Threading and NOV

266379_PixelRule.tiff ¬

Q: *What is `threading' and why is it good for you?*

A: I put this in first as a background question.

As news volume has grown in the last couple of years, the organization of the data presented to the user has become more important.   With the possibility of being presented with several hundred articles to read in a newsgroup, newer newsreaders have incorporated features that organize things by subject, and sort them in the order posted, and *thread* articles in groups according to who replied to who.   Note there is a difference between subject sorting and date sorting, and threading.

This subject sorting and threading can be expensive using normal NNTP commands, as it involves opening each article and transferring it over to the client, for a relatively small bit of information.   This led to newsreader authors extending the NNTP command set so that they could set up an auxiliary database of just the information they needed.   The database would be updated as news arrived on the server, and when a client requested thread information about a particular group, it would just open one file and send it out, which made things pretty zippy.

The downside of this, was that every newsreader had their own threading database, their own specialized NNTP extension, and it was customized to their *way* of doing threads.   Trn, NN, and tin are prime examples.   Each database occupies a fair amount of space, and in the case of trn and tin, the daemons that maintain the databases use up a nontrivial amount of system resources.

173805_PixelDottedRulePadded.tiff ¬

Q: *What is the Overview/NOV database?*

A: The Overview or News OverView (NOV) database was designed by Geoff Collyer.   Its a more generalized database, it it contains **no** threading information, just the information needed to thread.   The overview database is just a database of files, one for each group, containing a reference to each article, with seven or eight of the most popular headers.   Thus, the newsreader client can get most of the information it needs to thread, but it does the threading the way it wants to.   Yes, the downside is that now the threading occurs on the client's CPU, but that's not as expensive as it may seem, if done properly.

462132_PixelDottedRulePadded.tiff ¬

Q: *What is the format of the NOV database?*

A: (from Geoff Collyer's newsoverview(5) man page) Each newsgroup directory contains a file named   `.overview', containing one-line   summaries   of   articles in that group.   Fields are separated by tabs, and any tabs   or   newlines   in the   original articles   headers   have   been   replaced   with spaces.   The fields are, in   order: article   number   (file name),   subject,   author, date,
message-id, references, byte   count, line count, and optionally other headers, as arranged   locally   (none   are   supplied   by   the database maintenance software, as shipped).   The line-count and references field may   be   empty.   If the optional other headers are present, they include their header keyword and   colon;   if   they are absent entirely, the tab after the line-count field may also be absent.

ie:

```
artid|subject|From|Date|Message-Id|References|\
bytecount|linecount|Optional-Header: Stuff
```

where | represents a tab and \ a line continuation (there are no continuation lines in the Overview entry, it's just so that the above line fits in 80 columns).

Remember that header keywords in optional headers should be matched case insensitively.

683963_PixelDottedRulePadded.tiff ¬

Q: *What optional headers should I configure into the NOV database?*

A: It's nice to configure the pseudo-header Xref: in.   Xref: is a site specific header that, tells if present where that are article has been cross posted.   It is useful for newsreaders who want to kill cross posted articles, you've told them to kill.

172600_PixelDottedRulePadded.tiff ¬

Q: *What is XOVER?*

A: XOVER is an NNTP extension for accessing the Overview database.   It is becoming a defacto standard for accessing bulk header/thread information via NNTP.

It is used after establishing a current group with the GROUP command, and operations on that group.   It's syntax is

```
XOVER artnumber1[-artnumber2]
```

Where *artnumber1* is < *artnumber2*.   If successful, it returns a 224 reply code, and the overview data ± as in the Overview database€± for the requested articles.   It is terminated by a `.' (dot) on a line by itself.

NNTP servers implementing the XOVER command should try to ensure that the data is as consistant as possible, they should avoid handing out XOVER information for articles that don't exist, and should generate XOVER information on the fly for articles that do exist, but aren't found in the Overview database for some reason. While the job of XOVER is to provide access to the Overview database, validating the Overview data is best done in the NNTP server, rather than having the NNTP client go through contortions to do it.

90168_PixelDottedRulePadded.tiff ¬

Q: *How do I know if my NNTP server does XOVER?*

A: By initially trying it, and looking at the response code.   Before doing a GROUP in your session, you should get back a 500 (no such command) response if the server doesn't implement XOVER, and a 412 (not in a newsgroup) response, if it does.

