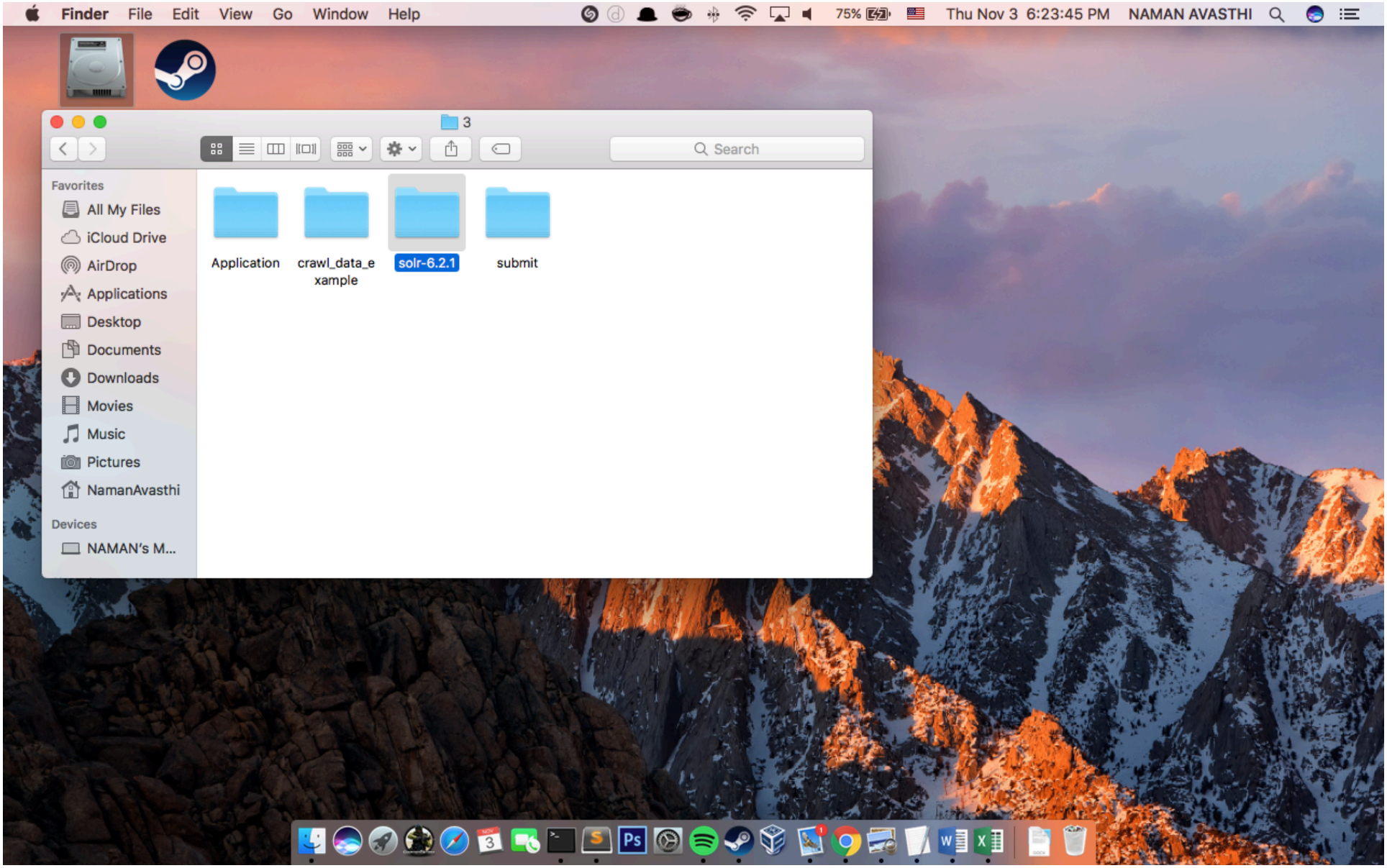
