



## Semantic Search - Faceted Search and Web 3.0 for Drupal

### ***User Guide***

This document provides users basic documentation for the Drupal module "Semantic Search"

### **CivicActions, LLC**

Jonathan Hendler

#### **Table of Contents**

Table of Contents.....	1	Development Status Summary.....	7
Mission.....	2	Successes.....	7
Dependencies.....	2	Current Challenges.....	8
Architecture.....	2	Development Roadmap Summary.....	8
Data Flow.....	3	Conclusion.....	9
Installation .....	3	Appendix.....	9
Setup.....	4	Revision History.....	9
Creating a Search .....	4	Appendix :: Tables.....	9
Adding Facets to a Search .....	7	Appendix : : Acknowledgments .....	9
Building the Search and Managing the		Appendix : : Diagrams.....	10
Cache.....	7	Appendix : : Links.....	10
Day to Day Use.....	7	Appendix : : External Resources.....	10
Import and Export.....	7	Appendix :: License and Copyright...	10



## Introduction

Semantic Search is a Drupal module for using Drupal content and meta-data in search and syndication. Included is a default Web 2.0 interface, administrative interfaces, and APIs for custom search and Web 3.0 applications.

Semantic Search is not for all kinds of search, or all kinds of data. Semantic Search is best used with data that has many fields (columns), but when the content of those fields is not consistent, known, or highly structured. Data that has one large text field or, alternatively, highly structured nodes can use Drupal's Views, custom SQL and searches, or Drupal's already well developed Search Module. Other reasons to *not* use Semantic Search are that Semantic Search is a new technology, may require programming, and has a number of dependencies.

Web 3.0 is also known as the Semantic Web. The Semantic Web creates highly structured meta data. The logo for Semantic Search is derived from Semantic Web concepts. The Semantic Search logo is the Drupal logo in the form of a "triple", which is the way that data on the Semantic web is generally stored.

Semantic Search can be licensed under either the GPL 3.0 or the AGPL (Aferro General Public License).

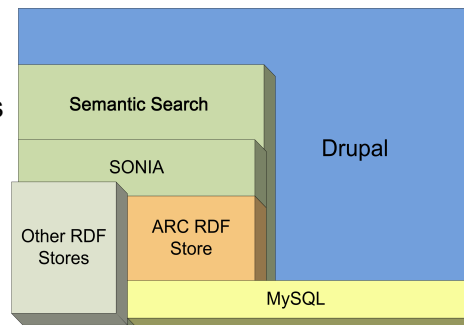
## Mission

Semantic Search makes the development of advanced search tools easier. The mission serves small organizations, whose websites are already running Drupal. Semantic Search provides a more structured tool than text search to take advantage of Drupal's taxonomy and other structured content. The long term mission is to help Drupal be ready for the next generation of the web known as Web 3.0 or "The Semantic Web".

## Dependencies

Semantic Search is a Drupal Module bundled with SONIA and ARC. Semantic Search needs