490468_PixelDottedRulePadded.tiff ¬

Q: *What does it buy me?*

A: Speed, if you have newsreader clients that can take advantage of this, disc space if you are currently supporting more than one version of a threading database.

If used by the NNTP daemon, XHDR can be speed up by **orders** of magnitude by using the Overview database.

499629_PixelDottedRulePadded.tiff ¬

Q: *How much disc space does the Overview database take up?*

A: About 10% of what your news spool space is.

571772_PixelDottedRulePadded.tiff ¬

Q: *What News Transport Systems support NOV/XOVER and where can I get it?*

A: INN 1.4 supports the Overview database and XOVER right out of the box.

The Performance release of C News (October 18th, 1994) supports the Overview database.   To use the XOVER extension with C News and NNTP,   you need Stan Barber's NNTP reference package   (via ftp via ftp.uu.net in /networking/news/nntp/nntp.1.5.12.1.tar.Z).

874284_PixelDottedRulePadded.tiff ¬

Q: *What Newsreaders Support NOV/XOVER?*

A: Among *officially released* newsreaders trn3.6 and tin-1.22 support both the Overview database and XOVER.   TaskArrowSm.tiff ¬ Win-vn 0.8 for MS-Windows claims to support XOVER [Alan Thew]. 833876_TaskArrowSm.tiff ¬   RNR v1.25 and up for MS-DOS support XOVER [Russell Schulz]. 854953_TaskArrowSm.tiff ¬ Geoff has modified *proof of concept* newsreaders in ftp.std.com:/src/news/ that support the Overview database (**not** XOVER). 952810_TaskArrowSm.tiff ¬ Version 6.5.0 of NN (which is in late beta) is fully NOV capable.   It is available from ftp://ftp.uwa.edu.au/pub/nn/beta/nn-6.5.0.b3.tar.gz. 872404_TaskArrowSm.tiff ¬ There are some excellent patches by Felix Lee that allow GNUS to support XOVER (plus some kool async transfer mods).   Perhaps he'll post them again soon.   Perhaps they will be integrated into GNUS.

(ALEXANDRA fully supports a NOV database. It uses the XHDR command if available and falls back to a configurable set of XOVER commands otherwise ± *erik*)

1PixelDottedRulePadded.tiff ¬

Q: *Is there a reference threading implementation for NOV data?*

A: Yes, in Geoff Collyer's package at
ftp.std.com:/src/news/nov.dist.tar.gz.

756601_PixelDottedRulePadded.tiff ¬

Q: *Any tips on configuring the Overview database?*

A: It's probably a good idea to configure in the optional pseudo
header Xref:, as many newsreaders like to see it.

Make sure you put the `.overview' files somewhere besides where
you put the articles (normally /usr/spool/news).   If, for example,
`/usr/spool/news/news/answers' has 10,000 articles plus the
`.overchan' file, then the open() call may need to search through up
to 10,001 filenames   to find the `.overview' file.   If, instead, you put
the overview database info `/usr/spool/news/over.view' then the
`.overview' file could be found almost instantly.   Part 3 of the INN
FAQ has tips on making this change for INN.

If newsreading is totally limited to NNTP, you might want to consider
putting your Overview database on another partition on another disc
(if you have space).   This buys you more room for news on your
spool disc, and also spreads the disc accesses out across multiple
drives, reducing the beating your spool partition is going to take.

460723_PixelDottedRulePadded.tiff ¬

Thanks to the following for contributions, additions, corrections, and
updates:

    Jerry Aguirre <jerry@strobe.ATC.Olivetti.Com>

Geoff Collyer <geoff@world.std.com>
Wayne Davison <davison@borland.com>
Christopher Davis <ckd@eff.org>
Jason Haar <j.haar@cantva.canterbury.ac.nz>
Damon Hart-Davis <d@exnet.co.uk>
Andrei Ivanov <iva@lynx.riga.lv>
Iain Lea <iain@anl433.erlm.siemens.de>
Tom Limoncelli <tal@plts.org>
Mark Linimon <linimon@nominil.lonesome.com>
Doug Sewell <doug@cc.ysu.edu>
Russell Schulz <russell@alpha3.ersys.edmonton.ab.ca>
Alan Thew <News.Admin@liverpool.ac.uk>

This posting, like much of Usenet, is maintained on a purely volunteer basis.   I welcome reactions, additions, and corrections via email at rob@agate.berkeley.edu.

374095_PixelRule.tiff ¬