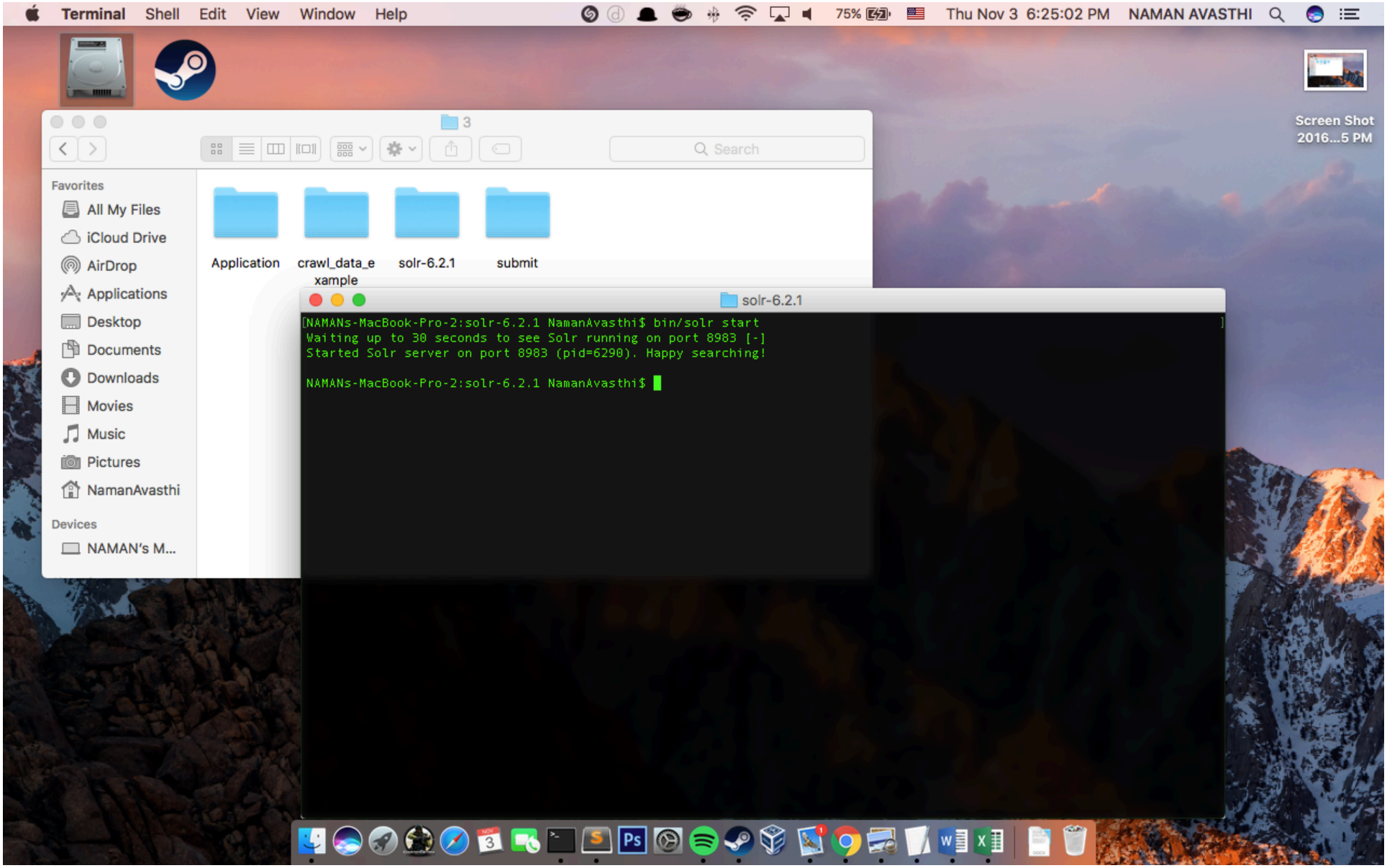
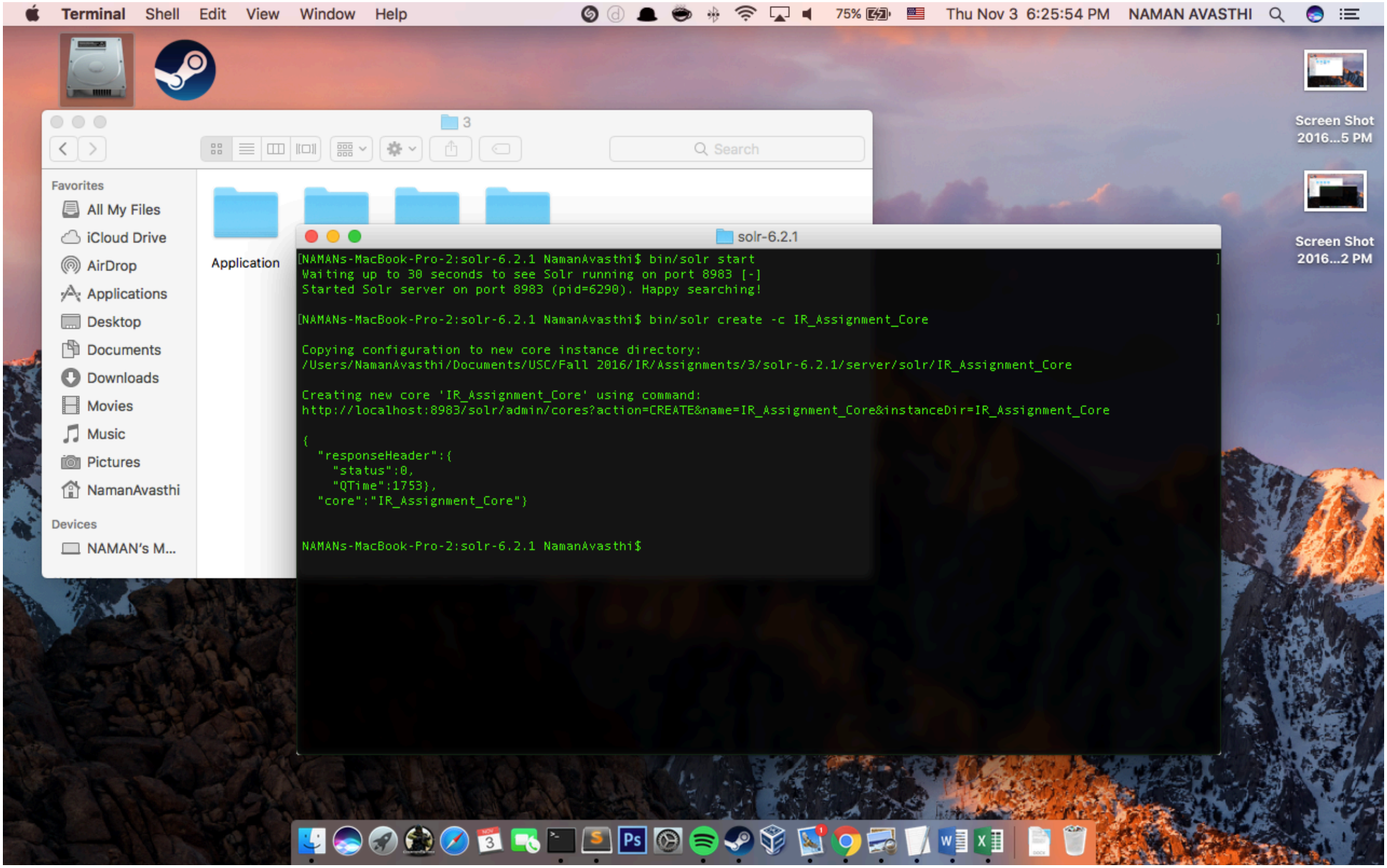