- Drupal 5.x.x
  - CCK, taxonomy, and other modules
- PHP 5.x
- MySQL 5

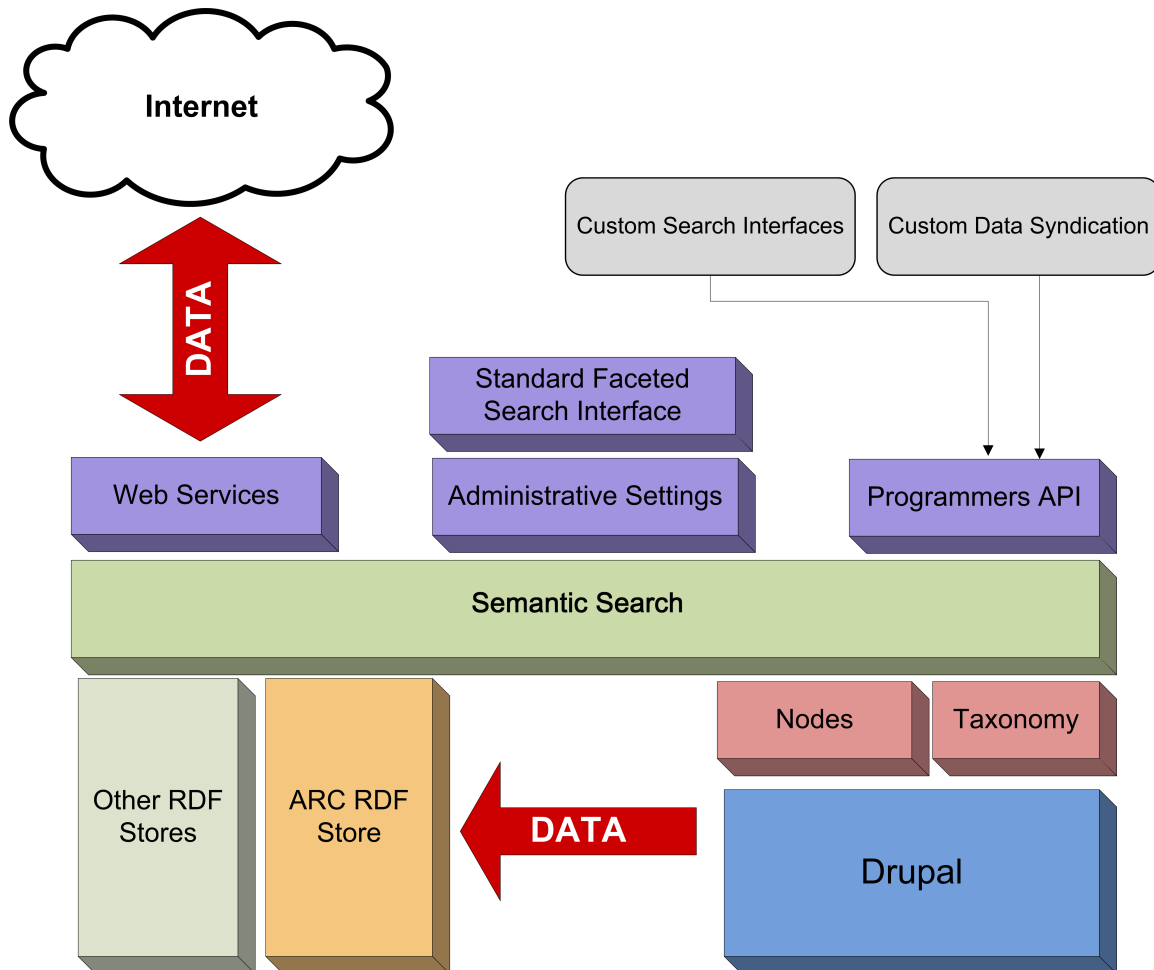


## Architecture

### Data Flow



## Installation



To install Semantic Search download the files from the project page.

Install via `admin/build/modules` the **Semantic Search** and **Semantic Search RPC** modules.

You will then see a message on most pages :

Please [configure SearchFacets](#)

Go to that page to create the basic settings needed to begin setting up Semantic Search. On this page you can “install database tables” again, set up the basic site information and the database settings for RDF Store(s).



Home > Administer

SearchFacets Search Configurations Settings Export Import

Please configure SearchFacets

The configuration options have been saved.

☒ Install SearchFacets database tables.  
You may need to install the tables for indexes, caches, and RDF data.  
(It is safe to run this after the initial install.)

Basic Settings

Settings for All Searches

Base Url:  
  
For this site it should be \$base\_url = http://fsearch

Base RDF namespace URI:  
  
The base URI used for the data on this site.

RDF Store Settings

SearchFacets uses SONIA (Simple Ontology and Inference Algorithms) as an abstraction layer for for various RDF stores.  
There is planned support for RDF stores running remotely, locally, or aggregated from across sites.

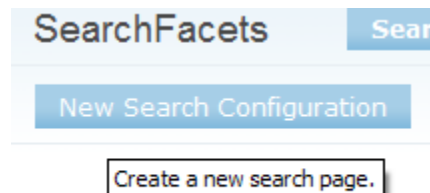
ARC Local Settings

Save configuration Reset to defaults

### Setup

Semantic Search lets you have multiple searches and their underlying data stores by using a web based administration. The interfaces are themable, and a default theme is included. For more advanced module developers, functionality can be called via a programmers API.

