

Ruby master - Misc #15806

Explicitly initialise encodings on init to remove branches on encoding lookup

04/27/2019 11:41 PM - methodmissing (Lourens Naudé)

Status:	Assigned
Priority:	Normal
Assignee:	nobu (Nobuyoshi Nakada)

Description

References Github PR <https://github.com/ruby/ruby/pull/2128>

I noticed that the encoding table is loaded on startup of even just miniruby (minimal viable interpreter use case) through this backtrace during ruby setup:

```
/home/lourens/src/ruby/ruby/miniruby(rb_enc_init+0x12) [0x56197b0c0c72] encoding.c:587
/home/lourens/src/ruby/ruby/miniruby(rb_usascii_encoding+0x1a) [0x56197b0c948a] encoding.c:1357
/home/lourens/src/ruby/ruby/miniruby(Init_sym+0x7a) [0x56197b24810a] symbol.c:42
/home/lourens/src/ruby/ruby/miniruby(rb_call_inits+0x1d) [0x56197b11afed] inits.c:25
/home/lourens/src/ruby/ruby/miniruby(ruby_setup+0xf6) [0x56197b0ec9d6] eval.c:74
/home/lourens/src/ruby/ruby/miniruby(ruby_init+0x9) [0x56197b0eca39] eval.c:91
/home/lourens/src/ruby/ruby/miniruby(main+0x5a) [0x56197b051a2a] ./main.c:41
```

Therefore I think it makes sense to instead initialize encodings explicitly just prior to symbol init, which is the first entry point into the interpreter loading that currently triggers `rb_enc_init` and remove the initialization check branches from the various lookup methods.

Some of the branches collapsed, cachegrind output, columns are `Ir Bc Bcm Bi Bim` with `Ir` (instructions retired), `Bc` (branches taken) and `Bcm` (branches missed) relevant here as there are no indirect branches (function pointers etc.):

(hot function, many instructions retired and branches taken and missed)

```
      .      .      .      .      .      rb_encoding *
      .      .      .      .      .      rb_enc_from_index(int index)
835,669      0      0      0      0      {
13,133,536  6,337,652  50,267      0      0      if (!enc_table.list) {
      3      0      0      0      0      rb_enc_init();
      .      .      .      .      .      }
23,499,349  8,006,202  293,161      0      0      if (index < 0 || enc_table.count <= (index &
= ENC_INDEX_MASK)) {
      .      .      .      .      .      return 0;
      .      .      .      .      .      }
30,024,494      0      0      0      0      return enc_table.list[index].enc;
1,671,338      0      0      0      0      }
```

(cold function, representative of the utf8 variant more or less too)

```
      .      .      .      .      .      rb_encoding *
      .      .      .      .      .      rb_ascii8bit_encoding(void)
      .      .      .      .      .      {
27,702      9,235      955      0      0      if (!enc_table.list) {
      .      .      .      .      .      rb_enc_init();
      .      .      .      .      .      }
9,238      0      0      0      0      return enc_table.list[ENCINDEX_ASCII].enc;
9,232      0      0      0      0      }
```

I think lazy loading encodings and populating the table is fine, but initializing it can be done more explicitly in the boot process.

Associated revisions

Revision 6546aed4 - 07/23/2019 07:45 AM - methodmissing (Lourens Naudé)

Explicitly initialise encodings on init to remove branches on encoding lookup

[Misc #15806]

Closes: <https://github.com/ruby/ruby/pull/2128>

Revision 009ec37a - 07/23/2019 07:45 AM - methodmissing (Lourens Naudé)

Let the index boundary check in `rb_enc_from_index` be flagged as unlikely

[Misc #15806]

Closes: <https://github.com/ruby/ruby/pull/2128>

History

#1 - 07/23/2019 07:58 AM - methodmissing (Lourens Naudé)

- *Status changed from Open to Closed*

Applied in changeset <git6546aed4757be07f4932326e1eb41a5d69141acf>.

Explicitly initialise encodings on init to remove branches on encoding lookup

[Misc #15806]

Closes: <https://github.com/ruby/ruby/pull/2128>

#2 - 08/29/2019 04:29 AM - mame (Yusuke Endoh)

- *Assignee set to nobu (Nobuyoshi Nakada)*

- *Status changed from Closed to Assigned*

usa said it might not work on windows when the install path includes non-ASCII characters. Please check it out, [nobu \(Nobuyoshi Nakada\)](#) .