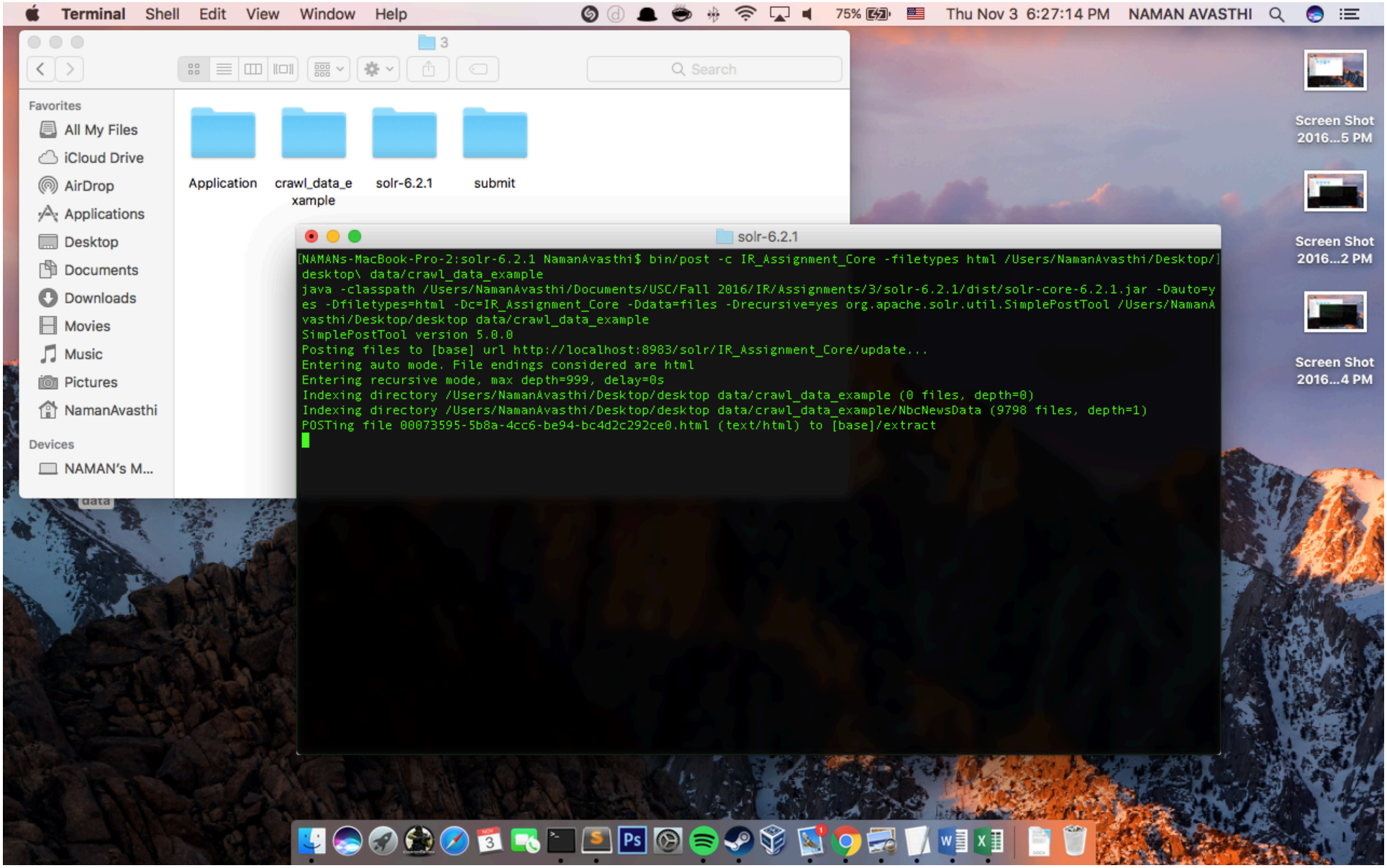

