SPECTRUM SERVICES 2007

László Dobos[1]**, Tamás Budavári**[2]**, István Csabai**[1]**, and Alex S. Szalay**[2]
[1]Eötvös University, Budapest, Hungary
[2]The Johns Hopkins University, Baltimore, USA

ABSTRACT

We present the Filter and Spectrum Services consisting of easy-to-use web applications and web services for searching, plotting and managing large collections of spectral energy distribution data and filter profiles as well as for performing various scientific operations on spectra in a unified framework. The services provide keyword search, advanced query forms and SQL query possibilities for selecting spectra or bandpass curves which may be retrieved in a variety of file formats including XML, VOTable and ASCII. All SDSS DR1-DR5 spectra had been loaded into a database as well as the entire 2dF catalog that adds up to more than 2.5 million SEDs of about a million spatial objects, but registered users can upload their own data making it available for the rest of the community and are free to modify or delete them at any time. Theoretical catalogs, such as the Bruzual-Charlot stochastic burst model spectrum library (100k spectra) and the BaSeL stellar library are also available. Scientific services allow building rest-frame composite spectra out of selected spectra; calculating synthetic magnitudes by convolving spectra with an arbitrary set of bandpass curves of optical instrument filters to generate simulated photometric catalogs on-the-fly; galactic extinction correction, fitting of the continuum using different set of templates (Bruzual-Charlot 03 templates, SDSS eigenspectra), line fitting. All scientific functionalities are available from the web user interface and via the SOAP web services for programmers. MySpectrum is a cross-platform version of the spectrum web service for setting up your own spectrum repository. It integrates into the main service allowing easy access to your data for the whole VO community.

The main idea behind our web services is to move scientific functionalities physically close to the database in order to spare network bandwidth. This way scientists may do research without setting up expensive hardware, downloading large datasets for days or weeks or installing complicated software.

Key words: Technique: spectroscopic; Virtual Observatory.

1. FILTER SERVICE

The Filter Service is a searchable database of optical instrument bandpass filter curves. The database contains about a hundred filter curves of the most important astronomical instruments. The Filter Service allows users to upload their own bandpass curves and share them with the VO community, and also shares its database with the Spectrum Services, so uploaded filters will appear in the Spectrum Services when, for example, synthetic magnitude calculation is required. In order to post your own filter curves, you have to register on the web site.

The Filter Service supports a simple method of searching by keyword in the filters descriptions. Click on the “search” link on the navigation bar at the top of the web page to get the search form. You can get the list of all available filters by clicking on the “List all filters” button, or get a subset by specifying a portion of the name of filter, for example SDSS in order to get all Sloan Digital Sky Survey filters.

Filters also can be plotted on a graph and downloaded in XML, VOTable or ASCII format.

The Filter Service is accessible from http://voservices.net/filter.

2. SPECTRUM SERVICES

The web user interface of the Spectrum Services provides much more functionality than the Filter Services does. You can access a broad range of different search forms to find your spectra of interest. The Spectrum Services are accessible from http://voservices.net/spectrum.

2.1. The Spectrum Database

The Spectrum Services incorporate a bunch of searchable databases of observed and theoretical spectra of spatial objects. The databases altogether contain about two million spectra of one million objects, mainly from the SDSS and 2dF surveys. Most of these spectra are calibrated, so broadband magnitudes of the original objects can be determined by convolving spectra with optical filter curves or more useful scientific information can be derived from them.

The different surveys are organized in different datasets referred as collections. Each collection contains data from a single sky survey. Collections may be located at different geographical locations; the central service can query them remotely and unify the data from the different catalogs.

2.2. Searching the spectrum databases

There are multiple search methods available:

Object Search is for finding spectra of spatial objects by coordinates. An "cross-match radius" must be specified (in arcseconds) in which the matching objects must reside in order to be returned in the result set. Coordinates must be given in decimal or hh:mm:ss, dd:mm:ss format, one on each line.

ID Search is for retrieving spectra by their database identifier. Database identifiers have two parts, one for identifying the dataset and one for identifying the individual observation.

Cone Search is for retrieving spectra of objects in a given circle on the sky. Cone search is a basic functionality of most VO services. A coordinate pair in degrees or in hh:mm:ss, dd:mm:ss format must be specified. The search radius must be in arcminutes.

Advanced Search allows searching the catalog by numerous criteria. Please be sure to specify criteria as strictly as possible, otherwise the result set may be very large. Typically, specifying a redshift range of 0.01 returns several thousand matches just within the Sloan catalog.

