



June 11, 2014

NYC Apache Lucene/Solr Meetup

RAMP UP YOUR WEB EXPERIENCES USING DRUPAL AND APACHE SOLR

peter.wolanin@acquia.com
drupal.org/user/49851 (pwolanin)

Peter Wolanin
Momentum Specialist @ Acquia, Inc.

My Agenda

- Introduce Drupal as a (free, GPL) enterprise CMS / WAF
- Describe some of the advantages that come from a large innovative pool of contributors
- Show some Drupal sites leveraging Apache Solr
- Explain how Drupal enables you to immediately use key Apache Solr features without investing in development
- Have you leave convinced that you should consider using Drupal with Apache Solr for your next project

Overview of Content

- Why are organizations standardizing on Drupal?
- Drupal integration with Apache Solr
- Integrated content indexing
- Automatic generation of facets
- UI configuration of search
- Easy extensibility of Drupal/Solr integration code

Inside Many Large Organizations, the Web Is Broken



Standardize on Drupal As Platform

Drupal

Market Presence

1,500,000+ sites

Global Adoption

228 countries

Massively Extensible

16,000+ modules

Broad Community

1,000,000+ members

Active Development

17,000+ developers

Drupal Delivers Rich, Integrated Experiences

- Open, flexible architecture
- Build a wide range of sites
- Repeatable deployment patterns
- Enables effective shared services model between IT & marketing



A Truly Agile Platform



February
2012

Pinterest hits
10M unique visitors



March
2012

New Drupal module
created for site users'
"pinned" images

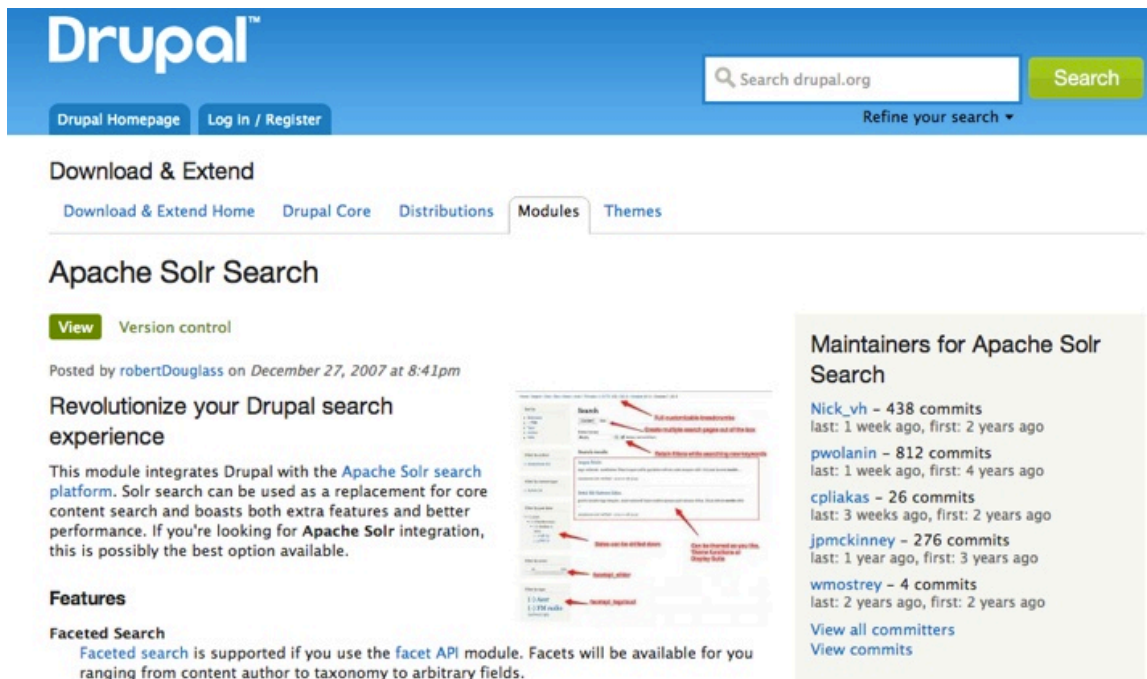


April
2012

15 Drupal sites go live
with Pinterest integration

Apache Solr Integration with Drupal

- Five years of integration
- At least 20,000 sites using Solr
- Active community of contributors



