BROWSING THE PDS IMAGE ARCHIVE WITH THE IMAGING ATLAS AND APACHE SOLR. K. M.

Grimes II[1], J. H. Padams[1], A. Stanboli[1], and K. L. Wagstaff[1]. [1]Jet Propulsion Laboratory, California Institute of Technology. {first}.{mi}.{last}@jpl.nasa.gov.

Introduction: The PDS Imaging Node [1] is home to tens of millions of images, nearly 30 million have rich metadata associated with them in the form of PDS labels. Naturally, this metadata varies from image to image, but examples of what it may contain include the following: when it was taken, what spacecraft and instrument took it, and the physical location of the spacecraft when it took the image. This information is stored in the image's label as key-value pairs.

A pressing need of scientists is to locate images with similar characteristics. For some characteristics, such as *mission name*, locating these images by hand may be a trivial process; however, for multi-mission characteristics such as *product type* and *spacecraft clock*, locating images with similar characteristics becomes much more difficult to do by hand. Complicated requests such as "show me all images taken in 2001 with wheels in them" approach impossibility. With the assistance of big data tools, however, such queries may be made.

In order to allow users to search the archive by any available characteristic, the Imaging Node leverages Apache Solr ("Solr") [2], an open-source indexing platform built upon another Apache product, Lucene [3]. The PDS Imaging Atlas ("the Atlas") [4] provides an interactive interface to Solr and allows users to easily and intuitively navigate the archive. Content-based search is enabled via a two-part process: first, a convolutional neural network identifies features in images; next, these features are indexed by Solr so that they can be searched by the Atlas.

Apache Solr: In order to effectively index PDS-compliant images, the Imaging Node creates a *collection* and defines *fields* within a *schema*. Each field in the schema corresponds to a key in the image's label.

Solr provides support for several different types of fields out-of-the-box, including *strings*, *multi-values* (arrays), and *floats*. Custom field types may also be created.
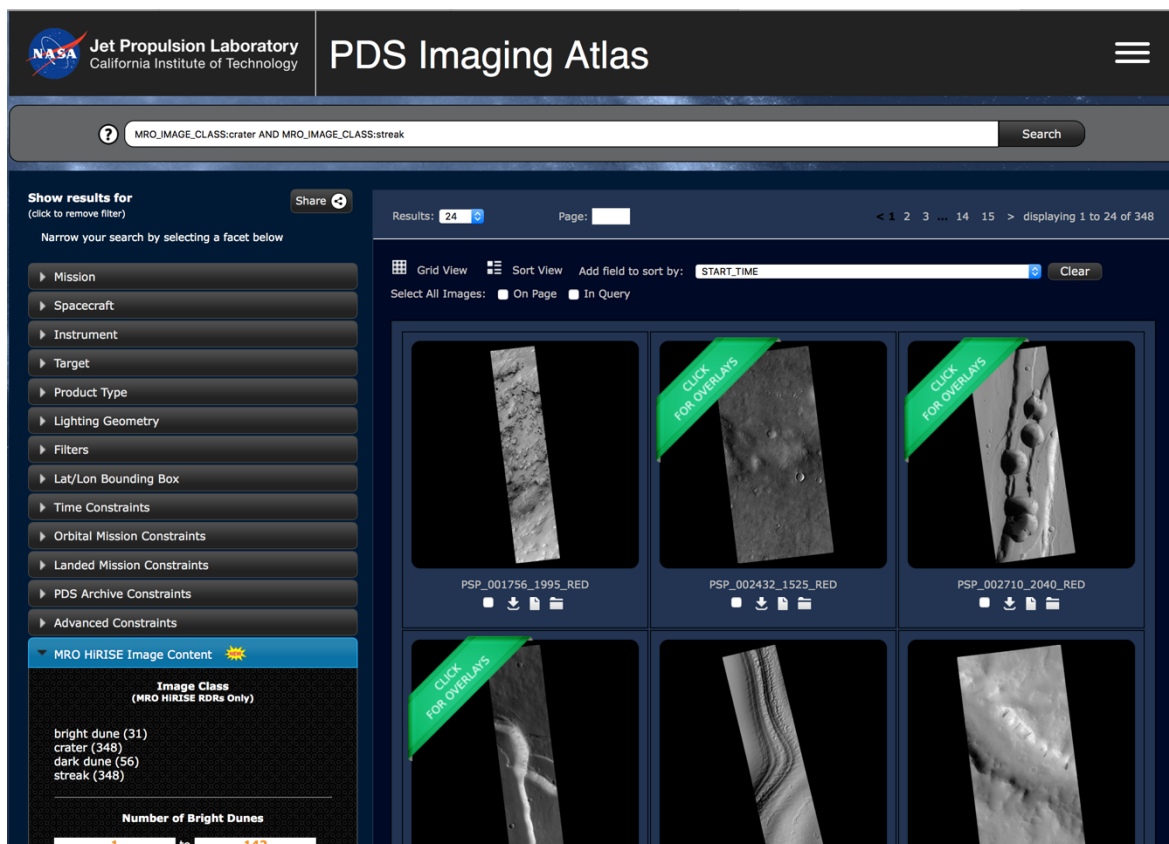

Figure 1: The PDS Imaging Atlas.
https://pds-imaging.jpl.nasa.gov/search

Once the schema has been defined, the main indexing procedure may begin. After some time, the indexing is completed, and the collection is ready to be searched.

