16,984 of them 16,977 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 16,984P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,033,954P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 3,580KB
Total distinct phrases: 16,977
Total time: 1 second(s)
Total performance: 16,984P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_09_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.09.wrd" 16000 y
Leprechaun_nonupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 9-grams (9-word phrases) with length 33..101 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
/; 00,013,274P/s; Phrase count: 13,274 of them 13,270 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 13,274P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,026,540P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.
```

```
Total memory needed for one pass: 3,058KB
Total distinct phrases: 13,270
Total time: 1 second(s)
Total performance: 13,274P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_10_01p.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.lst" "Gibson, William - Cyberpunk 1 - Neuromancer.txt.10.wrd" 16000 y
Leprechaun_decupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 10-grams (10-word phrases) with length 37..111 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,867
-; 00,010,315P/s; Phrase count: 10,315 of them 10,313 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 10,315P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,020,626P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 2,578KB
Total distinct phrases: 10,313
Total time: 1 second(s)
Total performance: 10,315P/s i.e. phrases per second
Leprechaun: Done.

Second copy x-grammed...

D:\Dumbino_r1>dir "Gibson, William - Neuromancer.txt"/b  1>"Gibson, William - Neuromancer.txt.lst"

D:\Dumbino_r1>Leprechaun_x-leton_32bit_01_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.01.wrd" 16000 y
Leprechaun_singleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 1-grams (1-word phrases) with length 1..31 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
```

```
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
/; 00,081,815P/s; Phrase count: 81,815 of them 8,851 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 81,815P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,017,702P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 830KB
Total distinct phrases: 8,851
Total time: 1 second(s)
Total performance: 81,815P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_02_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.02.wrd" 16000 y
Leprechaun_doubleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 2-grams (2-word phrases) with length 5..41 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
-; 00,065,845P/s; Phrase count: 65,845 of them 36,882 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 65,845P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,073,764P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 4,174KB
Total distinct phrases: 36,882
Total time: 1 second(s)
Total performance: 65,845P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_03_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.03.wrd" 16000 y
```

Leprechaun_tripleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 3-grams (3-word phrases) with length 9..41 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
\; 00,053,517P/s; Phrase count: 53,517 of them 47,469 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 53,517P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,094,938P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 5,371KB
Total distinct phrases: 47,469
Total time: 1 second(s)
Total performance: 53,517P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_04_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.04.wrd" 16000 y
Leprechaun_quadrupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 4-grams (4-word phrases) with length 13..51 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
|; 00,043,122P/s; Phrase count: 43,122 of them 42,053 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 43,122P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,084,106P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

```
Total memory needed for one pass: 5,577KB
Total distinct phrases: 42,053
Total time: 1 second(s)
Total performance: 43,122P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_05_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.05.wrd" 16000 y
Leprechaun_quintupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 5-grams (5-word phrases) with length 17..61 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
/; 00,034,456P/s; Phrase count: 34,456 of them 34,247 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 34,456P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,068,494P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 5,212KB
Total distinct phrases: 34,247
Total time: 1 second(s)
Total performance: 34,456P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_06_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.06.wrd" 16000 y
Leprechaun_sextupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 6-grams (6-word phrases) with length 21..71 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
-; 00,027,313P/s; Phrase count: 27,313 of them 27,249 distinct; Done: 64/64
```

```
Bytes per second performance: 460,486B/s
Phrases per second performance: 27,313P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,054,498P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.


Total memory needed for one pass: 4,680KB
Total distinct phrases: 27,249
Total time: 1 second(s)
Total performance: 27,313P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_07_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.07.wrd" 16000 y
Leprechaun_septupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 7-grams (7-word phrases) with length 25..81 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
\; 00,021,498P/s; Phrase count: 21,498 of them 21,474 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 21,498P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,042,948P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.


Total memory needed for one pass: 4,108KB
Total distinct phrases: 21,474
Total time: 1 second(s)
Total performance: 21,498P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>Leprechaun_x-leton_32bit_08_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.08.wrd" 16000 y
Leprechaun_octupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 8-grams (8-word phrases) with length 29..91 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
```

Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
|; 00,016,878P/s; Phrase count: 16,878 of them 16,871 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 16,878P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,033,742P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 3,558KB
Total distinct phrases: 16,871
Total time: 1 second(s)
Total performance: 16,878P/s i.e. phrases per second
Leprechaun: Done.

D:\Dumbino_r1>Leprechaun_x-leton_32bit_09_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.09.wrd" 16000 y
Leprechaun_nonupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 9-grams (9-word phrases) with length 33..101 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
/; 00,013,175P/s; Phrase count: 13,175 of them 13,171 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 13,175P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,026,342P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 3,035KB
Total distinct phrases: 13,171
Total time: 1 second(s)
Total performance: 13,175P/s i.e. phrases per second
Leprechaun: Done.