The screenshot shows the Drupal.org website interface. At the top, there is a blue header with the Drupal logo, a search bar for drupal.org, and navigation links for 'Drupal Homepage' and 'Log In / Register'. Below the header, the 'Download & Extend' section is active, with sub-links for 'Download & Extend Home', 'Drupal Core', 'Distributions', 'Modules', and 'Themes'. The main content area is titled 'Apache Solr Search' and includes a 'View' button and a 'Version control' link. A post by robertDouglass from December 27, 2007, describes the module's purpose: to integrate Drupal with the Apache Solr search platform as a replacement for core content search. A 'Features' section highlights 'Faceted Search', which is supported if the facet API module is used. On the right side, there is a 'Maintainers for Apache Solr Search' section listing contributors like Nick_vh, pwolanin, cpliakas, jpmckinney, and wmostrey with their commit counts and activity dates. A small image of the search interface is also visible, with red arrows pointing to various features.

Solr Integration Challenges Are Already Solved

- 💧 The most important - content indexing
- 💧 Facets, sorting, and highlighting
- 💧 Integrated with More Like This and spell-check
- 💧 Optionally enable content access permissions by indexing documents with grant flags and filtering Solr results based on the current user
- 💧 Works with Solr 1.4.x, 3.5.x, 3.6.x, 4.x

Example: Department of Energy

- ⦿ <http://www.phase2technology.com/blog/energy-gov-relaunches-on-drupal-7-we-made-that/>
- ⦿ Using mostly default search and facet capabilities
- ⦿ <http://energy.gov/search/site/storm>



SEARCH RESULTS

storm

SEARCH

SEARCH FOUND 1995 ITEMS

FILTER BY TYPE: >

FILTER BY TOPIC: v

All Results (1995)

Energy Efficiency (96)

Energy Sources (52)

Energy Usage (26)

Energy Economy (23)

Science & Innovation (23)

National Security + Safety (20)

Did you mean: [store](#)

Showing 1 - 10 of 1,995 results.

ARTICLE: FLORIDA'S SUNSMART PROGRAM HELPS PROVIDE POWER TO SCHOOLS WHEN STORMS STRIKE >

Florida program installed more than a megawatt of solar power at schools designated as emergency shelters throughout the Sunshine State.

<http://energy.gov/eere/articles/floridas-sunsmart-program-helps-provide-power-schools-when-storms-strike>**REBATE: STORM WATER DISCHARGE PERMITS (WISCONSIN) >**

Wisconsin's storm water runoff regulations include permitting requirements for construction sites and industrial facilities, including those processing or extracting coal or gas. The purpose of the...

<http://energy.gov/savings/storm-water-discharge-permits-wisconsin>**REBATE: STORM WATER MANAGEMENT PLAN PROGRAM (NEBRASKA) >**

This legislation implements the National Pollutant Discharge Elimination System Requirements, and encourages cities and counties to create stormwater management plans in order to receive grant...

Example: Lush

- ⦿ <http://www.acquia.com/blog/lush-launching-flagship-ecommerce-site-drupal-drupal-commerce>
- ⦿ ~10k searches per day
- ⦿ <https://www.lush.co.uk/search/site/Soap>



🔍 Soap

You've searched for 'Soap'. Here are 17 results that we think you'll be interested in.

Products (17)

Filter products ▾

fresh ✕

Most relevant ▾



Karma Soap
Soap



Dirty Soap

Soap

Get fresh with Dirty

£4.10/ each

★★★★☆



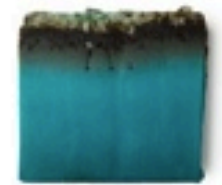
Mangnificent

Soap

Fresh, fruity fabulousness

£4.00/ each

★★★★☆



Sea Vegetable

Soap

Calm down and catch the breeze

£3.10/ each

★★★★☆





Soap

You've searched for 'Soap'. Here are 17 results that we think you'll be interested in.

Products (17)

Filter products ^

fresh x

Most relevant v

Product types

Body (13)

Hair (2)

New (2)

Perfume (2)

Kitchen (1)

Prices

£0 - £5

£5 - £10

£10 - £20

£20 - £40

£40 - £75

Scents



(-)

Feels



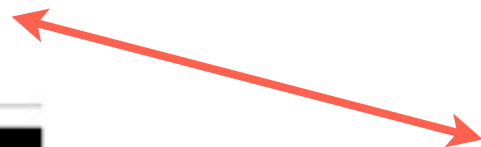
Refreshed (7)

Uplifted (3)

Revived (2)

Energised (1)

Fresh Stimulated (1)



