



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

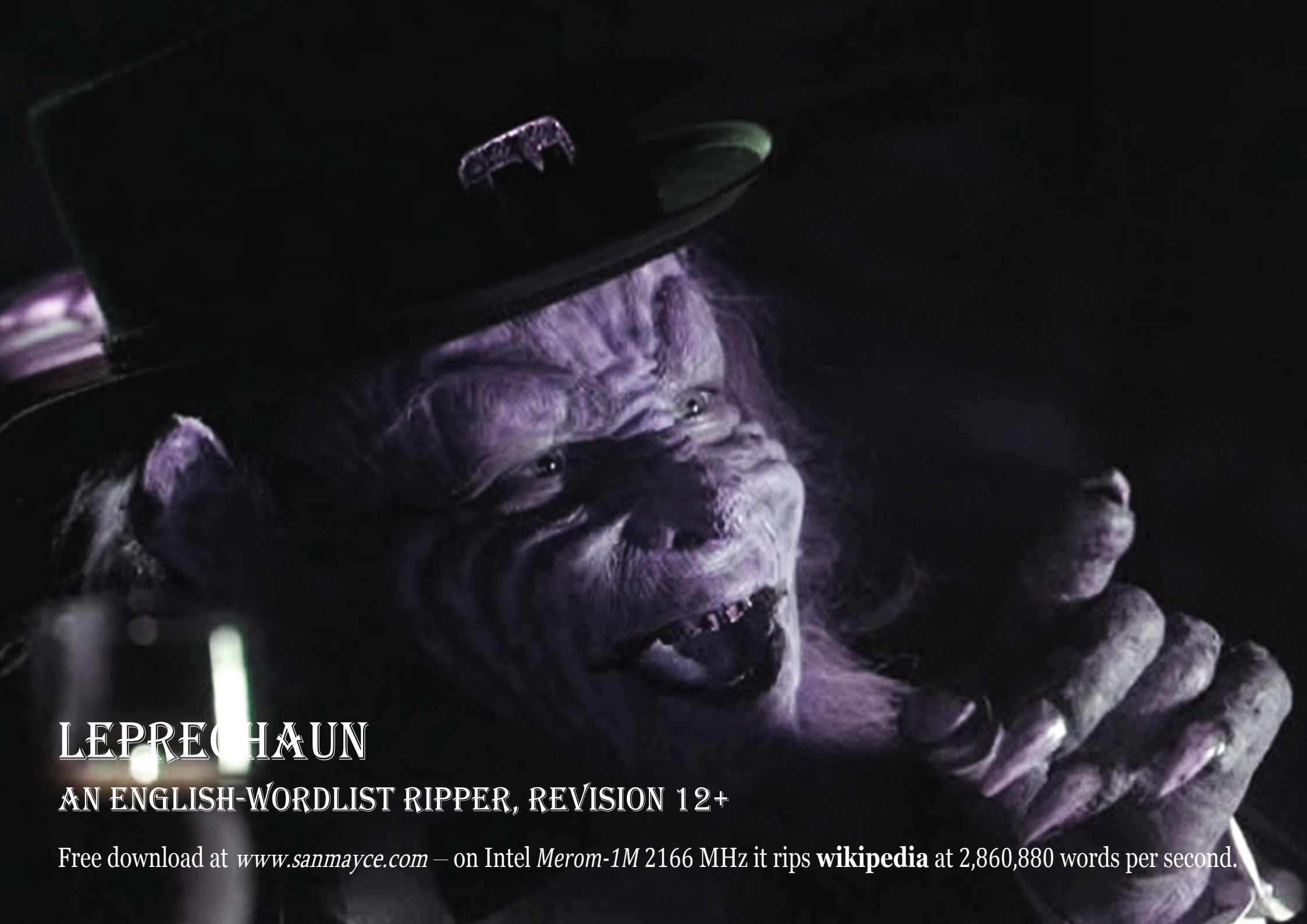
Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.



LEPRECHAUN

AN ENGLISH-WORDLIST RIPPER, REVISION 12+

Free download at www.sanmayce.com — on Intel Merom-1M 2166 MHz it rips **wikipedia** at 2,860,880 words per second.