```
D:\Dumbino_r1>Leprechaun_x-leton_32bit_10_01p.exe "Gibson, William - Neuromancer.txt.lst" "Gibson, William - Neuromancer.txt.10.wrd" 16000 y
Leprechaun_decupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 10-grams (10-word phrases) with length 37..111 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 35
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 16MB ... OK
Size of Input TEXTual file: 460,486
-; 00,010,225P/s; Phrase count: 10,225 of them 10,223 distinct; Done: 64/64
Bytes per second performance: 460,486B/s
Phrases per second performance: 10,225P/s
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,020,446P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 2,556KB
Total distinct phrases: 10,223
Total time: 1 second(s)
Total performance: 10,225P/s i.e. phrases per second
Leprechaun: Done.


D:\Dumbino_r1>QuickSortExternal_4+GB_32bit_ascending.exe "Gibson, William - Neuromancer.txt.01.wrd" /fast
QuickSortExternal_4+GB r.2+, written by Kaze.
Size of input file: 77,250
Counting lines ...
Allocated memory for pointers-to-words in MB: 1
Assigning pointers ...
Trying to allocate memory for the file itself in MB: 1 ... OK! Get on with fast internal accesses.
Uploading ...
Sorting 8,851 Pointers ...
Pass #1: Quicksort started ...
| RightEnd-LeftEnd: 000,000,000,022; NumberOfSplittings: 0,000,000,797 ...
Pass #2: Insertionsort started ...
/ i: 000,000,008,851 ...
NumberOfComparisons: 130,455
The time to sort 8,851 items via Quicksort+Insertionsort was 16 clocks.
Dumping the sorted data ...
Dumped 8,851 lines.
```

```
OK! Incoming and resultant file's sizes match.
Dumping the sorted data [deduplicated] ...
Dumped 8,851 distinct lines.
Total time: 109 clocks.
Performance: 75 KB/s.
Done successfully.

D:\Dumbino_r1>del QuickSortExternal_4+GB.distinct.txt

D:\Dumbino_r1>ren QuickSortExternal_4+GB.txt "Gibson, William - Neuromancer.txt.01.wrd.sorted"


D:\Dumbino_r1>QuickSortExternal_4+GB_32bit_ascending.exe "Gibson, William - Neuromancer.txt.02.wrd" /fast
QuickSortExternal_4+GB r.2+, written by Kaze.
Size of input file: 464,212
Counting lines ...
Allocated memory for pointers-to-words in MB: 1
Assigning pointers ...
Trying to allocate memory for the file itself in MB: 1 ... OK! Get on with fast internal accesses.
Uploading ...
Sorting 36,882 Pointers ...
Pass #1: Quicksort started ...
- RightEnd-LeftEnd: 000,000,000,028; NumberOfSplittings: 0,000,003,334 ...
Pass #2: Insertionsort started ...
/ i: 000,000,036,882 ...
NumberOfComparisons: 623,859
The time to sort 36,882 items via Quicksort+Insertionsort was 47 clocks.
Dumping the sorted data ...
Dumped 36,882 lines.
OK! Incoming and resultant file's sizes match.
Dumping the sorted data [deduplicated] ...
Dumped 36,882 distinct lines.
Total time: 343 clocks.
Performance: 453 KB/s.
Done successfully.

D:\Dumbino_r1>del QuickSortExternal_4+GB.distinct.txt

D:\Dumbino_r1>ren QuickSortExternal_4+GB.txt "Gibson, William - Neuromancer.txt.02.wrd.sorted"


D:\Dumbino_r1>QuickSortExternal_4+GB_32bit_ascending.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.01.wrd" /fast
QuickSortExternal_4+GB r.2+, written by Kaze.
Size of input file: 77,841
Counting lines ...
```

```
Allocated memory for pointers-to-words in MB: 1
Assigning pointers ...
Trying to allocate memory for the file itself in MB: 1 ... OK! Get on with fast internal accesses.
Uploading ...
Sorting 8,967 Pointers ...
Pass #1: Quicksort started ...
| RightEnd-LeftEnd: 000,000,000,022; NumberOfSplittings: 0,000,000,800 ...
Pass #2: Insertionsort started ...
/ i: 000,000,008,967 ...
NumberOfComparisons: 129,745
The time to sort 8,967 items via Quicksort+Insertionsort was 16 clocks.
Dumping the sorted data ...
Dumped 8,967 lines.
OK! Incoming and resultant file's sizes match.
Dumping the sorted data [deduplicated] ...
Dumped 8,967 distinct lines.
Total time: 109 clocks.
Performance: 76 KB/s.
Done successfully.

