UMSMailStat

Olaf Peters

COLLABORATORS			
	TITLE : UMSMailStat		
ACTION	*NAME*	*DATE*	*SIGNATURE*
WRITTEN BY	Olaf Peters	January 17, 2023	

REVISION HISTORY			
NUMBER	DATE	DESCRIPTION	NAME

Contents

Chapter 1

UMSMailStat

1.1   UMSMailStat

UMSMailStat 1.18 written by Olaf Peters

Copyright

What is UMSMailStat?

Usage of UMSMailStat

ums.config-Variables

Examples

Revision history
    =========
Remember:
=========

UMSMailStat is MailWare, if you like and/or use it, please send me a
little mail saying "Hi Olf, I'm using UMSMailStat!" to let me know I
didn't make the effort for myself alone! Hey, I know, you have the
ability to send mail, you're using UMS! ;-) My e-mail address is

here
.

1.2   UMSMailStat Copyright

UMSMailStat is MailWare, that means if you like this program, find a bug or
have suggestions for extensions, simply send me a mail (email or standard,
doesn't matter, but please send it!), to let me know there's someone out
there using my program!

1.3 About UMSMailStat

What is UMSMailStat?

UMSMailStat is a tool for creating hitlists about authors and subjects of a
newsgroup found in an UMS-messagebase. UMSMailStat has its roots in
'TopTen' by Kai Bolay and Jan Geissler, expanding functionality in many
points.

UMSMailStat is/has been written in Modula-2 (M2Amiga v4.3), the sourcecode
is available on request.

I can be reached via the following addresses:

e-mail: op@hb2.maus.de   (preferred, but only mails <16KB!)
        olf@informatik.uni-bremen.de

Snail: Olaf Peters
        Kulmer Str. 7
        D-28237 Bremen
        +49-421-6163352

1.4 Usage of UMSMailStat

            UMSMailStat can only be started from CLI and has the following ←
               arguments:

NAME/A,PASSWORD/A,SERVER/K,GROUP/A,DATE,DAYS/N,MU=MAXUSERS/N,
MS=MAXSUBJECTS/N, QUIET/S,RQ=REALBYTES/S,CD=CDATE/S,QCHK=QUOTECHECK/S,
QCHRS=QUOTECHARS,NS=NOSUBJECTS/S,NU=NOUSERS/S

If you want statistics to be in german or french language, you should copy
the supplied 'UMSMailStat.catalog' from the Locale/Catalogs/<language>
directory of the distribution archive to your corresponding
Locale:Catalogs/<language> directory (provided that you have installed
WorkBench v2.1 or higher).

USER/A

    the name (or alias) of a user in the UMS messagebase.

PASSWORD/A

    his/her password

SERVER/K

    the server, on which the user wants to login (default is "default" :-),
    the server keyword must be given, if you want to login on another
    server (e.g. "SERVER=other")

GROUP/A

    the group in your UMS messagebase, which should be scanned for the
    statistics. Instead of a group you can also give an AmigaDOS-pattern
    and all groups matching this pattern will be taken into the stats.

DATE,DAYS/N

    these two arguments determine the period about which the statictics
    will be built: DATE means the stats are starting with this day, DAYS is
    for the length of the period in days. If DATE is stated sole, stats
    will be built just for this one day. If you state DATE and DAYS, stats
    will be built for all messages that were written in the DAYS days after
    DATE (including DATE). If you just state DAYS, only messages of the
    past DAYS days, including today will be scanned. If you neither give
    DAYS nor DATE, all messages of that group will be scanned.

CD=CDATE/S

    this option acts the same way as DATE does, the only exception is that
    the messages are not selected by their import- but their creation-date,
    i.e. the date the message has really been written by the author. This
    option will only take effect with an UMSServer v11.22+, with v10 it
    falls back to DATE.

MU=MAXUSERS/N

    determines how many users will be in the list. If you state 10, you'll
    get a top-10 and so on. 0 means all users will appear in the list.

    This option will override the settings of the
                UMS-variable
                    "UMSMailStat.MaxUsers".

MS=MAXSUBJECTS/N

    same as MAXUSERS, but for the subject-list. 0 also means all subjects
    will appear.

    This option will override the settings of the
                UMS-variable
                    "UMSMailStat.MaxSubjects".

QUIET/S

    UMSMailStat will just show the statistics, but not copyright,
    progress-display etc.

RB=REALBYTES/S

    if this option is used, UMSMailStat determines the real length of the

messagetext instead of using the faster headerfield TextLen resulting
in more accurate byte result but also slows down scanning.

