EURO VO

Workshop

Garching, June 27 – July 1 2005

Metadata / UCD

Sebastien Derriere
CDS
derriere@astro.u-strasbg.fr

Outline

- Why do we need UCDs?
- What are UCDs?
- Where are they used in the VO?
- Who's in charge of curation?
- How to use them in applications?
- What tools are available?

Why UCDs?

- UCDs = Unified Content Descriptors
- UCDs are metadata attached to data
- They give a semantic description of astronomical values = what a value is, with a standardized vocabulary
- UCDs needed for
  - Homogenizing large collections of heterogeneous data
  - Interoperability for data exchange

Heterogeneous data (1)

214.386166	-57.767818	16.926	15.777	99.999	0.09	0.19	9.99	17.067	15.508	99.999
214.535889	-57.767764	16.458	15.562	99.999	0.07	0.17	9.99	16.496	15.457	99.999
214.401036	-57.767685	14.974	14.391	99.999	0.04	0.11	9.99	15.021	14.549	99.999
214.569711	-57.767623	17.971	15.777	99.999	0.18	0.19	9.99	17.394	15.553	99.999
214.349915	-57.767576	16.975	99.999	99.999	0.10	9.99	9.99	16.840	99.999	99.999
214.550993	-57.767487	16.801	15.716	99.999	0.09	0.18	9.99	16.605	15.682	99.999
214.557370	-57.767406	99.999	16.525	13.594	9.99	0.27	0.22	99.999	15.544	12.905
214.404212	-57.767370	15.848	14.973	99.999	0.05	0.13	9.99	15.654	15.197	99.999
214.296113	-57.767262	15.161	13.266	99.999	0.04	0.08	9.99	15.055	13.271	99.999
214.238914	-57.767254	15.363	14.061	99.999	0.04	0.10	9.99	14.916	14.106	99.999
214.286765	-57.767228	15.694	13.984	99.999	0.05	0.09	9.99	15.784	14.019	99.999
214.595510	-57.767131	17.716	16.170	99.999	0.15	0.23	9.99	17.274	16.150	99.999
214.466317	-57.767040	15.975	13.680	12.353	0.06	0.09	0.12	15.998	13.686	12.836
214.503014	-57.767008	17.436	99.999	99.999	0.13	9.99	9.99	17.548	99.999	99.999
214.483010	-57.766971	99.999	16.015	99.999	9.99	0.21	9.99	99.999	16.370	99.999
214.470701	-57.766933	16.031	13.917	12.708	0.06	0.09	0.14	16.025	13.909	13.395
214.38202	-57.766657	18.085	99.999	99.999	0.19	9.99	9.99	19.044	99.999	99.999
214.26135	-57.766521	17.167	15.984	99.999	0.11	0.20	9.99	17.209	16.248	99.999
214.61179	-57.766361	17.103	15.149	99.999	0.10	0.14	9.99	16.741	14.812	99.999
214.26777	-57.766321	17.388	15.712	99.999	0.13	0.18	9.99	16.713	15.023	99.999
214.53227	-57.766314	16.179	14.313	13.036	0.06	0.11	0.16	16.099	14.255	13.102
214.56522	-57.766304	17.249	15.955	99.999	0.11	0.20	9.99	17.213	17.242	99.999
214.47422	-57.765746	18.131	99.999	99.999	0.20	9.99	9.99	18.831	99.999	99.999

RA DEC magI magJ magK

Heterogeneous data (2)

- How can we find relevant data ?
- How do we (or a software) know that some data can be compared?
- How do we automate data processing?
- ... if we don't have a standard description of data contents?
- Description of tables with column labels (names) is definitely not sufficient

Heterogeneous data (3)

- Historically, UCDs come from a work on the VizieR catalogue service at CDS
  - over 4000 different catalogues
  - coming from various origins
    - digitization of photographic plates
    - observations over the whole spectrum
    - electronic publication of tables in journals
  - original descriptions and column names vary a lot
    - over 140 different names for a V magnitude !!

First version: UCD1

- ESO-CDS data mining project
  - manual exploration of 100,000 columns
  - creation of new words for new quantities
  - ~1500 words created
    - PHOT_EXTINCTION_ISM = Interstellar extinction
    - POS_GAL_LAT = Galactic Latitude
- UCD give a semantic description of column contents
  - allows to search catalogues on their contents
  - allows automated comparisons

IVOA: UCD1+

- UCD1 were defined before VO projects
- The IVOA defines standards to allow interoperability between the different VO projects and partners: interest in UCDs
- Migration to UCD1+
  - re-use an existing knowledge base (UCD1)
  - more flexible syntax
  - vocabulary controlled by IVOA
- http://cdsweb.u-strasbg.fr/UCD/
  - list of words, tools, ...