D:\Dumbino_r1>del QuickSortExternal_4+GB.distinct.txt

D:\Dumbino_r1>ren QuickSortExternal_4+GB.txt "Gibson, William - Cyberpunk 1 - Neuromancer.txt.01.wrd.sorted"


D:\Dumbino_r1>QuickSortExternal_4+GB_32bit_ascending.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.02.wrd" /fast
QuickSortExternal_4+GB r.2+, written by Kaze.
Size of input file: 465,519
Counting lines ...
Allocated memory for pointers-to-words in MB: 1
Assigning pointers ...
Trying to allocate memory for the file itself in MB: 1 ... OK! Get on with fast internal accesses.
Uploading ...
Sorting 37,067 Pointers ...
Pass #1: Quicksort started ...
- RightEnd-LeftEnd: 000,000,000,026; NumberOfSplittings: 0,000,003,355 ...
Pass #2: Insertionsort started ...
/ i: 000,000,037,067 ...
NumberOfComparisons: 629,536
The time to sort 37,067 items via Quicksort+Insertionsort was 47 clocks.
Dumping the sorted data ...
Dumped 37,067 lines.
OK! Incoming and resultant file's sizes match.
Dumping the sorted data [deduplicated] ...
Dumped 37,067 distinct lines.
```

```
Total time: 343 clocks.
Performance: 454 KB/s.
Done successfully.

D:\Dumbino_r1>del QuickSortExternal_4+GB.distinct.txt

D:\Dumbino_r1>ren QuickSortExternal_4+GB.txt "Gibson, William - Cyberpunk 1 - Neuromancer.txt.02.wrd.sorted"


D:\Dumbino_r1>Overlapper-Blender_r1+1300MB.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.01.wrd.sorted" english.dic_351116_wordlist.txt
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 77841
Size of 2nd input file: 4024155
Allocating 1300MB ...
Lines in 1st input file: 8967
Lines in 2nd input file: 351116
Allocated memory for pointers-to-words in MB: 2
Allocated memory for pointers-to-words in MB: 1
Sorting 360083 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 351830
Overlapped lines, i.e. lines common for both files: 8253
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 714

D:\Dumbino_r1>ren Unfamiliar.txt Unfamiliar.1.spell-checked.txt

D:\Dumbino_r1>Overlapper-Blender_r1+1300MB.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.01.wrd.sorted" "Gibson, William - Neuromancer.txt.01.wrd.sorted"
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 77841
Size of 2nd input file: 77250
Allocating 1300MB ...
Lines in 1st input file: 8967
Lines in 2nd input file: 8851
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 17818 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 9029
Overlapped lines, i.e. lines common for both files: 8789
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 178
```

```
D:\Dumbino_r1>ren Unfamiliar.txt Unfamiliar.1.txt

D:\Dumbino_r1>Overlapper-Blender_r1+1300MB.exe "Gibson, William - Cyberpunk 1 - Neuromancer.txt.02.wrd.sorted" "Gibson, William - Neuromancer.txt.02.wrd.sorted"
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 465519
Size of 2nd input file: 464212
Allocating 1300MB ...
Lines in 1st input file: 37067
Lines in 2nd input file: 36882
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 73949 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 37304
Overlapped lines, i.e. lines common for both files: 36645
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 422

D:\Dumbino_r1>ren Unfamiliar.txt Unfamiliar.2.txt

Done.
What-is-what notes:
The goal is to proof-read the file 'Gibson, William - Cyberpunk 1 - Neuromancer.txt' by getting all its words not to be found in another edition of the e-book 'Gibson, William
- Neuromancer.txt'. Alongside with the misspelled words ('Unfamiliar.1.spell-checked.txt') it eases the proof-reading.
File 'Unfamiliar.1.spell-checked.txt' contains all misspelled words in 'Gibson, William - Cyberpunk 1 - Neuromancer.txt'.
File 'Unfamiliar.1.txt' contains all 1-grams in 'Gibson, William - Cyberpunk 1 - Neuromancer.txt' not to be found in 'Gibson, William - Neuromancer.txt'.
File 'Unfamiliar.2.txt' contains all 2-grams in 'Gibson, William - Cyberpunk 1 - Neuromancer.txt' not to be found in 'Gibson, William - Neuromancer.txt'.