QCHK=QUOTECHECK/S

turns on the quote-analyzer to generate statictics about "quoting
habits". UMSMailStat will try to find out how many bytes of a message
have been quoted from another message and show these values in a
seperate column of the statistics. The quotecheck can be configured by
the QUOTECHARS commandline-option (see below) and the
                 UMS-variables
                     "UMSMailStat.Quote.Chars" and "UMSMailStat.Quote. ←
                         CheckFirstNChars".
See also
                 this example
                  for further inspiration :-)

Note: when using QUOTECHECK the REALBYTES option will automagically be
used. QUOTECHECK also slows down scanning speed.

QCHRS=QUOTECHARS

configures which characters will be considered as quote-introducers in
the messagebody. If one of the characters in this string is found in
the first 10 characters (see note below) of a line, that line will be
considered as a quoted line. By default, UMSMailStat just uses ">" (as
this appears to be the standard quote-introducer).

This option will override the settings of the
                 UMS-variable
                     "UMSMailStat.Quote.Chars".

Note: by using the
                 UMS-variable
                  "UMSMailStat.Quote.CheckFirstNChars"
you can configure how many chars at the beginning of a line shall be
checked for being quote-chars.

NS=NOSUBJECTS/S
NU=NOUSERS/S

if using one of these options, the subject-statictics (NOSUBJECTS) or
the author-statictics (NOUSERS) will not be printed. It is also
possible to use both option then only the global statistics will be
shown.

After calling UMSMailStat the statictics will be built, these can be
redirected into a file or i.e. to the printer via the AmigaDOS redirection
option. Just the statistics will be redirected, but not the copyrightnote,
progress-display etc., these will always be written into the
console-window. See also
                 Examples
                 .

1.5 ums.config-Variablen

Since version 1.14 UMSMailStat some variables in the ums.config have been
introduced:

    ·~UMSMailStat.MaxUsers (since v1.14)
    ·~UMSMailStat.MaxSubjects (since v1.14)
    ·~UMSMailStat.Quote.Chars (since v1.16)
    · UMSMailStat.Quote.CheckFirstNChars (since v1.16)

These options can be used to override some of UMSMailStat's hardcoded
limitations for statistics-generation, but will themselves be overridden by
their pendants on the commandline.

UMSMailStat.MaxUsers (CLI pendant: MAXUSERS/N)
UMSMailStat.MaxSubjects (CLI pendant: MAXSUBJECTS/N)

    I.e. if you insert the following lines into your ums.config (v11 notation)

        (UMSMailStat.MaxUsers "20")
        (UMSMailStat.MaxSubjects "0")

    UMSMailStat would now list 20 users in the users-statistics and all
    subjects in the subjectstatistics by default.

UMSMailStat.Quote.Chars (CLI pendant: QUOTECHARS)

    I.e. if you insert the following line into your ums.config (v11
    notation):

        (UMSMailStat.Quote.Chars ">%:")

    UMSMailStat would now consider ">", "%" and ":" as quote-introducers by
    default. This option will only be used if the QUOTECHECK option has
    been specified in the commandline.

UMSMailStat.Quote.CheckFirstNChars (no CLI pendant)

    When the QUOTECHECK-option is used, UMSMailStat by default checks the
    first 10 characters of each line of a message to see if it contains a
    quote-char (see above). If you think this is too much/low, you can use
    this to change the margin to your personal needs. I.e. the following
    line in your ums.config (v11 notation)

        (UMSMailStat.Quote.CheckFirstNChars "20")

    would let UMSMailStat check the first 20 characters of every line.

    Note: A large margin may (I haven't tested it) result in a further
    slow-down of scanning speed.

1.6 Examples

```
                Some example-calls:
~~~~~~~~~~~~~~~~~~~~
```

  UMSMailStat olf secret GROUP=de.comm.software.ums >topten.asc

Build statistics about the group de.comm.software.ums and write them into
file topten.asc.

  UMSMailStat olf secret GROUP=de.comm.software.ums DATE=23-Nov-93

Build statistics about the 23rd of november 1993 in de.comm.software.ums.

  UMSMailStat olf secret GROUP=de.comm.software.ums DATE=23-Nov-93 DAYS=7

Build statistics about the week starting with the 23rd of november 1993 and
ending on 29th of november.

  UMSMailStat olf secret GROUP=de.comm.software.ums DAYS=7

Builds statistics of the last week, including today as well.

  UMSMailStat olf secret GROUP=de.comm.software.ums QUOTECHECK QUOTECHARS=">|:"

Build statistics about the group de.comm.softeware.ums and also checks the
quoting habits, using ">", "|" and ":" as quote-introducers.