IVOA: UCD1+

- Same goal as UCD1
  - describe quantities: "what this is"
  - with a reasonable accuracy
  - allow interoperability
- New list of words (<500) with definitions
- New syntax
  - case-insensitive words
  - possibility to combine several words to form one UCD
  - the first word carries most of the meaning

Examples of UCD1+

- pos.eq.dec;meta.main = declination
- phot.mag = magnitude
- phot.mag;em.opt.V = magnitude measured in the optical between 500 and 600nm
- stat.error;phys.temperature = measurement error on the temperature
- phot.color;em.opt.B;em.opt.V = B-V colour
- meta.id;meta.main = source main identifier
- meta.id;instr.telescope = telescope's name

UCDs in the VO (1)

- Standard described on IVOA web site
  - http://www.ivoa.net/Documents/latest/UCD.html
- Discussions: ucd@ivoa.net
- Vocabulary not cast in concrete
  - improvements to the list of words
  - long-term curation, open to new domains
- Scientific Board: vocabulary curation
  - ucd-sci@ivoa.net
- Technical board: requirements for tools
  - ucd-tech@ivoa.net

UCDs in the VO (2)

- Use in other VO standards:
  - VOTable ("ucd" attribute to <FIELD>, <PARAM> and <GROUP>)
  - Registry: description of column contents in the <column> element of VODataService schema
  - VOEvent: use in <param> and <group>
- Use in Cone Search
  - ucd= "**ID_MAIN**"
  - ucd= "**POS_EQ_RA_MAIN**"
  - ucd= "**POS_EQ_DEC_MAIN**"
- Use in applications (Aladin, TopCat)