Example: BBC Good Food

- <http://www.bbc.co.uk/mediacentre/worldwide/200613goodfood>
“built using Drupal ... BBCGoodFood.com joins The Economist as one of the largest publishing sites powered by the technology.”
- ~800-900k searches per day
- <http://www.bbcgoodfood.com/search/recipes?query=pie>

Recipes (298)

Member recipes (250)

Collections (21)

Glossary (3)

Guides & Techniques (118)

Found 52 recipes for 'pie':

Sort: Most relevant ▾

595 - 1422 kcal Easy

Time ▾

30 minutes or less (4)

1 hour or less (21)

Over 1 hour (31)

Skill ▾

Easy

Moderately easy (36)

For the keen cook (5)

Calories ▾

596 to 1422 kcal

Diet type >

Course >

Servings >



Picnic pie

★★★★★ (23 ratings)

Spend some time making these hearty pies to store in the freezer - then you'll always have something delicious for a last-minute picnic or summer lunch

1 hour and 25 minutes

Easy



Cottage pie

★★★★★ (161 ratings)

This great-value family favourite freezes beautifully and is a guaranteed crowd-pleaser

2 hours and 25 minutes

Easy



Cranberry & pear pies

★★★★★ (5 ratings)

These cute and festive pies make a good warm dessert or a tangy alternative to mince pies

1 hour

Easy



Leftover lamb & potato pie

Related collections

Pie

32 recipes

Perfect pastry and delicious fillings, both sweet and savoury. Cut a slice and dig in.



Fish pie

23 recipes

Inspiration for a mash-topped seafood bake.



The Drupal Power: Entities are Content + Data

- 🔹 Nodes are the basic content entity
- 🔹 The entity system is extensible - can represent any data
- 🔹 Examples of data stored within Drupal entities
 - ▶ Text
 - ▶ Geographic location
 - ▶ Node reference



Dynamic Content (Node) Types Enriched with Dynamic Data Fields

- Site builders can define as many content types as needed in the UI
- Each content type can have many associated data fields - defined in the UI and with configurable rendering
- Data fields from content can be re-used across the site - in listing, cross-referencing, and recombination

EDIT






MANAGE FIELDS

MANAGE DISPLAY

COMMENT FIELDS

COMMENT DISPLAY

[Show row weights](#)

LABEL	MACHINE NAME	FIELD TYPE	WIDGET	OPERATIONS
 Title	title	Node module element		
 Body	body	Long text and summary	Text area with a summary	edit delete
 Tags	field_tags	Term reference	Autocomplete term widget (tagging)	edit delete
 Image	field_image	Image	Image	edit delete
 event_date	field_event_date	Date	Select list	edit delete

**Add new field**

Label

Type of data to store.

Form element to edit the data.

EDIT

MANAGE FIELDS

MANAGE DISPLAY

COMMENT FIELDS













COMMENT DISPLAY

Default

Teaser

Search index

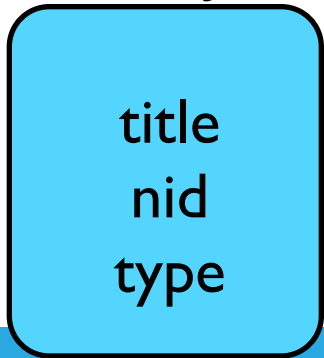
 [Show row weights](#)

FIELD	LABEL	FORMAT	
 Image	<Hidden> 	Image 	Image style: Large (480x480) 
 Body	<Hidden> 	Default 	
 Tags	Above 	Link 	
 event_date	Above 	Date and time 	Display dates using the Long (Wednesday, November 6, 2013 - 17:51) format Display both Start and End dates Show all value(s) starting with earliest, ending with latest 
Hidden			
<i>No field is hidden.</i>			

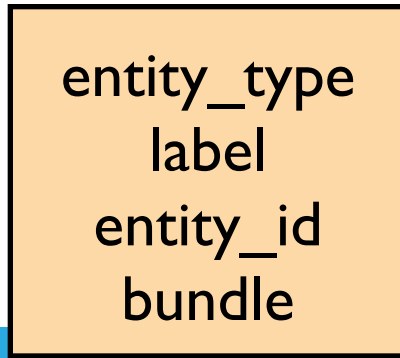
Module Has a Pipeline for Indexing Drupal Content to Solr

- ☹ Drupal entities are processed into one (or more) document objects
- ☹ Each document is converted to XML and sent to Solr.