D:\Dumbino_r1>dir Unfamiliar*.txt
 Volume in drive D is S640_Vol5
 Volume Serial Number is F85D-148B


 Directory of D:\Dumbino_r1


03/29/2012  11:33 PM             6,229 Unfamiliar.1.spell-checked.txt
03/29/2012  11:33 PM             1,280 Unfamiliar.1.txt
03/29/2012  11:33 PM             5,089 Unfamiliar.2.txt
               3 File(s)         12,598 bytes
               0 Dir(s)  27,881,725,952 bytes free


D:\Dumbino_r1>
```

```
D:\Dumbino_r1>dir/og/on

03/30/2012  05:28 AM    <DIR>          Logo
03/30/2012  07:18 AM             5,457 Am-am.bat
03/30/2012  07:18 AM             8,253 Dumbino_26Clash_4-grams.BAT
03/30/2012  07:18 AM                 0 Empty
03/30/2012  07:18 AM         4,024,155 english.dic_351116_wordlist.txt
03/30/2012  07:18 AM                 2 Enter
03/30/2012  07:18 AM           460,867 Gibson, William - Cyberpunk 1 - Neuromancer.txt
03/30/2012  07:18 AM           460,486 Gibson, William - Neuromancer.txt
03/30/2012  07:18 AM       409,829,386 googlebooks-eng-us-all-4gram-20090715-graffith_A_distinct
03/30/2012  07:18 AM       149,298,133 googlebooks-eng-us-all-4gram-20090715-graffith_B_distinct
03/30/2012  07:18 AM       151,969,755 googlebooks-eng-us-all-4gram-20090715-graffith_C_distinct
03/30/2012  07:18 AM        92,266,425 googlebooks-eng-us-all-4gram-20090715-graffith_D_distinct
03/30/2012  07:18 AM        83,849,606 googlebooks-eng-us-all-4gram-20090715-graffith_E_distinct
03/30/2012  07:18 AM       122,493,889 googlebooks-eng-us-all-4gram-20090715-graffith_F_distinct
03/30/2012  07:18 AM        48,570,461 googlebooks-eng-us-all-4gram-20090715-graffith_G_distinct
03/30/2012  07:18 AM       150,628,233 googlebooks-eng-us-all-4gram-20090715-graffith_H_distinct
03/30/2012  07:18 AM       213,094,578 googlebooks-eng-us-all-4gram-20090715-graffith_I_distinct
03/30/2012  07:18 AM        10,033,769 googlebooks-eng-us-all-4gram-20090715-graffith_J_distinct
03/30/2012  07:18 AM        12,627,224 googlebooks-eng-us-all-4gram-20090715-graffith_K_distinct
03/30/2012  07:18 AM        70,317,358 googlebooks-eng-us-all-4gram-20090715-graffith_L_distinct
03/30/2012  07:18 AM       119,240,995 googlebooks-eng-us-all-4gram-20090715-graffith_M_distinct
03/30/2012  07:18 AM        69,802,440 googlebooks-eng-us-all-4gram-20090715-graffith_N_distinct
03/30/2012  07:18 AM       240,428,287 googlebooks-eng-us-all-4gram-20090715-graffith_O_distinct
03/30/2012  07:18 AM       128,166,172 googlebooks-eng-us-all-4gram-20090715-graffith_P_distinct
03/30/2012  07:18 AM         6,345,892 googlebooks-eng-us-all-4gram-20090715-graffith_Q_distinct
03/30/2012  07:18 AM        86,795,262 googlebooks-eng-us-all-4gram-20090715-graffith_R_distinct
03/30/2012  07:18 AM       203,420,884 googlebooks-eng-us-all-4gram-20090715-graffith_S_distinct
03/30/2012  07:18 AM       560,863,997 googlebooks-eng-us-all-4gram-20090715-graffith_T_distinct
03/30/2012  07:18 AM        37,451,032 googlebooks-eng-us-all-4gram-20090715-graffith_U_distinct
03/30/2012  07:18 AM        22,690,873 googlebooks-eng-us-all-4gram-20090715-graffith_V_distinct
03/30/2012  07:18 AM       221,709,660 googlebooks-eng-us-all-4gram-20090715-graffith_W_distinct
03/30/2012  07:18 AM           128,966 googlebooks-eng-us-all-4gram-20090715-graffith_X_distinct
03/30/2012  07:18 AM        21,288,488 googlebooks-eng-us-all-4gram-20090715-graffith_Y_distinct
03/30/2012  07:18 AM           436,576 googlebooks-eng-us-all-4gram-20090715-graffith_Z_distinct
03/30/2012  07:18 AM             1,570 KAZE prompt.lnk
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_01_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_02_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_03_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_04_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_05_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_06_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_07_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_08_01p.c
```

