Easy comparison of dataset footprints with MOC

P. Fernique & T. Boch

Centre de données astronomiques de Strasbourg

Multi-Order Coverage map: A sky area definition based on HEALpix for building very fast access and comparison tools (Xmatch, database,...)

The background: new developments at CDS, notably Aladin allsky V7 and the new cross-match tool for very large catalogues, led the CDS team to define and use a new method to describe regions of the sky. We aimed at addressing use cases such as: - find the common area between multiple catalogues; - which objects of my list have potential counterparts in this dataset; -visualize the coverage of a survey. The major constraint is to answer in a very short time, ie a few milliseconds, the following question: "*Does this data set intersect such or such region of the sky?*". Astronomical data bases are obviously able to solve this problem. Usually the search region is a cone of a few arc-minutes radius, and typical response times are a few seconds, growing proportionally with the area covered by the cone. These response times are mainly due to the data base indexation method and also to the conversions and computations of spherical distances required to solve this kind of request.

The idea: In order to shorten this delay, we opted for a static and homogeneous description of data sets using the sky tesselation defined by the HEALPix(1) standard. HEALPix is a sky pixelisation method based on 12 diamonds, recursively subdivided into four sub-diamonds. HEALPix is well known notably by the users of the WMAP or PLANCK missions. It has the advantage, over its competitors QBOX(2) or HTM(1), to provide equal surface cells for a given level, and a constant computation time for any level. The key idea(3) is to use such a system to **define a region of the sky by the list of corresponding HEALPix cells, grouping them hierarchically** in order to reduce the volume required by the description. We called this system "**MOC**" for "**Multi-Order Coverage map**".

FITS

ASCII

The results: With the help of HEALPix developers, we have defined a **FITS encoded format** and an **ASCII format** in order to easily manipulate them. We also create a Java library for writing, reading, comparing and also to compute some simple operations such as union and intersection. We generated the MOC for each VizieR catalog (~ 10,000) and for some image surveys available in Aladin V7 (~ 50). These MOCs are already available and can be visualised in the Aladin beta version, which supports the MOC formats.

2.5s: for loading 10 000 VizieR Mocs (level 8)

7ms : for simple query
(which catalogs are concerned by this Hpx cell ?)

90ms : for Moc query
(which catalogs intersect this Moc (XMM => 4500 cells) ?

(realized on a basic 3 years old laptop)

VO: Such a system shared by several data centers, or even being accepted as a standard at the IVOA level, would significantly improve in the short term a lot of VO tools, for instance TOPcat, the well-known tool for handling astronomical tables. Currently, this tool generates more than half of the VizieR queries mainly to cross-correlate the user data set with such or such VizieR catalog. It would be infinitely more effective if TOPcat would compare first the both MOCs before operating the source cross-identification in order to consider only the sources present in the intersection of the data sets.

GALEX Moc
Hpx level 8
69487 cells
278KB

SDSS Moc
Hpx level 10
223574 cells
880KB

Links & references:
Java API : http://cds.u-strasbg.fr/resources/doku.php?id=moc
Moc repository: http://alasky.u-strasbg.fr/footprints/
Aladin displayer: http://aladin.u-strasbg.fr/java/AladinBeta.jnlp

(1) MPA/ESO/MPE Workshop 2001 – O'Mullane et al. "Splitting the Sky - HTM and HEALPix"
(2) QBOX - VizieR catalog service tesselation system – F.Ochsenbein
(3) PASP 2010 - Budavari et al. "Searchable Sky coverage..."