Use in VOTable

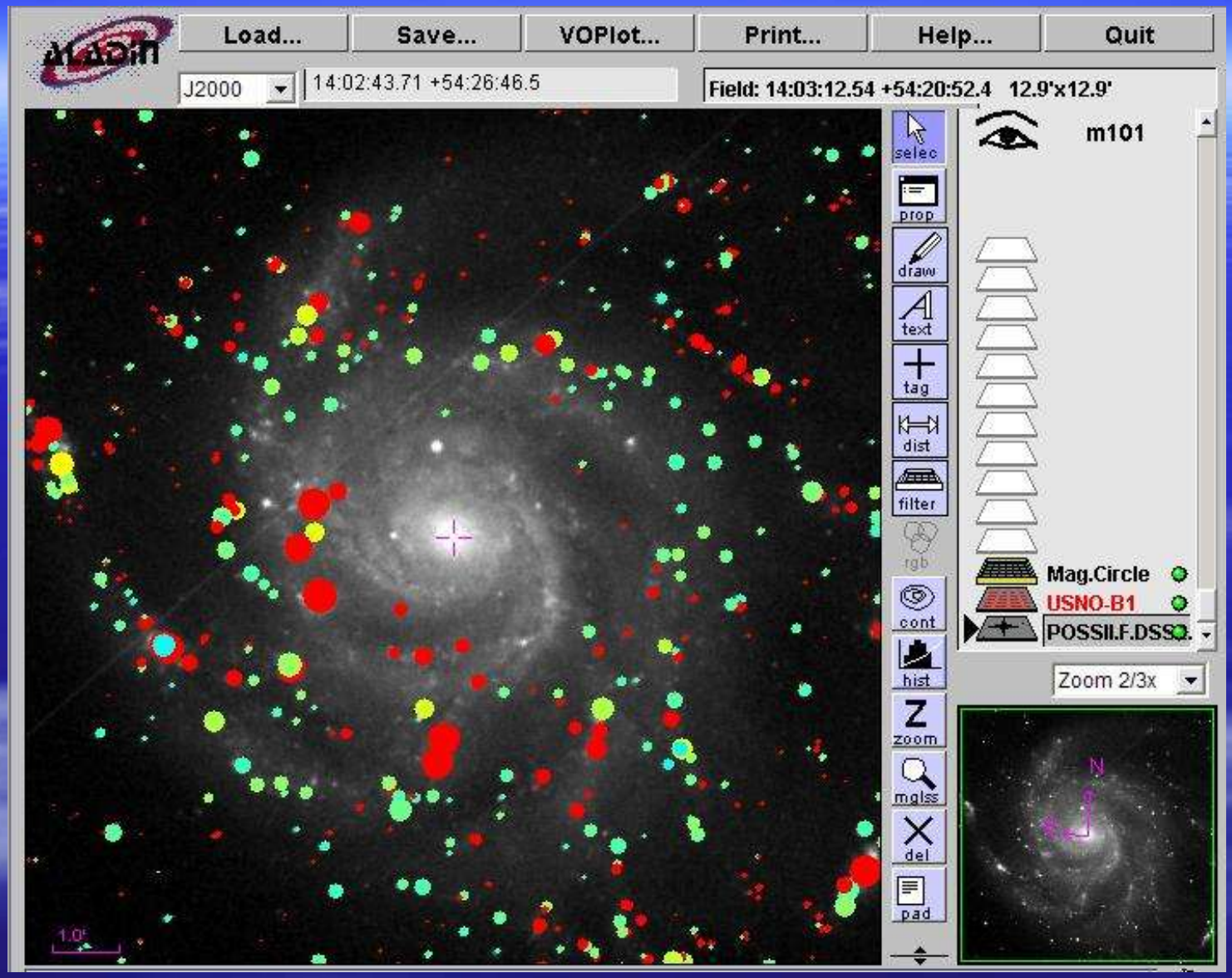
```
- <FIELD datatype="float" name="plx" precision="2" ucd="pos.parallax.trig" unit="mas" width="5">
    <DESCRIPTION>? Hipparcos <I/239> trigonometric parallax</DESCRIPTION>
  </FIELD>
- <FIELD arraysize="11*" datatype="char" name="SpType" ucd="src.spType">
    <DESCRIPTION>Spectral type and luminosity class</DESCRIPTION>
  </FIELD>
- <FIELD datatype="float" name="LDD" precision="2" ucd="phys.angSize" unit="mas" width="5">
    <DESCRIPTION>Limb-darkened disc diameter</DESCRIPTION>
  </FIELD>
- <FIELD datatype="float" name="e_LDD" precision="3" ucd="stat.error" unit="mas" width="6">
    <DESCRIPTION>Error on LDD</DESCRIPTION>
  </FIELD>
- <FIELD datatype="short" name="Teff" ucd="phys.temperature.effective" unit="K" width="4">
    <DESCRIPTION>Effective temperature</DESCRIPTION>
  </FIELD>
- <FIELD datatype="float" name="logg" precision="2" ucd="phys.gravity" unit="[cm/s2]" width="5">
    <DESCRIPTION>Surface gravity: log(g) with g in cm/s^2^</DESCRIPTION>
  </FIELD>
- <FIELD datatype="float" name="Kmag" precision="2" ucd="phot.mag;em.IR.K" unit="mag" width="5">
    <DESCRIPTION>Johnson's K magnitude</DESCRIPTION>
  </FIELD>
- <DATA>
  - <TABLEDATA>
    - <TR>
        <TD>1</TD>
        <TD>00 11 34.42</TD>
        <TD>-27 47 59.1</TD>
        <TD>720</TD>
        <TD>Simbad</TD>
        <TD>HIP</TD>
        <TD>00 11 34.4201</TD>
```

Use in the registry

- Allows to search VO resources based on their contents
  - more efficient than keyword search
- Key access method to identify relevant resources among large collections of heterogeneous datasets!
- See example in the tutorial session!

Use in applications

- When UCDs are present in a VOTable document, they allow some automated processing, e.g.:
  - identifying which column contain sky coordinates and plot the sources on an image
  - perform predefined generic operations without specific knowledge of the table structure
    - e.g. filters in Aladin


Load...
Save...
VOPlot...
Print...
Help...
Quit
J2000
14:02:43.71 +54:26:46.5
Field: 14:03:12.54 +54:20:52.4 12.9'x12.9'
m101
selec
prop
draw
text
tag
dist
filter
rgb
cont
hist
zoom
mglss
del
pad
Mag.Circle
USNO-B1
POSSII.F.DSS
Zoom 2/3x
N
1.0'