```
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_09_01p.c
03/30/2012  07:18 AM           315,203 Leprechaun_x-leton_10_01p.c
03/30/2012  07:18 AM           103,936 Leprechaun_x-leton_32bit_01_01p.exe
03/30/2012  07:18 AM           104,448 Leprechaun_x-leton_32bit_02_01p.exe
03/30/2012  07:18 AM           104,448 Leprechaun_x-leton_32bit_03_01p.exe
03/30/2012  07:18 AM           104,960 Leprechaun_x-leton_32bit_04_01p.exe
03/30/2012  07:18 AM           105,984 Leprechaun_x-leton_32bit_05_01p.exe
03/30/2012  07:18 AM           106,496 Leprechaun_x-leton_32bit_06_01p.exe
03/30/2012  07:18 AM           106,496 Leprechaun_x-leton_32bit_07_01p.exe
03/30/2012  07:18 AM           107,520 Leprechaun_x-leton_32bit_08_01p.exe
03/30/2012  07:18 AM           107,520 Leprechaun_x-leton_32bit_09_01p.exe
03/30/2012  07:18 AM           108,032 Leprechaun_x-leton_32bit_10_01p.exe
03/30/2012  07:18 AM            44,074 Overlapper-Blender_r1+.c
03/30/2012  07:18 AM            66,048 Overlapper-Blender_r1+1300MB.exe
03/30/2012  07:18 AM            70,656 QuickSortExternal_4+GB_32bit_ascending.exe
03/30/2012  07:18 AM            70,656 QuickSortExternal_4+GB_32bit_descending.exe
03/30/2012  07:18 AM           107,063 QuickSortExternal_4+GB_ascending.c
03/30/2012  07:18 AM           107,180 QuickSortExternal_4+GB_descending.c

D:\Dumbino_r1>Dumbino_26Clash_4-grams.BAT
Usage: Dumbino_26Clash_4-grams.BAT yourtextfile
Purpose: Creates three files:
- yourtextfile_overlapped_all_distinct.txt
- yourtextfile_unfamiliar_all_distinct.txt
- yourtextfile_progenitor_all_distinct.txt
First contains all 4-grams from yourtextfile to be found in the corpus being used.
Second contains all 4-grams from yourtextfile not to be found in the corpus being used.
Third contains all 4-grams from yourtextfile.


D:\Dumbino_r1>Dumbino_26Clash_4-grams.BAT "Gibson, William - Cyberpunk 1 - Neuromancer.txt"

Leprechaun_quadrupleton (Fast-In-Future Greedy n-gram-Ripper), rev. 15FIXFIX, written by Svalqyatchx.
Purpose: Rips all distinct 4-grams (4-word phrases) with length 13..51 chars from incoming texts.
Feature1: All words within x-lets/n-grams are in range 1..31 chars inclusive.
Feature2: In this revision 128MB 1-way hash is used which results in 16,777,216 external B-Trees of order 3.
Feature3: In this revision 1 pass is to be made.
Feature4: If the external memory has latency 99+microseconds then !(look no further), IOPS(seek-time) rules.
Pass #1 of 1:
Size of input file with files for Leprechauning: 49
Allocating HASH memory 134,217,793 bytes ... OK
Allocating memory 293MB ... OK
Size of Input TEXTual file: 460,867
|; 00,043,272P/s; Phrase count: 43,272 of them 42,208 distinct; Done: 64/64
Bytes per second performance: 460,867B/s
Phrases per second performance: 43,272P/s
```

```
Time for putting phrases into trees: 1 second(s)
Flushing UNsorted phrases: 100%; Shaking trees performance: 00,084,416P/s
Time for shaking phrases from trees: 1 second(s)
Leprechaun: Current pass done.

Total memory needed for one pass: 5,597KB
Total distinct phrases: 42,208
Total time: 1 second(s)
Total performance: 43,272P/s i.e. phrases per second
Leprechaun: Done.

Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 409829386
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 17981107
Allocated memory for pointers-to-words in MB: 69
Allocated memory for pointers-to-words in MB: 1
Sorting 18023315 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 18021681
Overlapped lines, i.e. lines common for both files: 1634
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42205
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 149298133
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 6571872
Allocated memory for pointers-to-words in MB: 26
Allocated memory for pointers-to-words in MB: 1
Sorting 6614080 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 6613475
Overlapped lines, i.e. lines common for both files: 605
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 151969755
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 6212540
Allocated memory for pointers-to-words in MB: 24
Allocated memory for pointers-to-words in MB: 1
Sorting 6254748 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 6254383
Overlapped lines, i.e. lines common for both files: 365
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 92266425
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 3856617
Allocated memory for pointers-to-words in MB: 15
Allocated memory for pointers-to-words in MB: 1
Sorting 3898825 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 3898466
Overlapped lines, i.e. lines common for both files: 359
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 83849606
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 3424994
Allocated memory for pointers-to-words in MB: 14
Allocated memory for pointers-to-words in MB: 1
Sorting 3467202 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 3467039
Overlapped lines, i.e. lines common for both files: 163
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42205
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 122493889
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 5282784
Allocated memory for pointers-to-words in MB: 21
Allocated memory for pointers-to-words in MB: 1
Sorting 5324992 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 5324511
Overlapped lines, i.e. lines common for both files: 481
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42205
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 48570461
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 2116401
Allocated memory for pointers-to-words in MB: 9
Allocated memory for pointers-to-words in MB: 1
Sorting 2158609 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 2158358
Overlapped lines, i.e. lines common for both files: 251
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 150628233
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 6760278
Allocated memory for pointers-to-words in MB: 26
Allocated memory for pointers-to-words in MB: 1
Sorting 6802486 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 6801085
Overlapped lines, i.e. lines common for both files: 1401
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 213094578
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 9449270
Allocated memory for pointers-to-words in MB: 37
Allocated memory for pointers-to-words in MB: 1
Sorting 9491478 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 9490385
Overlapped lines, i.e. lines common for both files: 1093
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 10033769
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 444251
Allocated memory for pointers-to-words in MB: 2
Allocated memory for pointers-to-words in MB: 1
Sorting 486459 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 486408
Overlapped lines, i.e. lines common for both files: 51
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 12627224
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 569361
Allocated memory for pointers-to-words in MB: 3
Allocated memory for pointers-to-words in MB: 1
Sorting 611569 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 611476
Overlapped lines, i.e. lines common for both files: 93
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42205
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 70317358
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 3123807
Allocated memory for pointers-to-words in MB: 13
Allocated memory for pointers-to-words in MB: 1
Sorting 3166015 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 3165664
Overlapped lines, i.e. lines common for both files: 351
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 119240995
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 5180952
Allocated memory for pointers-to-words in MB: 20
Allocated memory for pointers-to-words in MB: 1
Sorting 5223160 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 5222856
Overlapped lines, i.e. lines common for both files: 304
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 69802440
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 3075105
Allocated memory for pointers-to-words in MB: 12
Allocated memory for pointers-to-words in MB: 1
Sorting 3117313 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 3117127
Overlapped lines, i.e. lines common for both files: 186
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 240428287
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 10718140
Allocated memory for pointers-to-words in MB: 42
Allocated memory for pointers-to-words in MB: 1
Sorting 10760348 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 10759527
Overlapped lines, i.e. lines common for both files: 821
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 128166172
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 5222828
Allocated memory for pointers-to-words in MB: 21
Allocated memory for pointers-to-words in MB: 1
Sorting 5265036 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 5264800
Overlapped lines, i.e. lines common for both files: 236
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 6345892
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 257343
Allocated memory for pointers-to-words in MB: 2
Allocated memory for pointers-to-words in MB: 1
Sorting 299551 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 299544
Overlapped lines, i.e. lines common for both files: 7
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 86795262
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 3565405
Allocated memory for pointers-to-words in MB: 14
Allocated memory for pointers-to-words in MB: 1
Sorting 3607613 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 3607384
Overlapped lines, i.e. lines common for both files: 229
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 203420884
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 8736465
Allocated memory for pointers-to-words in MB: 34
Allocated memory for pointers-to-words in MB: 1
Sorting 8778673 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 8777651
Overlapped lines, i.e. lines common for both files: 1022
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 560863997
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 24309233
Allocated memory for pointers-to-words in MB: 93
Allocated memory for pointers-to-words in MB: 1
Sorting 24351441 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 24348709
Overlapped lines, i.e. lines common for both files: 2732
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42206
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 37451032
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 1640327
Allocated memory for pointers-to-words in MB: 7
Allocated memory for pointers-to-words in MB: 1
Sorting 1682535 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 1682407
Overlapped lines, i.e. lines common for both files: 128
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42205
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 22690873
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 957759
Allocated memory for pointers-to-words in MB: 4
Allocated memory for pointers-to-words in MB: 1
Sorting 999967 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 999910
Overlapped lines, i.e. lines common for both files: 57
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 221709660
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 9738971
Allocated memory for pointers-to-words in MB: 38
Allocated memory for pointers-to-words in MB: 1
Sorting 9781179 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 9780054
Overlapped lines, i.e. lines common for both files: 1125
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 128966
Allocating 1300MB ...
```