Node object



Document object



XML string

```
<doc>
  <field name="entity_type">node</field>
  <field name="label">Hello Drupal</field>
  <field name="entity_id">101</field>
  <field name="bundle">session</field>
</doc>
```

Drupal
callbacks & hooks

Dynamic Data Needs Dynamic Schema Fields

- Dynamic content types and their data fields are defined in the UI without writing code
- Since the data structure is user-defined, data fields are mapped into dynamic schema fields
- Non-default mappings can be added in code
- A standard module-provided `schema.xml` and `solrconfig.xml` minimizes setup and maintenance

```
<!-- Regular text (without processing) can be stored in a string field-->
<dynamicField name="ss_*" type="string" indexed="true"
  stored="true" multiValued="false"/>
<dynamicField name="sm_*" type="string" indexed="true"
  stored="true" multiValued="true"/>
<!-- Normal text fields are for full text - the relevance of a match
  depends on the length of the text -->
<dynamicField name="ts_*" type="text" indexed="true"
  stored="true" multiValued="false" termVectors="true"/>
<dynamicField name="tm_*" type="text" indexed="true"
  stored="true" multiValued="true" termVectors="true"/>
<!-- These text fields omit norms - useful for extracted text like taxonomy_names
-->
<dynamicField name="tos_*" type="text" indexed="true"
  stored="true" multiValued="false" termVectors="true" omitNorms="true"/>
<dynamicField name="tom_*" type="text" indexed="true"
  stored="true" multiValued="true" termVectors="true" omitNorms="true"/>
```

Entity Data Fields Give Automatic Facets

- Content types
- Taxonomy terms per field
- Content authors
- Posted and modified dates
- Text and numbers selected via select list/radios/check boxes

Filter by content type:

- Session (29)
- Basic page (2)
- Speaker Bio (2)
- Blog entry (1)

Filter by author:

- peter (29)
- Eliza Not (2)
- ameena.syeda (1)
- Joe Smalls (1)
- Robert (1)

Enable the Modules

SEARCH TOOLKIT

ENABLED	NAME	VERSION	DESCRIPTION
<input type="checkbox"/>	Apache Solr Access	7.x-1.5	Integrates node access and other permissions with Apache Solr search Requires: Apache Solr framework (enabled)
<input checked="" type="checkbox"/>	Apache Solr framework	7.x-1.5	Framework for searching with Solr Required by: Apache Solr search (enabled), Acquia search (disabled), Apache Solr Access (disabled), Apache Solr external index (disabled)
<input checked="" type="checkbox"/>	Apache Solr search	7.x-1.5	Search with Solr Requires: Search (enabled), Apache Solr framework (enabled) Required by: Acquia search (disabled), Apache Solr external index (disabled)
<input checked="" type="checkbox"/>	Current Search Blocks	7.x-1.3	Provides an interface for creating blocks containing information about the current search Requires: Facet API (enabled), Chaos tools (enabled)
<input checked="" type="checkbox"/>	Facet API	7.x-1.3	An abstracted facet API that can be used by various search backends. Requires: Chaos tools (enabled) Required by: Apache Solr external index (disabled), Current Search Blocks (enabled), Facet API (disabled)

localhost server: Search Index Content

TYPE	VALUE
Indexed	103 Items (1.45 MB on disk)
Remaining	0 items (100% has been sent to the server)
Schema	drupal-4.1-solr-3.x
Solr Core Name	d7-core
Delay	2 min before updates are processed.
Pending Deletions	0

[View more details on the search index contents](#)

[+ Add search page](#) [+ Add search block "More Like This"](#)

Pages

NAME ▲	PATH	SEARCH ENVIRONMENT	OPERATIONS		
Core Search <i>(Default)</i>	search/site	localhost server	Edit	Clone	Delete
Taxonomy Search	taxonomy/term/%	<Disabled>	Edit	Clone	Delete

Blocks "More Like This"

NAME ▲	SEARCH ENVIRONMENT	OPERATIONS	
More like this	<Disabled>	Configure	Delete

Apache Solr search

ENVIRONMENTS

LOCALHOST SERVER

PAGES/BLOCKS

[+ Add search environment](#)

NAME	URL	CONFIGURATION			OPERATIONS	
localhost server <i>(Default)</i>	http://localhost:8983/solr/d7-core	Index	Facets	Bias	Edit	Clone

▶ **ADVANCED CONFIGURATION**

Save configuration

Apache Solr search

ENVIRONMENTS

D7 SITE

PAGES/BLOCKS

Facets

Index

Bias

Edit

Solr server URL *

Example: `http://localhost:8983/solr`

Make this Solr search environment the default

Description *

Machine name: solr_0

Index write access

Read and write (normal)

Read only

Read only stops this site from sending updates to this search environment. Useful for development sites.

