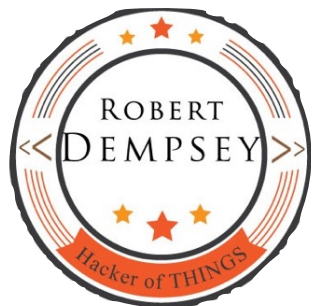
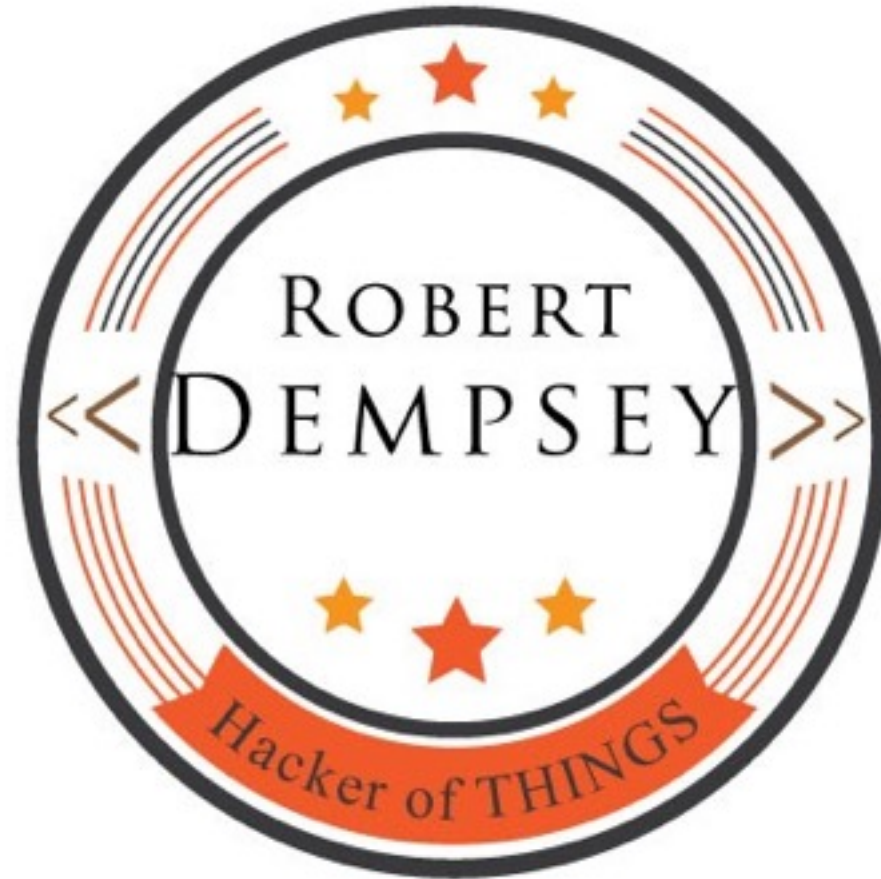


# Using PyParsing For Web Scraping, Application Control and Data Wrangling

Robert Dempsey  
[robertwdempsey.com](http://robertwdempsey.com)



[robertwdempsey.com](http://robertwdempsey.com)



[rdempsey](#)



[robertwdempsey](#)



[rdempsey](#)



<https://github.com/rdempsey/pyparsing-dcpython>

# PyParsing

An alternative approach to creating and executing simple grammars

- Extract data from a text file or web page
- Use in applications to allow for user-defined commands or search strings
- Regex makes your brain hurt and you want an alternative

# Install It

Latest version: 2.0.1

- `easy_install pyparsing`
- `pip install pyparsing`

# Basic PyParsing App

- Import names from PyParsing
- Define the grammar
- Use the grammar
- Process the results

# Elements of a Grammar

- Word
- Literal
- OneOrMore
- oneOf
- alphas

Hello, World!



# Parsing User Input

# Web Scraping



Web Scrapping



Scrapy

[scrapy.org](https://scrapy.org)

# Web Scraping With Scrapy

# Built Like This

- Specify data (Item)
- Create the spider (Spider)
- Find your data elements (XPath Selectors)
- Extract the data
- Process the data

# Data Acquisition and Wrangling with Python Workshop



## Overview

Eighty percent or more of the time spent on data science projects is spent acquiring data, cleaning it, and preparing it for analysis. That data can come from a variety of sources, including APIs or individual web pages. However, not all data is created equal. Once we have automated its acquisition, much of it requires lengthy cleaning and formatting before it can be used. In this course you will learn how to obtain, clean, and mashup data in preparation for analysis.