UCDs in VO services

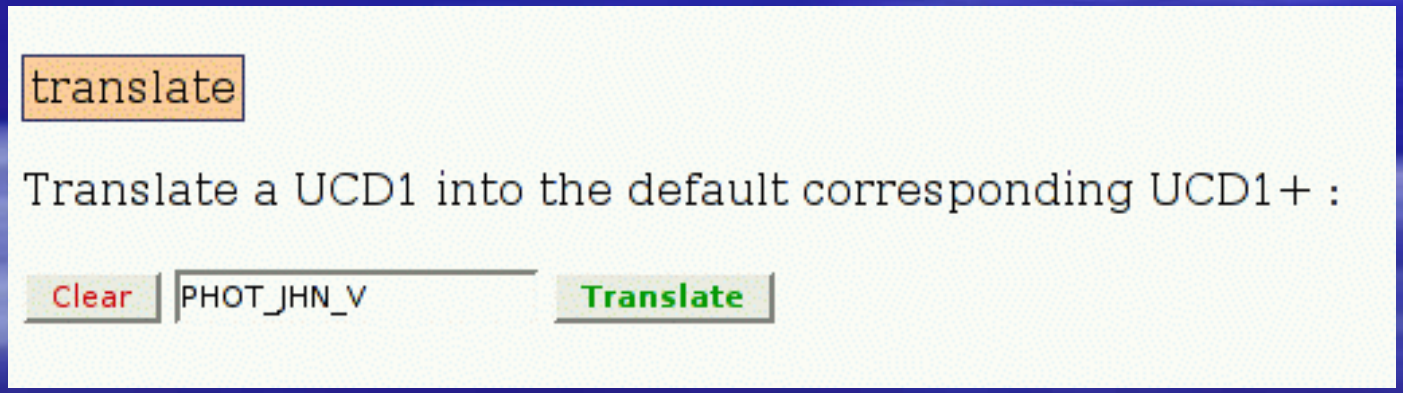
- You are not obliged to rename all your parameters with UCDs :o)
- UCDs are only needed during data exchange, so you can use a translation layer:
  - attach UCDs to your data, e.g. when exporting VOTable documents (for data providers)
  - interpret UCDs present in a query in terms of expected input parameters (for a VO service)

UCD1+ tools

- A set of tools has been developed to:
  - Ease transition from UCD1 to UCD1+
  - Manipulate UCD1+
- Tools available as CGI scripts and Web Services
  - http://vizier.u-strasbg.fr/UCD/tools.htx
  - http://cdsweb.u-strasbg.fr/cdsws/ucdClient.gml

translate

- Translate a UCD1 into the default UCD1+
- For data providers who already use UCD1



validate

- Check if a UCD1+ is correctly written
- Possible actions depending on the return status

validate

Check if a UCD1+ is valid :

Clear | pos.pm;ivoa:pos.eq.ra | Validate

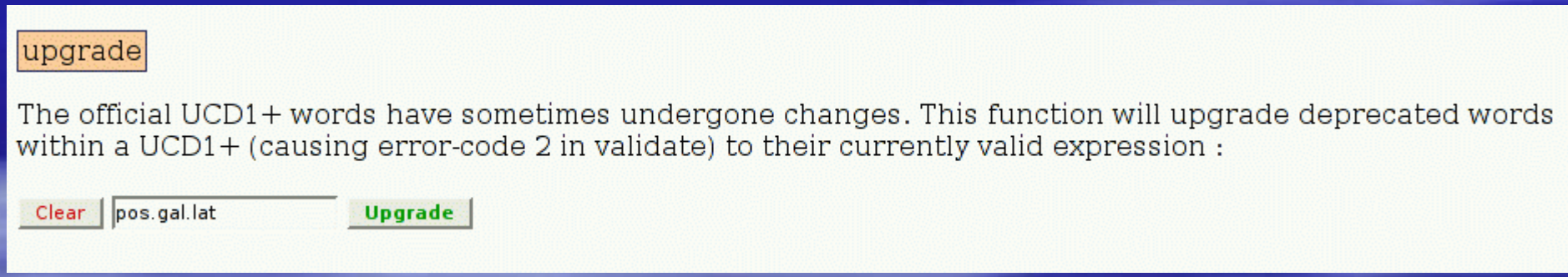
The first word of the result is an error code, possibly followed by an explanation of the error.
This function will return 0 if the UCD1+ is fully valid.
The error-code results from the combination (logical OR) of the following values:

- 1: warning indicating use of non-standard namespace (not ivoa:)
- 2: use of deprecated word
- 4: use of non-existing word
- 8: syntax error (extra space or unallowed character)

Example: an error-code value of 10 will indicate the use of a deprecated word and a syntax error.