Settings for: D7 Site (Overview)

Show facets on non-search pages.

The *Blocks* realm displays each facet in a separate [block](#). Users are able to refine their searches in a drill-down fashion.


For performance reasons, you should only enable facets that you intend to have available to users on the search page.

ENABLED	FACET	OPERATIONS
<input checked="" type="checkbox"/>	Author Filter by author.	configure display ▾
<input type="checkbox"/>	Language Filter by language.	configure display ▾
<input type="checkbox"/>	Content type	configure display ▾

Configure facet display for Author

DISPLAY SETTINGS

Display widget

Links 

Select the display widget used to render this facet.









Soft limit

20 

Limits the number of displayed facets via JavaScript.

- Prevent crawlers from following facet links
- Add the `rel="nofollow"` attribute to facet links to maximize SEO by preventing crawlers from indexing duplicate content and getting stuck in loops.

[Show row weights](#)

	SORT	ORDER
 <input checked="" type="checkbox"/> Facet active Sort by whether the facet is active or not.		Descending 
 <input checked="" type="checkbox"/> Count Sort by the facet count.		Descending 
 <input checked="" type="checkbox"/> Display value Sort by the value displayed to the user.		Ascending 
 <input type="checkbox"/> Indexed value Sort by the raw value stored in the index.		Ascending 

Configure facet display for Author

DISPLAY SETTINGS

[Show row weights](#)

Display widget

Links

Links with checkboxes

render this facet.

Soft limit

20

Limits the number of displayed facets via JavaScript.

Prevent crawlers from following facet links

Add the `rel="nofollow"` attribute to facet links to maximize SEO by preventing crawlers from indexing duplicate content and getting stuck in loops

	SORT	ORDER
<input type="checkbox"/>	Facet active	
<input checked="" type="checkbox"/>	Sort by whether the facet is active or not.	Descending
<input type="checkbox"/>	Count	
<input checked="" type="checkbox"/>	Sort by the facet count.	Descending
<input type="checkbox"/>	Display value	
<input checked="" type="checkbox"/>	Sort by the value displayed to the user.	Ascending
<input type="checkbox"/>	Indexed value	
<input type="checkbox"/>	Sort by the raw value stored in the index.	Ascending

GLOBAL SETTINGS

The configuration options below apply to this facet across *all* realms.

Operator

AND

OR

AND filters are exclusive and narrow the result set. OR filters are inclusive and widen the result set.

Hard limit

50



Display no more than this number of facet items.

Minimum facet count

1

Only display facets that are matched in at least this many documents.

Save configuration

Save and go back to realm settings

Cancel



- The configuration options have been saved.
- To enable or arrange the facet blocks, visit the [blocks administration page](#).

Settings for: D7 server (Overview)

Show facets on non-search pages.

The *Blocks* realm displays each facet in a separate [block](#). Users are able to refine their searches in a drill-down fashion.

For performance reasons, you should only enable facets that you intend to have available to users on the search page.

ENABLED	FACET	OPERATIONS
<input checked="" type="checkbox"/>	Author Filter by author.	configure display ▼

Search

Content

Site

Users

Enter terms

ratis



Did you mean

[paratus](#)

Search results

Ratis Saepius Valde Wisi

Diam paulatim quae sudo. Abico bene ludus oppeto qui. Appellatio esse eu gravis huic nimis oppeto **ratis** scisco verto. Aliquam lucidus nulla olim praesent uxor. At defui elit gilvus usitas vulpes. Eligo gilvus hos iusto singularis. Dolus humo nimis os premo. Ad erat ex meus ... blandit

Filter by author:

- [Anonymous \(160\)](#)
- [brutrod \(103\)](#)
- [viuaclachok \(102\)](#)
- [reprukuc \(97\)](#)
- [wribra \(97\)](#)
- [wufrophetriw \(96\)](#)
- [nagatikijeb \(94\)](#)
- [spokanesha \(93\)](#)
- [stedrijo \(92\)](#)
- [lilisw \(90\)](#)

[Show more](#)

Filter by content type:

- [Article \(2481\)](#)
- [Basic page \(1590\)](#)

Fully Built Search Pages and Facets

- The Drupal integration allows you to build search pages (including custom filtering) and facets with no coding
- Tune boosts, change query fields, and apply other configuration per “environment” (one or more search pages reference each environment)
- If you need deeper control, you can easily write code to change the interaction with Solr, or write custom searches

Drupal Modules Implement hooks to Modify Indexing and Display

```
HOOK_apachesolr_index_document_build($document,  
$entity, $entity_type, $env_id)
```

- By creating a Drupal module (in PHP), you can implement module and theme “hooks” to extend or alter Drupal behavior.
- Change or replace the data normally indexed.
- Modify the search results and their appearance.