Solr Search. One images have been indexed, clients may query for them by any of the fields that they declared in the schema. For example, if a user desires to search for all images that were taken by the Cassini spacecraft and are EDRs, they could make a query to Solr like the following:

```
MISSION:CASSINI AND PRODUCT_TYPE:EDR
```

Query 1: Request for all EDR products produced by the Cassini spacecraft.

Suppose that we want to make a content-based search, for example, of all images taken by the Mars Science Laboratory rover from spacecraft clock counts greater than or equal to 397,000,000 that contain wheels. To do so, we could make the following query to Solr:

```
MISSION:"mars science laboratory" AND
SPACECRAFT_CLOCK_START_COUNT:[39700000
   0 TO *] AND MSL_IMAGE_CLASS:wheel
```

Query 2: Request for products produced by the MSL lander after spacecraft clock count 397,000,000 that contain wheels in them.

Solr supports more complicated queries, including nested conditionals and negations.

Browsing the Archive: Because Solr's query syntax can be a bit daunting to users, PDS Imaging Node provides the Atlas as a frontend service to Solr. The Atlas is an interactive interface that allows users to make queries to Solr and interact with the images it returns in an intuitive way.

Facet tabs. Users are able to refine their search by clicking on one of the various facet constraints found in the left hand side view of the Atlas. The available facets to choose from include Mission, Spacecraft, Instrument, and Product Type. When users select a given constraint, a query is made to Solr for documents matching what the user click on.

For example, if the user clicked on the "Cassini" constraint under the "Mission" facet tab and then proceeded to click on the "RDR" constraint under the "Product Type" facet tab, a query similar to *Query 1* above is made to Solr, and matching images are displayed to the user.

Search box. In addition to refining searches using the facet tabs on the left, users may also enter their own custom queries in the search box at the top of the Atlas.

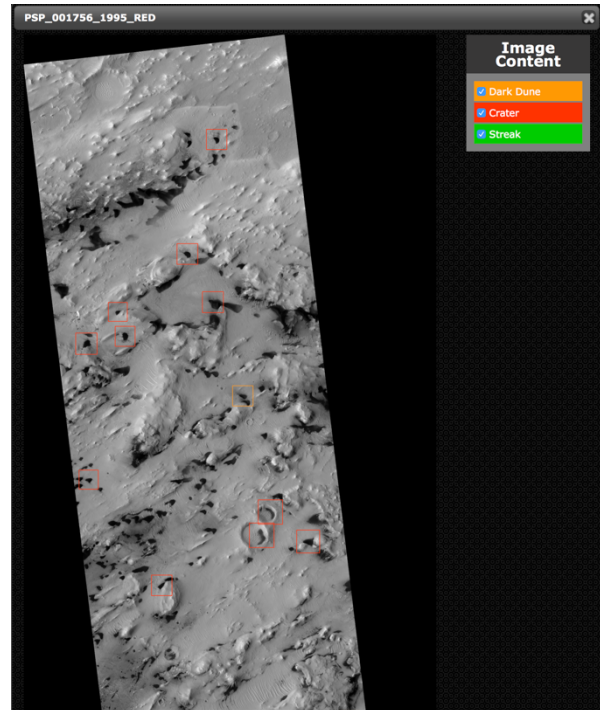For example, both *Query 1* and *Query 2* could be entered into the search box directly and Solr would re-

Figure 2: Feature recognition of Martian landscape in the PDS Imaging Atlas.

turn matching documents. This functionality is especially useful in cases where users want to search for constraints on facets that do not have their own facet tab.

MRO HiRISE Overlays. In collaboration with machine learning teams at Jet Propulsion Laboratory [5], the Atlas allows users to not only search for images by the features they contain but also to identify these features with bounding boxes on the image.

For example, if a user selects an image taken by the HiRISE instrument that has contains features such as dunes or craters, an overlay will be displayed on top of the image indicating where these dunes and craters are. An illustration of this functionality can be found in *Figure 2*.

References: [1] Planetary Data Systems Imaging Node: https://pds-imaging.jpl.nasa.gov. [2] Apache Solr: https://lucene.apache.org/solr. [3] Apache Lucene: https://lucene.apache.org. [4] PDS Imaging Atlas: https://pds-imaging.jpl.nasa.gov/search. [5] Jet Propulsion Laboratory: https://jpl.nasa.gov.