D:\Leprechaun12+>Leprechaun.exe

Leprechaun(Fast and Greedy Word_Ripper), revision 12+, written by Svalqyatchx.
Leprechaun: 'Oh, well, didn't you hear? Bigger is good, but jumbo is dear.'

'The Little Monster' short notes:

Note1: I wish to thank to R.N. Horspool, Ranjan Sinha, Dmitry Shkarin, Michael Abrash, J. Bentley, R. Sedgewick, Igor Pavlov for sharing their knowledge to public.

Note2: Run it without parameters to get usage and short notes:

Note3: This simple amateurish(more over I am not versed well neither in C nor in mathematics nor in english language, but I am persistent in INDEXING GBs of english TEXTS) tool is written in ANSI C(at least its source is compileable for CL(Windows) and GCC(Linux)), and its purpose is to create a wordList for a group of files(given via filelist).
Its name comes(according to Heritage Dictionary) from 'low corpus' or 'little body', in fact from amazing movie saga 'Leprechaun 1-2-3-4-5-6' starring by Warwick Davis.

Note4: Only words up to 31 chars are proceeded.

Note5: Cursor hiding in C - mission impossible for me.

Note6: By default(third parameter is 1023) allocated memory is 364MB.
Due to 'malloc()' limitation under WINDOWS, maximum value of third parameter is 5400 which is 1920MB allocated block.

Note7: File Leprechaun.LOG is a log, where new statistics are appended.

Note8: Revision 12+ can handle files larger than 4GB.

Note9: For 'H:\>Leprechaun.exe static.wikipedia.org_downloads_2008-06_en.lst
wikipedia-en-html.tar.wrd 5400'

where 223,674,511,360 wikipedia-en-html.tar
on laptop Toshiba Pentium T3400 2166 MHZ with
Motherboard Name:

Toshiba Satellite L305

CPU Type: Mobile DualCore Intel Pentium, 2166 MHZ (13 x 167)

CPU Alias: Merom-1M

L1 Code Cache: 32 KB per core

L1 Data Cache: 32 KB per core

L2 Cache: 1 MB (On-Die, ECC, ASC, Full-Speed)

Bus Type: Dual DDR2 SDRAM

Bus Width: 128-bit

Real Clock: 333 MHZ (DDR)

Effective Clock: 666 MHz
EVEREST v5.00.1650 Memory Copy: 3725MB/s with timings 5-5-5-13
result is logged to 'Leprechaun.LOG':
Bytes per second performance: 20,658,955B/s
Words per second performance: 2,860,880W/s
Input File with a list of TEXTual Files:
static.wikipedia.org_downloads_2008-06_en.lst
Size of all TEXTual Files: 223,674,511,360
Word count: 30,974,750,142 of them 12,561,874 distinct
Number Of Files: 1
Number Of Lines: 2088618575
Allocated memory in MB: 1920
Words with length 01 occupy 0,033KB of 0,349KB given i.e. 09% utilization
Words with length 02 occupy 0,033KB of 0,349KB given i.e. 09% utilization
Words with length 03 occupy 0,037KB of 0,697KB given i.e. 05% utilization
Words with length 04 occupy 0,151KB of 0,871KB given i.e. 17% utilization
Words with length 05 occupy 0,744KB of 1,568KB given i.e. 47% utilization
Words with length 06 occupy 1,470KB of 3,136KB given i.e. 46% utilization
Words with length 07 occupy 2,605KB of 5,923KB given i.e. 43% utilization
Words with length 08 occupy 3,296KB of 6,968KB given i.e. 47% utilization
Words with length 09 occupy 3,714KB of 6,968KB given i.e. 53% utilization
Words with length 10 occupy 3,483KB of 6,968KB given i.e. 49% utilization
Words with length 11 occupy 3,235KB of 5,923KB given i.e. 54% utilization
Words with length 12 occupy 2,691KB of 4,181KB given i.e. 64% utilization
Words with length 13 occupy 2,230KB of 3,484KB given i.e. 64% utilization
Words with length 14 occupy 1,718KB of 3,484KB given i.e. 49% utilization
Words with length 15 occupy 1,357KB of 2,613KB given i.e. 51% utilization
Words with length 16 occupy 1,063KB of 2,613KB given i.e. 40% utilization
Words with length 17 occupy 0,814KB of 1,742KB given i.e. 46% utilization
Words with length 18 occupy 0,617KB of 1,742KB given i.e. 35% utilization
Words with length 19 occupy 0,485KB of 1,742KB given i.e. 27% utilization
Words with length 20 occupy 0,402KB of 1,742KB given i.e. 23% utilization
Words with length 21 occupy 0,327KB of 1,742KB given i.e. 18% utilization
Words with length 22 occupy 0,274KB of 1,742KB given i.e. 15% utilization
Words with length 23 occupy 0,224KB of 1,394KB given i.e. 16% utilization
Words with length 24 occupy 0,190KB of 1,394KB given i.e. 13% utilization
Words with length 25 occupy 0,162KB of 1,394KB given i.e. 11% utilization