### Creating a Search



Searches are created by creating a configuration.

The searches can be edited for basic settings, cache and Drupal import settings, which Facets display in the search form, Import, Export, and how the search is managed.



New Search Configuration

Animal Facets

Cache Settings

Main Settings

Main Settings for Creating or Modifying a Search

### Search Name:

animals

**VERY IMPORTANT** readable name like "animals", "birds", or "fish" .  
Is used for the cache name, index name, rdf tables name etc.

### Search Path:

animals

The url for users to access the search. Eg `search/myanimals`

### Search Type:

default

- **default**: a default interface. Can be themed.
- **custom api**: creates an API available for RPC calls.
- **syndicated**: custom module for aggregating content dynamically.

### Search Title:

Animal Facets

The human-friendly, display name for the search. Like "My Search", or "Animals"

### Search Description:

Search Animals by different descriptions, tags, and Taxonomy

Description of the search.

Data is imported via cron, or manually.



### ▼ Data Settings

Settings for what Drupal data will be searched.

#### ☐ **NODE SYNC**

Select node and data types below. All data and related taxonomy terms will be imported and indexed. This could take a VERY long time.

#### ☐ **Clear Search Cache and RDF data.**

Also, clear ALL the data in NINA - only use this when starting over!!!

#### **Batch Size:**

1000

There are currently 0 of 0 total nodes imported this search.

It takes a long time to build RDF. This lets you build in batches.

### Node Synchronization

#### ☐ Store CCK type **Blog entry** in NINA.

Should CCK type *Blog entry* (A blog is a regularly updated journal or diary made up of individual posts shown in reversed chronological order. Each member of the site may create and maintain a blog.) be stored in NINA to be available for RDF search and export"?

#### ☐ Store CCK type **Page** in NINA.

Should CCK type *Page* (If you want to add a static page, like a contact page or an about page, use a page.) be stored in NINA to be available for RDF search and export"?

#### ☐ Store CCK type **Story** in NINA.

Should CCK type *Story* (Stories are articles in their simplest form: they have a title, a teaser and a body, but can be extended by other modules. The teaser is part of the body too. Stories may be used as a personal blog or for news articles.) be stored in NINA to be available for RDF search and export"?

#### ☐ Store CCK type **test 1** in NINA.

Should CCK type *test 1* (asfd) be stored in NINA to be available for RDF search and export"?

### ▼ test 1 Field Options

Should CCK field the field display in the search results?

☒ sdfsd : Display in search Results?

Weight



## Adding Facets to a Search

**SearchFacets**   Search Configurations   Settings   Export   Import

New Search Configuration   **Animal Facets**

—▷ Cache Settings

—▷ Main Settings

—▷ Data Settings

▽ Default Interface Settings

Settings for "default" and "module" interface searches.

**Number of Results:**

Number of search results to show per page.

▽ The Search Form

Create the search form here.

Facet Name	Description		
Nomenclature Taxonomy	The latin binomial nomenclature	Edit	Delete
Free Tags	our users tag the data	Edit	Delete
Text Search	Advanced text search of descriptions	Edit	Delete

Add Facet

—▷ Import

—▷ Export

## Building the Search and Managing the Cache

### Day to Day Use

### Import and Export

## Development Status Summary

Semantic Search is still in Beta as of this writing. Although an experimental module, Semantic Search has been used successfully in some projects:

### Successes

- “faceted search” is a promising concept for making search more navigable
- Drupal data integration can happen at a deep level with Nodes, CCK and Taxonomy without disrupting Drupal's core functionality or data.



- RDF stores and searches, though slower in performance and development time, provide improved
  - construction of search through SPARQL
  - import/export/syndication options
  - a consistent data structure for unstructured data
- Improved installation and configuration

### Current Challenges

- ease of installation
- reaching same level of functionality as current Drupal toolsets

### Development Roadmap Summary

Semantic Search is part of a strategy to bring advanced information search and sharing to smaller organizations using Drupal.

