Project '**GAMERA**' *revision 14*

(Semi-Powerful Console (Windows & Linux) Tools & gigabytes of English texts, downloadable from www.sanmayce.com)

WHERE THE WORD COUNTS

Yoshi (Filelist creator and more, 32bit console application), revision **06**
GRAFFITH (Text decompressor-finder-dumper, 32bit console application), revision **2++_Graphein**
Leprechaun[_quadrupleton] (Fast and Greedy Word_Ripper, 32bit console application), revision **13_7pluses_FIX**

WinRAR archive in 25 624MB volumes • Required HDD space: 14.9 GB (*ready to go when extracted on* **D:**\) • 2011 APR 21

Leprechaun rips wikipedia-en-html.tar at 5,235,758 words-per-second rate (Obtained with Toshiba Satellite L305 (Intel Pentium T3400 2.16GHz))
GRAFFITH decodes at 2x10MB/s and searches with wildcards at 70+MB/s rate (Obtained with Toshiba Satellite L305 (Intel Pentium T3400 2.16GHz))

LBL stands for Line-By-Line (GRAMMATICAL ENGLISH LINES) i.e. sentences not merely CRLF or LF lines!
.LBL files are made from .TXT files which are made from respective .DOC, .RTF, .LIT, .PDF, .CHM, .HTM[L], .DJV[U] files;
Number and size of *.LBL files: 562,504 files (26GB or 27,991,747,152 bytes);
Number and size of *.LBL (with max-line-length=2048chars) files: 1,204 LBL/BSC (27,709,189,566/5,731,810,202 bytes) files;
Lines and words in *.LBL files: 424,754,717 lines (with 4,582,451,898 words of them 9,177,221 distinct);
Lines and words in *.LBL (with max-line-length=2048chars) files: **424,711,401** lines (with 4,550,603,240 words of them 8,625,789 distinct);
Quadruplets/4grams in **1,204** BSC (9,848,192,722 bytes) files: 2,099,828,905 distinct-per-file (48,322,089,371 bytes) from **2,710,601,882** all;
Add-on: Quadruplets/4grams with simple-ranking in **164** BSC (3,571,395,313 bytes) files: **815,381,888** distinct (27,944,239,656 bytes).