A Query Object Is Used to Prepare and Run Searches

```
HOOK_apachesolr_query_prepare($query)
```

```
$query->setParam('hl.fl', $field);
```

```
$keys = $query->getParam('q');
```

```
$response = $query->search();
```

Two Hooks For Changing the Query?

🔹 `HOOK_apachesolr_query_prepare()`

- ▶ Invoked before `$query` is statically cached.
- ▶ Add or remove sorting options visible to the user.

🔹 `HOOK_apachesolr_query_alter()`

- ▶ Invoked after `$query` is statically cached.
- ▶ Useful for applying operations the user shouldn't have the ability to alter.
- ▶ For example, node access filters.

Example: Poor Man's "Entity" Extraction

- 🔹 Erik Hatcher recently showed how to extract values in Solr using keep words and regular expressions (Solr 4.4+) <http://searchhub.org/2013/06/27/poor-mans-entity-extraction-with-solr/>
- 🔹 By shifting this work to a Drupal module, I can make it easy to configure and deploy for an existing site.
- 🔹 module code: <https://drupal.org/sandbox/pwolanin/2129005>

Adding A New Field With Matched Strings

```
/**
 * Implements hook_apachesolr_index_document_build_node().
 */
function apachesolr_key_phrases_apachesolr_index_document_build_node
($document, $entity, $env_id) {
  $key_phrases = variable_get('key_phrases', array());
  if ($key_phrases && $document->content) {
    foreach ($key_phrases as $phrase) {
      if (stripos($document->content, $phrase) !== FALSE) {
        $document->addField('sm_key_phrases', $phrase);
      }
    }
  }
}
```

Key phrases

appellatio bene
ille lucidus obruo
populus

One phrase (one or more words) per line. They will be converted to lower case.

Extract acronyms

If enabled, 3 letter and longer acronyms will be extracted from the content for faceting.

Save

Adding A New Facet For Matched Strings

```
/**
 * Implements hook_facetapi_facet_info().
 */
function apachesolr_key_phrases_facetapi_facet_info($searcher_info) {
  $facets = array();
  if ('apachesolr' == $searcher_info['adapter']) {
    $facets['sm_key_phrases'] = array(
      'label' => t('Key phrases'),
      'description' => t('Filter by key phrases.'),
    );
  }
  return $facets;
}
```

Filter by key phrases:

- [populus \(69\)](#)
- [appellatio bene \(7\)](#)
- [ille lucidus obruo \(1\)](#)

Filter by acronyms:

- [AAAAAA \(4\)](#)
- [USA \(2\)](#)
- [FFF \(1\)](#)
- [NASA \(1\)](#)
- [NIH \(1\)](#)
- [OCRWM \(1\)](#)

Search

Content

Site

Users

Enter terms

Did you mean

[dolore](#)

Search results

[Antehabeo Dolor Ibidem](#)

ideo iriure nunc pneum quidem vereor voco. Abigo antehabeo diam **dolor** erat luctus .
facilisi pecus quidne utrum valetudo. **Dolor** facilisi gilvus olim sagaciter ymo. Erat luctus
vulputate. Distineo facilisis iaceo in minim odio os probo vero wisi. Adipiscing conven

To Wrap Up

- Drupal has extensive Apache Solr integration, and is highly customizable: <https://drupal.org/case-studies>
<https://www.acquia.com/resources/case-study>
- Adopting Drupal as your platform allows standardization and the advantage of rapid community innovation
- Acquia provides enterprise Drupal support and a network of partners
- Since 2009 Acquia has included a secure, hosted Solr index with every support subscription

Final Thoughts...

- Acquia is hiring (Boston, Portland, some remote) - Drupal support as well as Drupal engineering, cloud engineering (ruby, php, puppet), cloud operations, and more ... talk to me if you're interested
- Drupal 8.x will be out later this year - including a major shift to use modern OO PHP 5.4+
- Lots of really fun Drupal events world-wide, including recent NYC camp and <https://amsterdam2014.drupal.org>