```
Lines in 1st input file: 42208
Lines in 2nd input file: 6593
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 48801 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 48801
Overlapped lines, i.e. lines common for both files: 0
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42208
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 21288488
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 1000248
Allocated memory for pointers-to-words in MB: 4
Allocated memory for pointers-to-words in MB: 1
Sorting 1042456 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 1042180
Overlapped lines, i.e. lines common for both files: 276
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42207
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 436576
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 19684
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 61892 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 61892
Overlapped lines, i.e. lines common for both files: 0
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 42208

 Volume in drive D is S640_Vol5
 Volume Serial Number is F85D-148B
```

```
 Directory of D:\Dumbino_r1

03/30/2012  07:19 AM            32,128 googlebooks-eng-us-all-4gram-20090715-graffith_A_overlapped
03/30/2012  07:19 AM            12,018 googlebooks-eng-us-all-4gram-20090715-graffith_B_overlapped
03/30/2012  07:19 AM             7,686 googlebooks-eng-us-all-4gram-20090715-graffith_C_overlapped
03/30/2012  07:19 AM             7,072 googlebooks-eng-us-all-4gram-20090715-graffith_D_overlapped
03/30/2012  07:19 AM             3,482 googlebooks-eng-us-all-4gram-20090715-graffith_E_overlapped
03/30/2012  07:19 AM             9,861 googlebooks-eng-us-all-4gram-20090715-graffith_F_overlapped
03/30/2012  07:19 AM             4,932 googlebooks-eng-us-all-4gram-20090715-graffith_G_overlapped
03/30/2012  07:20 AM            27,669 googlebooks-eng-us-all-4gram-20090715-graffith_H_overlapped
03/30/2012  07:20 AM            20,390 googlebooks-eng-us-all-4gram-20090715-graffith_I_overlapped
03/30/2012  07:20 AM             1,015 googlebooks-eng-us-all-4gram-20090715-graffith_J_overlapped
03/30/2012  07:20 AM             1,829 googlebooks-eng-us-all-4gram-20090715-graffith_K_overlapped
03/30/2012  07:20 AM             6,991 googlebooks-eng-us-all-4gram-20090715-graffith_L_overlapped
03/30/2012  07:20 AM             6,060 googlebooks-eng-us-all-4gram-20090715-graffith_M_overlapped
03/30/2012  07:20 AM             3,703 googlebooks-eng-us-all-4gram-20090715-graffith_N_overlapped
03/30/2012  07:20 AM            15,726 googlebooks-eng-us-all-4gram-20090715-graffith_O_overlapped
03/30/2012  07:20 AM             5,059 googlebooks-eng-us-all-4gram-20090715-graffith_P_overlapped
03/30/2012  07:20 AM               144 googlebooks-eng-us-all-4gram-20090715-graffith_Q_overlapped
03/30/2012  07:21 AM             4,882 googlebooks-eng-us-all-4gram-20090715-graffith_R_overlapped
03/30/2012  07:21 AM            21,089 googlebooks-eng-us-all-4gram-20090715-graffith_S_overlapped
03/30/2012  07:21 AM            54,294 googlebooks-eng-us-all-4gram-20090715-graffith_T_overlapped
03/30/2012  07:21 AM             2,368 googlebooks-eng-us-all-4gram-20090715-graffith_U_overlapped
03/30/2012  07:21 AM             1,202 googlebooks-eng-us-all-4gram-20090715-graffith_V_overlapped
03/30/2012  07:22 AM            22,639 googlebooks-eng-us-all-4gram-20090715-graffith_W_overlapped
03/30/2012  07:22 AM                 0 googlebooks-eng-us-all-4gram-20090715-graffith_X_overlapped
03/30/2012  07:22 AM             5,175 googlebooks-eng-us-all-4gram-20090715-graffith_Y_overlapped
03/30/2012  07:22 AM                 0 googlebooks-eng-us-all-4gram-20090715-graffith_Z_overlapped
              26 File(s)        277,414 bytes
               0 Dir(s)  21,766,774,784 bytes free

googlebooks-eng-us-all-4gram-20090715-graffith_A_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_B_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_C_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_D_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_E_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_F_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_G_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_H_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_I_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_J_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_K_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_L_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_M_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_N_overlapped
```

```
googlebooks-eng-us-all-4gram-20090715-graffith_O_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_P_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_Q_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_R_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_S_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_T_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_U_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_V_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_W_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_X_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_Y_overlapped
googlebooks-eng-us-all-4gram-20090715-graffith_Z_overlapped
        1 file(s) copied.

Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 277414
Size of 2nd input file: 0
Allocating 1300MB ...
Lines in 1st input file: 13970
Lines in 2nd input file: 0
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 13970 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 13970
Overlapped lines, i.e. lines common for both files: 0
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 0

Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 0
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 0
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 42208 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 42208
Overlapped lines, i.e. lines common for both files: 0
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 0
```

```
Overlapper-Blender r.1+1300MB, written by Kaze.
Size of 1st input file: 928133
Size of 2nd input file: 277414
Allocating 1300MB ...
Lines in 1st input file: 42208
Lines in 2nd input file: 13970
Allocated memory for pointers-to-words in MB: 1
Allocated memory for pointers-to-words in MB: 1
Sorting 56178 Pointers ...
Deduplicating duplicates and dumping all into 'Blended.txt' ...
Dumping deduplicated duplicates into 'Overlapped.txt' ...
Dumping all-from-first-file except deduplicated duplicates into 'Unfamiliar.txt' ...
Blended lines, i.e. combined lines from both files: 42208
Overlapped lines, i.e. lines common for both files: 13970
Unfamiliar lines, i.e. lines from 1st file not encountered in 2nd file: 28238


 Volume in drive D is S640_Vol5
 Volume Serial Number is F85D-148B


 Directory of D:\Dumbino_r1


03/30/2012  07:22 AM           277,414 Gibson, William - Cyberpunk 1 - Neuromancer.txt_overlapped_all_distinct.txt
03/30/2012  07:22 AM           928,133 Gibson, William - Cyberpunk 1 - Neuromancer.txt_progenitor_all_distinct.txt
03/30/2012  07:22 AM           650,719 Gibson, William - Cyberpunk 1 - Neuromancer.txt_unfamiliar_all_distinct.txt
               3 File(s)      1,856,266 bytes
               0 Dir(s)  21,765,726,208 bytes free

The current time is:  7:18:50.23
Enter the new time:
The current time is:  7:22:08.70
Enter the new time:

D:\Dumbino_r1>type "Gibson, William - Cyberpunk 1 - Neuromancer.txt_unfamiliar_all_distinct.txt"|more
a_awake_in_straylight
a_background_of_twisted
a_bad_hangover_as
a_bahamian_orbital_bank
a_bal_loon_tired
a_bama_rapid_deployment
a_band_of_printed
a_band_to_match
a_bank_in_wichita
a_bar_for_professional
a_bar_she_knew
```

```
a_baroque_thing_for
a_battered_tricycle_truck
a_bedside_ashtray_after
a_beer_vendor_was
a_big_tube_and
a_biochemical_governing_the
a_black_automatic_pistol
a_black_clinic_in
a_black_expanse_where
a_black_glass_bank
a_black_hilton_tray
a_black_nylon_shoulder
a_black_sensor_set
a_black_storage_unit
a_black_velvet_slipper
a_block_down_baiitsu
a_block_from_deane
a_block_of_polycarbon
a_blond_lightning_bolt
a_blood_flecked_bag
a_bloody_hand_down
a_blue_derm_inside
a_blue_neon_replica
a_blue_plastic_syrette
a_blue_sanyo_vacuum
a_blunt_white_spindle
a_blur_of_blond
a_boardroom_the_size
a_body_grown_in
a_bonbon_and_stripped
a_booted_foot_up
a_bootheel_scraped_the
a_bored_researcher_who
a_bought_twenty_world
a_bracelet_of_flesh
a_brand_of_hypnotic
a_brass_plate_mounted
a_brass_table_beside
a_braun_coffee_maker
a_brazilian_kid_called
a_brazilian_payroll_net
a_breast_brushed_his
a_breeze_caught_at
a_brick_of_wage
a_bridge_or_overpass
```

```
a_briefcase_of_soft
a_bright_magenta_splinter
a_bright_nine_pointed
a_bright_red_mug
a_brilliant_slash_of
a_broad_puddle_of
a_broad_rectangular_pond
a_broad_shallow_curve
^C
D:\Dumbino_r1>
```