

- ⇒ [MetaMap Home](#)
- ⇒ [FAQ Home](#)
- ⇒ [Help/Questions](#)

Where do I find a UMLS vocabulary's code fields?

MetaMap does not provide UMLS vocabularies' source-specific codes; however, that information is readily available in the 14th field of the UMLS MRCONSO.RRF file.

For example, running MetaMap with [Fielded MMI Output](#) on the input text *rash* generates this output:

```
00000000|MMI|11.49|Exanthema|C0015230|[sosy]|["RASH"-tx-1-"rash"-noun-0]|TX|0:4|C17.800.257
```

How then do you determine the vocabulary-specific code field corresponding to that CUI (C0015230)? Suppose you want the SNOMEDCT codes for C0015230. A simple shell command is all you need:

```
<flang@indlx6> 6 : look C0015230 MRCONSO.RRF | \
    awk -F'|' '$12 == "SNOMEDCT_US" {printf "%s|%s|%s\n", $12, $14, $15}'
SNOMEDCT_US|112625008|Exanthema
SNOMEDCT_US|112625008|Exanthemata
SNOMEDCT_US|271807003|Eruption
SNOMEDCT_US|1806006|Eruption, NOS
SNOMEDCT_US|112625008|Exanthem
SNOMEDCT_US|271807003|Exanthem
SNOMEDCT_US|139684003|Rash
SNOMEDCT_US|267183006|Rash
SNOMEDCT_US|112625008|Rash
SNOMEDCT_US|271807003|Rash
SNOMEDCT_US|112625008|Rash, NOS

. . . output truncated . . .
```

The rows in the MRCONSO.RRF file are sorted by pipe-separated fields, so the `look` command retrieves all and only those rows beginning with C0015230, and the `awk` command restricts the output to those rows whose 12th field is `SNOMED_CT`, and extracts the 12th (SAB; source vocabulary), 14th (CODE), and 15th (STR; string) fields.

To restrict the results to English-only rows, simply add `$2 == "ENG" &&` before `$12 == "SNOMEDCT_US"` in the `awk` call.

The mapping from a UMLS CUI to a UMLS source is usually one to many, and often one to **very** many. E.g., RXNORM has 753 distinct codes for C0874083 and 583 for C2949404.