Feedback from users:

Assigning UCD1+ to VIZIER

APM - *INAF*
FO - *CDS*

VIZIER:
status on 11May2006

- 5,356 catalogues
- 12,983 tables
- 172,781 columns
- In-flux: about 1,000 columns/month

VIZIER

- method for the past:
  - UCD1(old)-UCD1+(new) translation (file)
- method for the present/future:
  - Assign both, cross-check
- plus:
  - Safer approach, simplicity
- minus:
  - Loss of the semantic richness of ucd1+

translated vs. assigned:

ucd1+	# if translated	# if assigned
stat.error	2	194
meta.code.error	1	145
meta.note	3	59
meta.number	2	21
phot.color	53 [171]	212
phot.count	7 [12]	46
phot.flux	43 [98]	242
phot.mag	47 [191]	313 [139]
pos.eq	14	121
spect.line.eqWidth	1	48
time.epoch	1	48
time.expo	2	18

feedback

- feedback on the use of web tools (log)
  - The log is inspected on a regular basis to check for failures or queer behaviors
- A standard list of description is assigned after s/w upgrades
- People integrating tables in VIZIER log failures

assign1p

assign1p assign ucd1p-words from list of [key]words
and build UCD1+ from word(s).
If present, a suggested old UCD1 is default-translated.

USE: assign1p [options] [<] **input-file.tsv**

tab-separated input-file fields:
    0: Cat/Tab   (Table name)
    1: Data Type (I,F,A)
    2: Col_Name (Title of the column)
    3: Col_units (no units= ---)
    4: Col_description (Free text)
    5: UCD1/ucd1+ || notes

Options:
    -h[elp] : this help
    -d : print revised description used by FindWord after the application of
        syntax/semantic rules (def=print input descr.)
    -l : list all words with matching score > 5 (def=only top P/QECV/S scores)
    -k : print elements of association-tables containing input
        description [key]words (def=no)
    -r : do not apply syntax/semantic rules to list of keywords (def=yes)
    -s : print some statistics at the end (def=no)
    -t : find also the traditional UCD.
        Generates a duplicate output line with old UCD (def=no)
    -u : do not force use of suggested ucd (def=yes)
    -v : verbose output, sets also -l -u (def=no)
    -n : prefix output with record-number (def=no)
  -nn : start at record 'nn' in input-file (def=0)

Output:

tab-separated output fields:
   [nn]0, 1, 2, ucd, 3, 4

The result of the assign1p procedure is flagged in field 1, as
   (Type,flag), where flag can be:
   nn = best score shown (if only description is used),
   _ = ColumnName used,
   % = Units used,
   _% = CN+U used,
   f = used suggested ucd1+,
   f1 = used suggested ucd1,
   ? = unable to assign,
   ! = forced note.

assignator

- **assign1p** exploits the information in:
  - Description
  - Units (not fully implemented!)
  - Column name
  - Notes
- to generate a working description to which **semantic rules** (>200) are applied
- finds ucd-words best **matching** the std list
- builds ucd1+ (~100 **syntactic rules**)

assignator

- **plus:**
  - speed: all vizier in < 8' (173,000 assignations)
  - efficiency: assigns in 99,3% of cases
  - "magics":

  Count rate in the **hard** ROSAT window 0.1-2.4keV

  $\Rightarrow$*arith.rate;phot.count;em.X-ray.soft*

  uncertainty flag Y-Z

  $\Rightarrow$*meta.code.error;phot.color;em.opt.B;em.opt.V*

assignator

- **minus:**
  - speed: all vizier in < 8' (but weeks to check!!)
  - efficiency: real rate <80%
  - "magics" (?):

    $RVel \Rightarrow spect.dopplerVeloc$

    $RV \Rightarrow ??$

  code written by amateur
  incomplete set of rules
  units not fully exploited
  [dimensions] not yet used