Harvard Faculty Finder (HFF) APIs

Introduction

The Harvard Faculty Finder (HFF) website creates an institution-wide view of the breadth and depth of Harvard faculty and scholarship, and it helps students, faculty, administrators, and the general public locate Harvard faculty according to research and teaching expertise. More information about the HFF website and the data it contains can be found on the Harvard University Faculty Development & Diversity website. HFF is a Semantic Web application, which means its content can be read and understood by other computer programs. This enables the data associated with a person, such as titles, contact information, and publications to be shared with other institutions and appear on other websites. Below are the technical details for building a computer program that can export data from HFF.

Technical Overview

As a Semantic Web application, HFF uses the Resource Description Framework (RDF) data model. In RDF, every entity (e.g., person, publication, concept) is given a unique URI. (A URI is similar to a URL that you would enter into a web browser.) Entities are linked together using "triples" that contain three URIs--a subject, predicate, and object. For example, the URI of a Person can be connected to the URI of a Concept through a predicate URI of hasResearchArea. HFF contains millions of URIs and triples. Semantic Web applications use an ontology, which describes the classes and properties used to define entities and link them together. HFF uses the VIVO Ontology, which was developed as part of an NIH-funded grant to be a standard for academic and research institutions. A growing number of sites around the world are adopting research networking platforms that use the VIVO Ontology. Because RDF can link different triple-stores that use the same ontology, software developers are able to create tools that span multiple institutions and data sources. When RDF data is shared with the public, as it is in HFF, it is called Linked Open Data (LOD).

There are four types of application programming interfaces (APIs) in HFF.

- RDF crawl. Because HFF is a Semantic Web application, every person has both an HTML page and a corresponding RDF document, which contains the data for that page in RDF/XML format. Web crawlers can follow the links embedded within the RDF/XML to access additional content.
- SPARQL endpoint. SPARQL is a programming language that enables arbitrary queries against RDF data. This provides the most flexibility in accessing data; however, the downsides are the complexity in coding SPARQL queries and performance. In general, the XML Search API (see below) is better to use than SPARQL. The URL of the SPARQL API is http://api.facultyfinder.harvard.edu/public/SPARQL/ProfilesSPARQLAPI.svc/search.
- XML Search API. This is a web service that provides support for the most common types of queries. It is designed to be easier to use and to offer better performance than SPARQL, but at the expense of fewer options. It enables full-text search across all entity types, faceting, pagination, and sorting options. The request message to the web service is in XML format, but the output is in RDF/XML format. The URL of the XML Search API is http://api.facultyfinder.harvard.edu/public/Search/ProfilesSearchAPI.svc/search.
- "Beta" Search API. This is an older API, which is no longer supported. The XML Search API is the recommended method of querying HFF. The URL of the Beta Search API is http://api.facultyfinder.harvard.edu/public/beta/ProfileService.svc/ProfileSearch.

Detailed Documentation

HFF is based on the open source Profiles Research Networking Software (RNS) platform. Below are links to the API documentation files for Profiles RNS. Make sure you use the API URLs listed in the above section rather than the defaults listed in the documentation.

- First read the main API Documentation file.
- The API Examples file will help you get started.
- The VIVO ontology is complex, but the HFF Ontology Overview file illustrates the key classes and properties.
- The full ontology can be downloaded as three Web Ontology Language (OWL) files: VIVO, PRNS, and Catalyst.

HFF Statistics

HFF contains 2,636,517 RDF nodes and 11,570,908 triples. Though, these numbers are continually increasing as we import new data sources into HFF and faculty edit their pages. Below are counts for selected classes in HFF.

Class Name	Class URI	Items
Book	http://purl.org/ontology/bibo/Book	2,164
Course	http://vivoweb.org/ontology/core#Course	18,263
DASH Article	http://profiles.catalyst.harvard.edu/ontology/catalyst#DashArticle	6,846
Department	http://vivoweb.org/ontology/core#Department	91
Division	http://vivoweb.org/ontology/core#Division	83
HOLLIS Keyword	http://profiles.catalyst.harvard.edu/ontology/catalyst#HollisKeyword	1,863
Organization	http://xmlns.com/foaf/0.1/Organization	187
Patent	http://purl.org/ontology/bibo/Patent	911
Person	http://xmlns.com/foaf/0.1/Person	11,018
Project	http://vivoweb.org/ontology/core#Project	866
Web of Science Article	http://profiles.catalyst.harvard.edu/ontology/prns#WebOfScienceArticle	244,505