Finder File Edit View Go Window Help
75%
Thu Nov 3 6:23:45 PM
NAMAN AVASTHI
3
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
Application
crawl_data_example
solr-6.2.1
submit


Terminal Shell Edit View Window Help
75%
Thu Nov 3 6:25:02 PM
NAMAN AVASTHI
Screen Shot 2016...5 PM
3
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
Application
crawl_data_example
solr-6.2.1
submit
solr-6.2.1
NAMANs-MacBook-Pro-2:solr-6.2.1 NamanAvasthi$ bin/solr start
Waiting up to 30 seconds to see Solr running on port 8983 [-]
Started Solr server on port 8983 (pid=6290). Happy searching!
NAMANs-MacBook-Pro-2:solr-6.2.1 NamanAvasthi$


Terminal Shell Edit View Window Help
75%
Thu Nov 3 6:25:54 PM
NAMAN AVASTHI
Screen Shot 2016...5 PM
Screen Shot 2016...2 PM
3
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
Application
solr-6.2.1
NAMANs-MacBook-Pro-2:solr-6.2.1 NamanAvasthi$ bin/solr start
Waiting up to 30 seconds to see Solr running on port 8983 [-]
Started Solr server on port 8983 (pid=6290). Happy searching!
NAMANs-MacBook-Pro-2:solr-6.2.1 NamanAvasthi$ bin/solr create -c IR_Assignment_Core
Copying configuration to new core instance directory:
/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignments/3/solr-6.2.1/server/solr/IR_Assignment_Core
Creating new core 'IR_Assignment_Core' using command:
http://localhost:8983/solr/admin/cores?action=CREATE&name=IR_Assignment_Core&instanceDir=IR_Assignment_Core
{
"responseHeader":{
"status":0,
"QTime":1753},
"core":"IR_Assignment_Core"}
NAMANs-MacBook-Pro-2:solr-6.2.1 NamanAvasthi$