SQL Search is for the advanced users who are familiar with the SQL database query language. The database schema can be found in the documentation section of the web site. SQL Search allows you to specify more sophisticated search criteria than what the web forms provide.

Redshift Search is useful when spectra of all objects in a given redshift range are required. Be careful when specifying the range, since a large collection may contain hundreds of thousands of observations.

The *Get Whole Collection* option is useful when you want to download all the spectra in a predefined set, for example a set of theoretical templates. Note, that large collections cannot be downloaded this way due to bandwidth and performance restrictions.

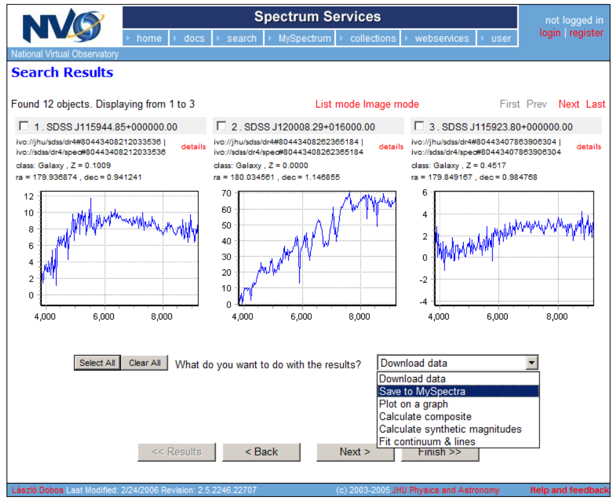

Figure 1. Screenshot from the Spectrum Services website displaying a list of search results.

3. SCIENTIFIC FUNCTIONALITY

3.1. Preprocessing Spectra

Before any further analysis, spectra should be preprocessed in order to be in an appropriate form for the scientific analysis. The Spectrum Services library provides the following operations for preprocessing spectra:

Dereddening: We use Schlegel's dust map and Cardelli et.al and O'Donell galactic extinction curve to deredden galaxy spectra on-the-fly. The extinction coefficient is automatically looked up at the coordinates associated with the object.

Redshift: Takes the original spectrum and multiplies the wavelength values to represent a spectrum of an object as it was at z = 0 (for the rest-frame) or at any given redshift. The web user interface only allows converting spectra into the rest frame.

Vacuum-to-air wavelength conversion: SDSS spectra are represented in vacuum wavelengths while most theoretical models use air wavelengths, so this conversion is required before fitting templates to observed spectra.

Normalization: Normalizes galaxy or QSO spectra using a special algorithm to deal with noise and masks in observed spectra.

Resampling: Resampling can be done in two ways: either specifying an arbitrary list of wavelength values or specifying the start wavelength, the increment and the number of required points. Logarithmic resampling is also possible. Resampling can use nearest neighbor or linear interpolation. Resampling is done automatically when using the web user interface. Explicit resampling to a given grid of wavelength values can only be done using the web service interface from a client program.

3.2. Continuum and Line Fitting

Spectrum Services provide an implementation of the non-negative and the standard least-squares fitting methods for continuum fitting. You can choose from different sets of theoretical templates, and different masking methods are available to mask strong emission lines, the night sky line and use mask values from the spectra.

To try fitting, run a search as detailed previously, pick some spectra, choose “Fit continuum and lines“ from the drop-down menu and click “Next“. After setting the preprocessing options, the fit options form will come up. You can choose from non-negative or standard least-squares fitting. The non-negative least-squares method is required for theoretical template fitting, when only a positive coefficient linear combination of templates is physically meaningful. The standard method can be used, for example, to expand spectra on a predefined basis, such as the first five eigenspectra of the SDSS galaxies.

You can select “Weigh with errors“ in order to reduce the significance of noisy parts of the spectrum, while the different masking options allow you to completely omit certain parts of the wavelength range when fitting the continuum. Numerous well known emission/absorption lines are predefined but the web user interface does not allow custom lines to be defined; for this use the web service interface instead from a client program.

Velocity dispersion should be set to a starting value for the line fitting algorithm, but it is also used to determine the line width for masking.

Different predefined template sets can be chosen from the drop-down list. Fitting to an arbitrary set of templates is only available through the web service interface from a client program. Choose a template set and pick some templates to fit to. Click “Next“ to get to the target selection form. You can view results in HTML format or download them in XML for future use.