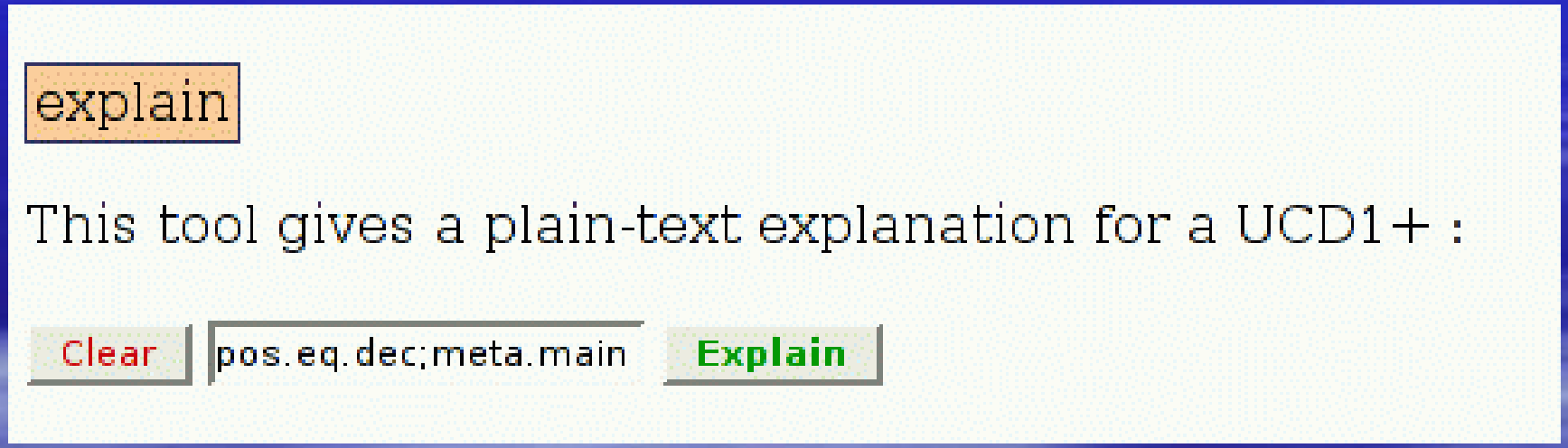
upgrade

- Convert deprecated UCD1+ to the new recommended expression
- Useful to take into account changes in vocabulary



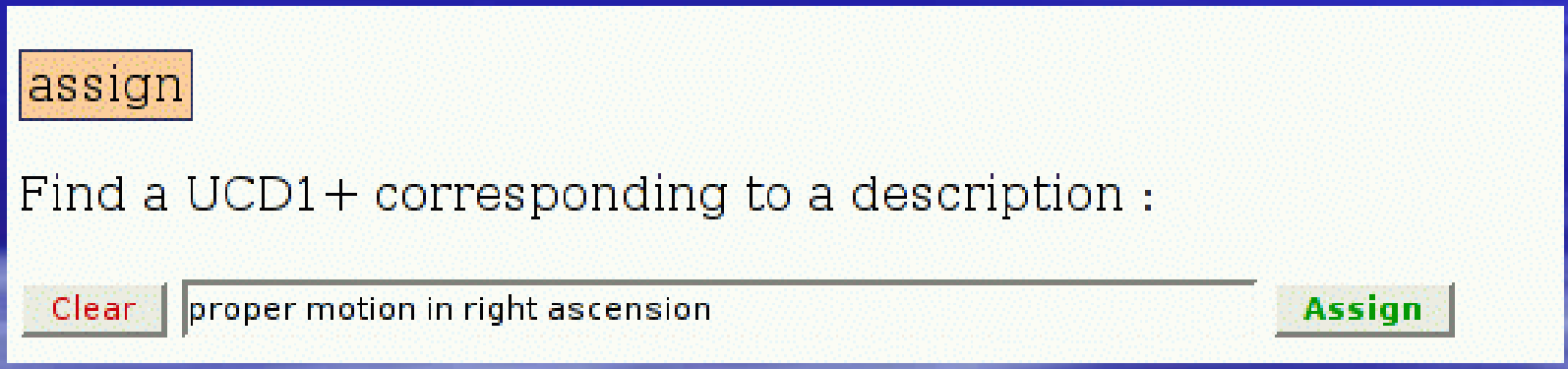
explain

- Returns explanation for a given UCD1+



assign

- Returns a UCD1+ corresponding to a plain text description (difficult job!)



Interactive assign

UCD builder

Use this interface to find the UCD corresponding to a description:

Enter a description in natural language:

Clear Form | Reset Form | maximum value of R magnitude | Build UCD

Last updated Thu May 12 15:54:16 2005

Suggested complete UCD: phot.mag;em.opt.R;stat.max

Refine your search:

The following words matched your query. You might force the selection of words matching precisely your query, and rebuild a UCD using these words:

Clear Form | Reset Form

word	definition	flag	score
☐ phot.mag	Photometric magnitude	E	60
☐ phot.mag.bol	Bolometric magnitude	Q	50
☐ phys.magAbs.bol	Bolometric absolute magnitude	Q	30
☐ em.opt.R	Optical R band between 600 and 750 nm	S	15
☐ stat.max	Maximum or upper limit	S	30

Rebuild UCD using selected words.

Thanks !

See you during the hands-on tutorial!