Our focus will be on achieving two goals:

1. Understanding more about your customers from their social profiles.
2. Pulling data off the web (screen scraping) for market research and getting it into a database.

## Date & Time:

Saturday, May 9th 2015 9am-5pm

## Location:

640 Massachusetts Avenue NW  
Washington, DC 20001

**Regular Price: \$300**

**Early Bird Price: \$250**

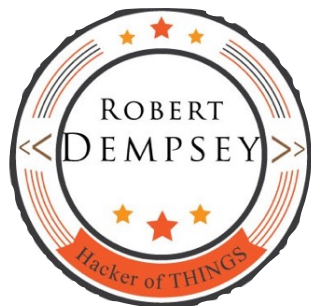
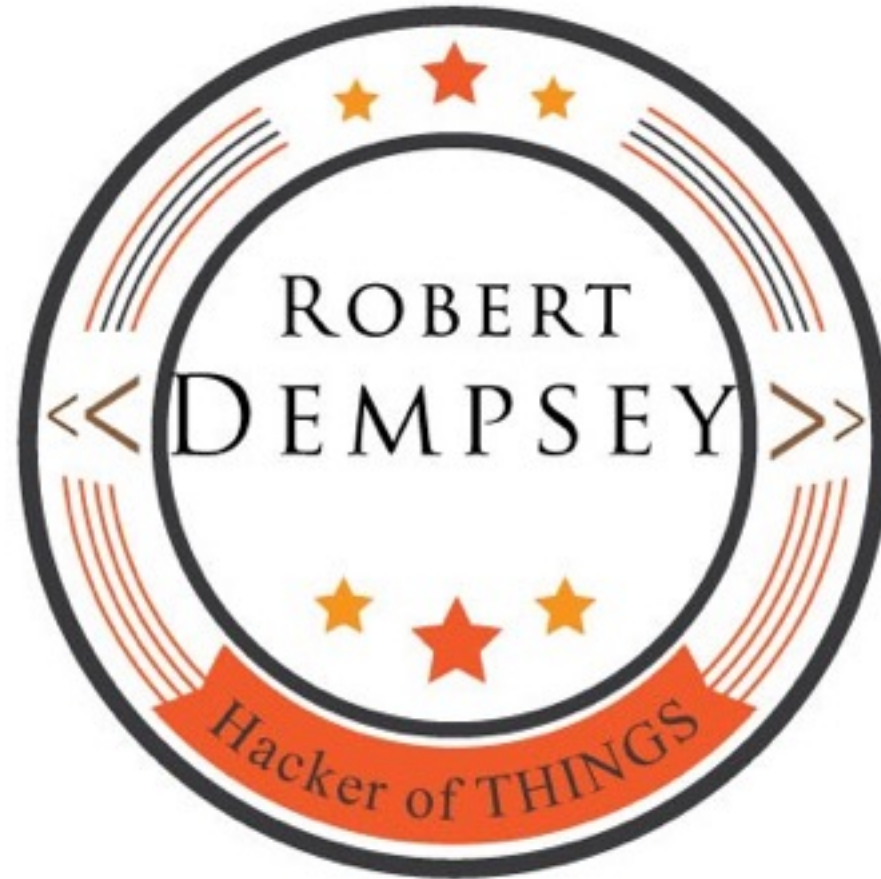
(expires 4/25/2015)

**REGISTER**

## Prerequisites

- Basic programming knowledge of Python
- Anaconda with Python 3.4 (for Part 1)
- Anaconda with Python 2.7 (for Part 2)
- MySQL 5.7
- iPython Notebook (included with the Anaconda install)

<http://robertwdempsey.com/ddl-datawrangling>



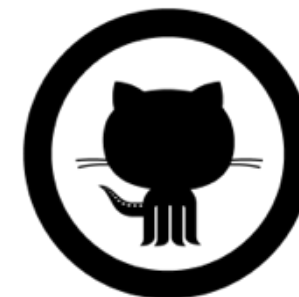
[robertwdempsey.com](http://robertwdempsey.com)



[rdempsey](#)



[robertwdempsey](#)



[rdempsey](#)