- **Version 1.0.5** is a stable basic release. The framework is set up, and performance characteristics are comparable to other Drupal solutions.
- **Version 1.1** is a bugfix and code stabilization release.
- **Version 1.5** maxes out the potential of search running on one website.
- **Version 2.0** is the full vision, integrated with the Semantic Web.

	<i>v. 1.0.5 Drupalcon Barcelona</i>	<i>v. 1.1</i>	<i>v. 1.5</i>	<i>v. 2.0 Semantic Web</i>
Installation /Configuration	Completely web based (nearly complete)			
Default Search	Web Configurable Default search		Data visualization	
API	Open API	Formalize API binary file handling improvements	logically grouped searches	P2P, OpenID
Import Export	CVS, RDF	RSS	MicroFormats, FOAF	Cross-site/Live
Documentation				
Demos	Demo site with large dataset,		API demos Even larger data sets	Cross-site API demos
RDF Stores	ARC speed issues improved caching	Custom Local for indexing	ARC upgrade, Sesame, Joseki, etc, Remote	Remote aggregation
Other	Drupal project page update, Code review by CA	Project Funded by Clients, bugfix release,	Verify IL8 and accessibility, Automated	



	<b>v. 1.0.5</b> <i>Drupalcon Barcelona</i>	<b>v. 1.1</b>	<b>v. 1.5</b>	<b>v. 2.0</b> <i>Semantic Web</i>
		media mover integration	tests, Drupal 6	
Community	Begin RFC process			

## Conclusion

## Appendix

### Revision History

Author	Date	Revision
Jonathan Hendler	11/09/06	First Draft
"	01/07/07	Rough Draft preparation for team meeting
	01/18/07	Reformatting, adding hours
	04/30/07	Name Change from NINA
	05/04/07	Changed to a user guide and re organized
	05/13/07	Added Installation and Configuration

### Appendix :: Tables

### Appendix : : Acknowledgments

Besides the amazing support, insight, and interest of the whole CivicActions team, a number of individuals have helped tremendously as the module took shape.

See how they rock: go to <http://civicaactions.com/people>

- Owen Barton  
Owen is one of CivicAction's Drupal gurus, and provided many key insights into the deeper workings of Drupal so the integration work would go smoothly.
- Gregory Heller  
Gregory has recently begun working with me to further the missions of "findability" of the data for many grassroots organizations
- Zoey Kroll  
Zoey is an Information Architect with CA and worked tirelessly to bring a less technical face to the technology for 2 of our clients. Some of her work on interfaces can be found in the default search interface.
- Fen Labalme  
Fen has the imagination and experience that helped dream up the technical possibilities – to get Drupal into the Semantic Web. Fen found me on a mailing list and brought me into CivicActions.
- Benjamin Nowack  
Benjamin wrote ARC, the RDF store used in Semantic Search. He provided many critical improvements and guided the integration of Semantic Search with an RDF Store. He is



## Semantic Search - User Guide

License: *Creative Commons Attribution-NonCommercial-Share Alike 3.0 License*

2006, 2007 - Jonathan Hendler, CivicActions, LLC

6/4/2007

Page 10

currently working with the W3C to continue the development of ARC.

- **Judilee Reed**  
Supported CivicActions work and vision with the LINC project, the National Information Network for Artists – also known as NINA. Seen at <http://lincnet.net> and <http://ahirc.org>
- **Dan Robinson**  
  
Dan helped guide Semantic Search through its earliest phases – fostering a unique vision to bring high technology to grassroots organizations.
- **Jenn Srammek**  
  
Jenn worked closely with all the above to keep this project rolling and her dedication and clarity and focus helped hold it together.

### ***Appendix : : Diagrams***

### ***Appendix : : Links***

- <http://drupal.org/project/nina>
- <http://www.lincnet.net/aa>
- <http://civicactions.com>

### ***Appendix : : External Resources***

### ***Appendix :: License and Copyright***

<http://creativecommons.org/licenses/by-nc-sa/3.0>

“You are free:

- **to Remix** — to adapt the work
- **to Share** — to copy, distribute and transmit the work

**Under the following conditions:**

- **Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial.** You may not use this work for commercial purposes.
- **Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.



- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.”

For the full text see <http://creativecommons.org/licenses/by-nc-sa/3.0/legalcode>