3.3. Synthetic Magnitude Calculation

To compute synthetic magnitudes by convolving spectra with a set of filter curves, first run a search that returns some spectra. Select some of them by clicking on the checkboxes, or click on the “Select all“ button. Choose “Calculate synthetic magnitudes“ from the drop-down list.

Defining the preprocessing options is exactly the same as described in the previous section. After the preprocessing parameter pages a form with the list of available filter curves will come up. You can select one or more filters to convolve spectra with. Note that filters are usually wider in the wavelength domain than spectrographs, so convolution yields to invalid magnitude values. If a filter covers a wider range than the selected spectra a notification will be displayed saying “too wide“ in the right column of the filter list.

After clicking the “Next“ button the list of computed magnitudes is displayed. The web interface does not currently allow saving results in any format: use a web service client instead.

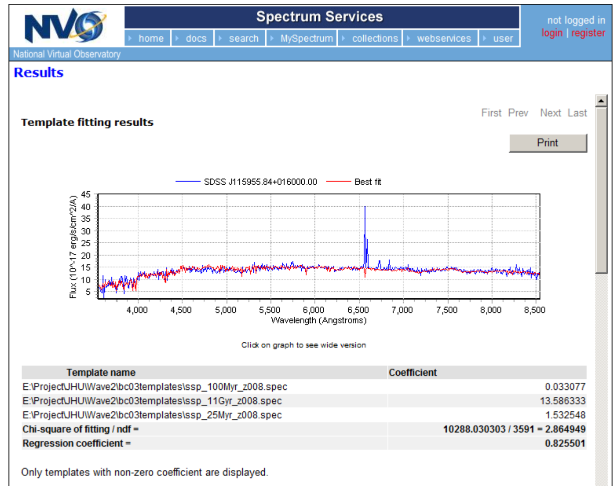

Figure 2. Screenshot from the Spectrum Services website displaying the results of line fitting.

3.4. Composite Spectrum

After converting spectra into the rest-frame and normalizing them, preprocessed spectra can be stacked using different mathematical operations, like average, median or sum. Composite calculation shows the real advantage of using Spectrum Services instead of ordinary tools: composites of thousands of spectra can be computed on the remote server and only the result travels through the network.

4. MY SPECTRUM

In order to enable data providers to incorporate their data into the Spectrum Services as a remote collection, we provide an implementation of the web service server interface.

MySpectrum is a simplified version of the Spectrum Services without the scientific functions. It is a platform independent spectrum repository that provides the same query interface as the Spectrum Services for seamless integration.

The MySpectrum program is written in the C# language and runs on the Linux platform using the Mono framework. On Windows the Microsoft .Net Framework is needed. It can access Microsoft SQL and MySQL database servers, but other DBMSs can be integrated easily in the future. We added hierarchical triangular mesh (HTM) support to the MySpectrum package, so spatial indexing is also available on MySQL.

Scientists can download and install the program, set up their own database and load their data into it. The main Spectrum Services portal allows registered users to add their own data source to the list of available collections. It requires only registering a single URL pointing to the MySpectrum web service endpoint.

5. ONGOING AND FUTURE DEVELOPMENTS

Spectrum Services is under continuous development and new versions are coming out periodically. Since the underlying databases are large, numerous software engineering problems have to be solved. Current development issues include upgrading the whole system to the latest version of the application server (ASP.NET 2.0) which will significantly increase the performance. A new streaming spectrum processing model is also under development that will allow scientists to process a really large amount of data simultaneously on multiple processors with minimal memory usage to serve as many clients as possible.

The current stable version does not conform to the final IVOA spectrum data model standard that was accepted in May 2007. Migration of the system to the IVOA standard data model is also an ongoing process.

Some new search methods are also under development. Since the Footprint Service (for describing spatial coverage of sky surveys) is on-line, it is possible to connect with the Spectrum Services and search for spectra of the objects in a precisely defined complex area of the sky. The other search method is what we call similarity search. It is going to find similar spectra to an uploaded sample spectrum using multi-dimensional search techniques. A typical application of this feature is to find similar spectra in a simulated catalog, thus to find the physical properties of observed galaxies.

6. URLS OF INTEREST

Spectrum Services can be found at `http://voservices.net/spectrum`. The Filter Service is at `http://voservices.net/filter`.

ACKNOWLEDGMENTS

The development of Spectrum Services was partially supported by the following grants: OTKA-T047244, MSRC-2005-038, NAP-2005/ KCKHA005 and MRTN-CT-2004-503929.