```
Klick
                here
                 for an example of the standard-statistics and
                here
                 for the
quote-statictics.
```


1.7 Standard example

```
--------------------------------------------------
Mailstatistics for group comp.sys.amiga.programmer
--------------------------------------------------

Statistics from 16.04.1995 to 18.04.1995.

Bytes:          166730

Authors            67   Bytes/Author:        2488
Messages:         105   Bytes/Message:       1587
Subjects:          65   Bytes/Subject:       2565
```

The most bytes in one message, being 5197, wrote
Doug Walker about 'Re: Explain .library'.

Top 3 Authors
=============

```
------------------------------+------------+-------------+-------------+
  Pl. Name                     | Messages   |    Bytes    | Bytes/Msg.  |
------------------------------+------------+-------------+-------------+
  1. Doug Walker               |    6,  5%  | 13483,  8%  | 2247    660 |
  2. Jan Krutisch              |    4,  3%  |  9430,  5%  | 2357    770 |
  3. James Cooper              |    4,  3%  |  6628,  3%  | 1657     70 |
------------------------------+------------+-------------+-------------+
  E  Sum/Average               |   14, 13%  | 29541, 17%  | 2110    523 |
------------------------------+------------+-------------+-------------+
```

Top 3 Subjects
==============

```
------------------------------+------------+-------------+-------------+
  Pl. Subject                  | Messages   |    Bytes    | Bytes/Msg.  |
------------------------------+------------+-------------+-------------+
  1. Zoom gadgets              |    5,  4%  |  7939,  4%  | 1587      0 |
  2. Interrupts                |    4,  3%  |  7454,  4%  | 1863    276 |
  3. A brand new OS            |    4,  3%  |  6406,  3%  | 1601     14 |
------------------------------+------------+-------------+-------------+
  E  Sum/Average               |   13, 12%  | 21799, 13%  | 1676     89 |
------------------------------+------------+-------------+-------------+
```

                              UMSMailStat 1.16 (18.4.95) by Olaf Peters

```
-----8<----------8<----------8<----------8<----------8<----------8<---------
```

What do the fields mean?

```
                                 1)  2)          3)  4)       5)        6)

  2. Interrupts                |    4,  3%  |  7454,  4%  | 1863    276 |
```

1) Number of messages about this subject (or this author)
2) percentage of all messages
3) Amount of bytes about this subject (or this author)
4) percentage of all messages
5) average bytes per message
6) difference between 5) and the overall average of bytes per message

The E-row at the end of the table states the sum/averages for the chosen
number of table-entries. It doesn't have to correspond with the global
values!

1.8 Quote example

```
--------------------------------------------------
```
Mailstatistics for group comp.sys.amiga.programmer

```
------------------------------------------------

Statistics from 16.04.1995 to 18.04.1995.

Bytes:           115281 (30% quoted)

Authors             67   Bytes/Author:          1720
Messages:          105   Bytes/Message:         1097
Subjects:           65   Bytes/Subject:         1773

The most bytes in one message, being 4739, wrote
Doug Walker about 'Re: Explain .library'.
```

Top 3 Authors
=============

```
-------------------------+----------+------------+------------+------+
 Pl. Name                | Messages |   Bytes    | Quotebytes | B/Msg |
-------------------------+----------+------------+------------+------+
  1. Doug Walker         |   6,  5% | 10600,  9% |  3646, 34% | 1766 |
  2. Jan Krutisch        |   4,  3% |  7346,  6% |  3176, 43% | 1836 |
  3. James Cooper        |   4,  3% |  4658,  4% |  2628, 56% | 1164 |
-------------------------+----------+------------+------------+------+
  E  Sum/Average         |  14, 13% | 22604, 19% |  9450, 41% | 1614 |
-------------------------+----------+------------+------------+------+
```

Top 3 Subjects
==============

```
-------------------------+----------+------------+------------+------+
 Pl. Subject             | Messages |   Bytes    | Quotebytes | B/Msg |
-------------------------+----------+------------+------------+------+
  1. Zoom gadgets        |   5,  4% |  5464,  4% |  2472, 45% | 1092 |
  2. Interrupts          |   4,  3% |  5415,  4% |  1622, 29% | 1353 |
  3. A brand new OS      |   4,  3% |  4331,  3% |   916, 21% | 1082 |
-------------------------+----------+------------+------------+------+
  E  Sum/Average         |  13, 12% | 15210, 13% |  5010, 32% | 1170 |
-------------------------+----------+------------+------------+------+
```

```
                        UMSMailStat 1.16 (18.4.95) by Olaf Peters

-----8<----------8<----------8<----------8<----------8<----------8<---------

What do the fields mean?