Terminal Shell Edit View Window Help
75%
Thu Nov 3 6:27:14 PM
NAMAN AVASTHI
3
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
Application
crawl_data_example
solr-6.2.1
submit
solr-6.2.1
[NAMANs-MacBook-Pro-2:solr-6.2.1 NamanAvasthi$ bin/post -c IR_Assignment_Core -filetypes html /Users/NamanAvasthi/Desktop/]
desktop\ data/crawl_data_example
java -classpath /Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignments/3/solr-6.2.1/dist/solr-core-6.2.1.jar -Dauto=y
es -Dfiletypes=html -Dc=IR_Assignment_Core -Ddata=files -Drecursive=yes org.apache.solr.util.SimplePostTool /Users/NamanA
vasthi/Desktop/desktop data/crawl_data_example
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/IR_Assignment_Core/update...
Entering auto mode. File endings considered are html
Entering recursive mode, max depth=999, delay=0s
Indexing directory /Users/NamanAvasthi/Desktop/desktop data/crawl_data_example (0 files, depth=0)
Indexing directory /Users/NamanAvasthi/Desktop/desktop data/crawl_data_example/NbcNewsData (9798 files, depth=1)
POSTing file 00073595-5b8a-4cc6-be94-bc4d2c292ce0.html (text/html) to [base]/extract
Screen Shot 2016...5 PM
Screen Shot 2016...2 PM
Screen Shot 2016...4 PM

Solr Admin | PHP Solr Form | PHP Solr Form | Homework3.pdf | CSCI 572 Home Pa | GradingGuidelines | Solr Admin

localhost:8983/solr/#/

Solr

- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump

Core Selector

Instance

Start	3 minutes ago

Versions

solr-spec	6.2.1
solr-impl	6.2.1 43ab70147eb494324a1410f7a9f16a896a59bc6f - shalin - 2016-09-1
lucene-spec	6.2.1
lucene-impl	6.2.1 43ab70147eb494324a1410f7a9f16a896a59bc6f - shalin - 2016-09-1