Words with length 26 occupy 0,136KB of 1,220KB given i.e. 11% utilization
Words with length 27 occupy 0,119KB of 1,046KB given i.e. 11% utilization
Words with length 28 occupy 0,107KB of 0,871KB given i.e. 12% utilization
Words with length 29 occupy 0,091KB of 0,697KB given i.e. 13% utilization
Words with length 30 occupy 0,080KB of 0,523KB given i.e. 15% utilization
Words with length 31 occupy 0,076KB of 0,523KB given i.e. 14% utilization
Total pseudo(including hash table) memory utilization: 42%
Total real(wordlist's words VS allocated block) memory utilization: 60/1000
Used value for third parameter in KB: 5400
Use next time as third parameter: 3475-
Time for making unsorted wordlist: 10827 second(s)
Time for sorting unsorted wordlist: 10 second(s)

Usage: Leprechaun InFile OutFile [BufferSize] [SortMethod]
<InFile>: Input file with files for Leprechauning, in WINDOWS console
you can create it by 'E:\KAZEHOME>dir *.txt/s/b>Leprechaun.lst'
<OutFile>: Output WORDLIST(sorted since r.9, CRLF) file
<BufferSize>: Optional Dynamic RAM buffer in KB, default(and minimum
in the same time) is 1023, i.e. omit or specify greater one
<SortMethod>: Optional Sort Method, default is 'D',
A - InsertionSort
B - InsertionX26Sort
C - MultiKeyQuickSortSort by J. Bentley, R. Sedgewick
D - MultiKeyQuickSortX26Sort' by J. Bentley, R. Sedgewick

Have a nice Leprechauning.
For contacts: sanmayce@hotmail.com
Sanmayce Svalqyatchx 'Kaze', 2005 Feb 07(rev.12+: 2010 Feb 21).

D:\Leprechaun12+>

Visual Studio 2008 Com...

Visual Studio 2008 x64 ...

foobar2000

UltraEdit

XviD4PSP 5.0

AIWebMenus 5 PRO

Mozilla Firefox

Visual C++ Toolkit 20...

7-Zip File Manager

Visual C++ Toolkit 2003 Command Prompt

Volume Serial Number is 1C0C-035D

Directory of H:\

02/21/2010 09:01 PM 2 ENTER

02/23/2010 02:36 AM 61,440 Leprechaun.exe

02/22/2010 05:10 PM 3,000 Leprechaun.LOG

02/22/2010 12:26 AM 23 static.wikipedia.org_downloads_2008-06_en.lst

02/21/2010 09:03 PM 55 TestReadSpeed.bat

06/18/2008 08:02 PM 223,674,511,360 wikipedia-en-html.tar

02/22/2010 05:10 PM 146,973,879 wikipedia-en-html.tar.wrd.old

7 File(s) 223,821,549,759 bytes

0 Dir(s) 96,171,352,064 bytes free

H:\>Leprechaun.exe static.wikipedia.org_downloads_2008-06_en.lst wikipedia-en-html.tar.wrd 5400

Leprechaun(Fast and Greedy Word Ripper), revision 12+, written by \$valqyatchx.