                          1)  2)        3)  4)        5)  6)        7)

  2. Interrupts          |   4,  3% |  5415,  4% |  1622, 29% | 1353 |


1) Number of messages about this subject (or this author)
2) percentage of all messages
3) Amount of bytes about this subject (or this author)
4) percentage of all messages
5) number of bytes from 3) that have been considered as quotes in the
   messages.
```

```
6) percentage of quotes relative to 3)
7) average bytes per message
```

The E-row at the end of the table states the sum/averages for the chosen
number of table-entries. It doesn't have to correspond with the global
values!

1.9 Revision history

```
v1.18 (16.8.95)
```

```
- NOSUBJECTS and NOUSERS options (as requested by Luc Gibert)
- french documentation included (many thanks to Luc Gibert!)
```

```
v1.17 (13.6.95)
```

```
- fixed a nasty bug that has been lurking in REALBYTES since v1.09.
  UMSMailStat had problems with mails larger than 32767 bytes, now we're
  able to handle 2 GB messages (if you got the mem for it :-D ). Thanks to
  Luc Gibert for finding this bug!
```

```
v1.16 (18.4.95)
```

```
- french catalog included. Big THANK YOU to Luc Gibert! :-)
- new options QUOTECHECK and QUOTECHARS and variables
  "UMSMailStat.Quote.Chars" and "UMSMailStat.Quote.CheckFirstNChars" for
  generation of statictics about the "quoting habits" of the authors (as
  requested by Luc Gibert).
- changed the way the progressmeter is printed. (as requested by Ralf
  Garbade)
- revised the documentation.
```

```
v1.15 (28.3.95)
```

```
- instead of printing 0% in the table UMSMailStat now prints the
  decimal-point and the first digit of the fraction-part: .5% instead of 0%
  (as requested by Luc Gibert)
- CLI-Usage now completely conforming the login-guide.
```

```
v1.14 (14.3.95)
```

```
- the defaults for MaxUsers and MaxSubject are now read from the
  ums.config-variables UMSMailStat.MaxUsers and UMSMailStat.MaxSubjects.
  Only if these do not exist, UMSMailStat will fall back to its hardcoded
  values.
```

```
v1.13 (25.2.95)
```

```
- new CDATE-Option. Using this the messages will be selected by their
  creation-date instead of their import-date. This option is only available
  with an UMSServer v11.22+, with v10 this option has no effect.
```

```
v1.12 (23.02.95)
```

```
- bumped revision number: the last public release (v1.11) contained the
```

  1.09 binary :-((Thanks to Matthias Scheler for the hint)

v1.11 (13.02.95)

- if two or more authors/subjects have the same ranking (see v1.10) they
  will be printed in alphabetical order.
- break-handling removed, could lead into bad crashes :-(
- some layoutchanges to the documentation

v1.10 (11.02.95)

- if two or more authors/subjects have the same amount of messages *and*
  bytes they will now have the same ranking.
- RB-shortcut for REALBYTES

v1.09 (15.1.95)

- cursor will now be switched off after the commandline has been parsed
  (requested by Christoph Viethen)
- some UMS-called have not been checked for errors.
- REALBYTES/S option to get more accurate byte results.

v1.08 (07.11.94)

- changed directory-structure of the distribution to fit Stefan Becker's
  guidelines
- revised the documentation
- GROUP "" didn't state that the statistics were about the personal mail
- changed cursorswitching a bit

v1.07 (03.11.94)

- memorymanagement now using memorypools, reducing fragmentation of
  mainmemory.
- lokalization with OS 2.1+, so the keyword ENGLISH is not needed anymore.

v1.06 (09-Apr-94)

- MS can now be used for MAXSUBJECTS, MU for MAXUSERS.
- some changes in the docs.

v1.05 (20-Jan-94)

- discovered a very nasty bug, occuring with entries (subject or name)
  >= 25 chars, making the whole statistics unreliable. Fixed!
- scanning-speed slightly increased.

v1.04 (13-Jan-94)

- changed the period-calculation again. DAYS now really states who many
  days are considered in the stats, so DAYS=7 really means "the last week"
  or "the week after (and including) DATE. What did I consider the other
  way to calculate the period to be more logical?!? :-)

v1.03 (11-Jan-94)

- Happy New Year :-)

- removed a bug that appeared when specifying DATE & DAYS: the end of
  that period was at time 00:00:00 and not at time 23:59:59, in fact a
  whole day of the period was missing (as if you would have specified
  DAYS-1)

v1.02 (19-Dez-93)

- GROUP can now also be a pattern: all groups matching this pattern are
  taken into the stats.

v1.01 (09-Dec-93)

- little "bugs" in the docs removed... ;-)

v1.00 (08-Dec-93)

- first public release