Leprechaun: 'Oh, well, didn't you hear? Bigger is good, but jumbo is dear.'

Size of input file with files for Leprechauning: 23

Allocated memory in MB: 1920

Size of Input TEXTual file: 223,674,511,360

Word count: 30,974,750,142 of them 12,561,874 distinct; Done: 64/64

Flushing unsorted words ...

Time for making unsorted wordlist: 10827 second(s)

Deallocated memory in MB: 1920

Allocated memory for words in MB: 141

Allocated memory for pointers-to-words in MB: 48

Sorting(with 'MultiKeyQuickSortX26Sort' by J. Bentley and R. Sedgewick) ...

Sort pass 26/26 ...

Flushing sorted words ...

Time for sorting unsorted wordlist: 10 second(s)

Leprechaun: Done.

H:\>dir

Volume in drive H is Sanmayce_Hitachi-298GB

Volume Serial Number is 1C0C-035D

Directory of H:\

02/21/2010 09:01 PM 2 ENTER

02/23/2010 02:36 AM 61,440 Leprechaun.exe

02/23/2010 05:51 AM 6,007 Leprechaun.LOG

02/22/2010 12:26 AM 23 static.wikipedia.org_downloads_2008-06_en.lst

02/21/2010 09:03 PM 55 TestReadSpeed.bat

06/18/2008 08:02 PM 223,674,511,360 wikipedia-en-html.tar

02/23/2010 05:51 AM 146,973,879 wikipedia-en-html.tar.wrd

02/22/2010 05:10 PM 146,973,879 wikipedia-en-html.tar.wrd.old

8 File(s) 223,968,526,645 bytes

0 Dir(s) 96,024,371,200 bytes free

H:\>

Lavalys EVEREST CPUID

Processor	Mobile DualCore Intel Pentium T3400	65 nm
Code Name	Merom-1M	
Platform	Socket 479	
Stepping	M0	
CPUID Vendor	GenuineIntel	
CPUID Name	Intel(R) Pentium(R) Dual CPU T3400 @ 2.16GHz	

CPUID Rev.	6 F D	Core Voltage
CPU Clock	2161.5 MHz	L1 Instr. Cache
Multiplier	13x	L1 Data Cache
FSB Clock	166.3 MHz	L2 Cache
FSB Speed	665.1 MHz	L3 Cache

Instruction Set	x86, x86-64, MMX, SSE, SSE2, SSE3, SSSE3	
Motherboard	Toshiba Satellite L305	
Chipset	Intel Cantiga GL40	
Integr. Video	Active (Intel GMA 4500M)	
Memory Type	Dual Channel DDR2-667 SDRAM (5-5-5-13)	
Memory Clock	332.5 MHz	DRAM:FSB

CPU #1 / Core #1 EVEREST v5.00.1650

Lavalys EVEREST Cache & Memory Benchmark

	Read	Write	Copy	Latency
Memory	4627 MB/s	3362 MB/s	3689 MB/s	88.1 ns
L1 Cache	34566 MB/s	34397 MB/s	68807 MB/s	1.4 ns
L2 Cache	16394 MB/s	12852 MB/s	15754 MB/s	2.9 ns
L3 Cache				

CPU Type	Mobile DualCore Intel Pentium T3400 (Merom-1M, Socket 479)		
CPU Clock	2161.5 MHz (original: 2166 MHz)		
CPU FSB	166.3 MHz (original: 166 MHz)		
CPU Multiplier	13x	CPU Stepping	M0

Memory Bus	332.5 MHz	DRAM:FSB Ratio	10:5
Memory Type	Dual Channel DDR2-667 SDRAM (5-5-5-13)		
Chipset	Intel Cantiga GL40		
Motherboard	Toshiba Satellite L305		

EVEREST v5.00.1650 / BenchDLL 2.4.258.0 (c) 2003-2009 Lavalys, Inc.

Start

Visual C++ Toolkit 2003 C...

EVEREST CPUID

EVEREST Cache & ...

EN 1.4 37 30 40 40 5:58 AM