JVM

Runtime	Oracle Corporation Java HotSpot(TM) 64-Bit Server VM 1.8.0_101 25.101-b13
Processors	4
Args	-DSTOP.KEY=solrrocks -DSTOP.PORT=7983 -Djetty.home=/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignm -Djetty.port=8983 -Dsolr.install.dir=/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assig -Dsolr.solr.home=/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assig -Duser.timezone=UTC -XX:+CMSParallelRemarkEnabled -XX:+CMSScavengeBeforeRemark -XX:+ParallelRefProcEnabled -XX:+PrintGCApplicationStoppedTime -XX:+PrintGCDateStamps -XX:+PrintGCDetails -XX:+PrintGCTimeStamps -XX:+PrintHeapAtGC -XX:+PrintTenuringDistribution -XX:+UseCMSInitiatingOccupancyOnly -XX:+UseConcMarkSweepGC

System 5.98 4.21 4.08

Physical Memory 99.6%

7.97 GB

8.00 GB

File Descriptor Count 1.7%

176

10240

JVM-Memory 27.2%

133.48 MB

490.69 MB

490.69 MB

Solr Admin | PHP Solr Form | PHP Solr Form | Homework3.pdf | CSCI 572 Home Pa | GradingGuidelines | Solr Admin

localhost:8983/solr/#/IR_Assignment3_Core

Use original UI

Solr

- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump

IR_Assignment...

- Overview
- Analysis
- Dataimport
- Documents
- Files
- Ping
- Plugins / Stats
- Query
- Replication
- Schema
- Segments info

Statistics

Last Modified:	16 days ago
Num Docs:	15976
Max Doc:	15976
Heap Memory Usage:	-1
Deleted Docs:	0
Version:	610
Segment Count:	18
Optimized:	✔
Current:	✔

Instance

CWD:	/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignments/3/solr-6.2.1/server
Instance:	/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignments/3/solr-6.2.1/server/solr/IR_Assignment3_Core
Data:	/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignments/3/solr-6.2.1/server/solr/IR_Assignment3_Core/data
Index:	/Users/NamanAvasthi/Documents/USC/Fall 2016/IR/Assignments/3/solr-6.2.1/server/solr/IR_Assignment3_Core/data/index
Impl:	org.apache.solr.core.NRTCachingDirectoryFactory

Replication (Master)

	Version	Gen	Size
Master (Searching)	1476827781682	97	590.95 MB
Master (Replicable)	-	-	-

Healthcheck

Ping request handler is not configured with a healthcheck file.

Admin Extra

Documentation | Issue Tracker | IRC Channel | Community forum | Solr Query Syntax

Solr Admin | PHP Solr Form | PHP Solr Form | Homework3.pdf | CSCI 572 Home Pa | GradingGuidelines | Solr Admin

localhost:8983/solr/#/IR_Assignment3_Core/query

Solr

- Dashboard
- Logging
- Core Admin
- Java Properties
- Thread Dump

IR_Assignment...

- Overview
- Analysis
- Dataimport
- Documents
- Files
- Ping
- Plugins / Stats
- Query
- Replication
- Schema
- Segments info

/select

common

q

:

fq

sort

start, rows

0 | 10

fl

df

Raw Query Parameters

key1=val1&key2=val2

wt

json

- [x] indent
- [ ] debugQuery
- [ ] dismax
- [ ] edismax
- [ ] hl
- [ ] facet
- [ ] spatial
- [ ] spellcheck

Execute Query


```
{
  "responseHeader":{
    "status":0,
    "QTime":87,
    "params":{
      "q":"*:*",
      "indent":"on",
      "wt":"json",
      "_":"1478222889213"}},
  "response":{"numFound":15976,"start":0,"docs":[
      {
        "id":"/Users/NamanAvasthi/Desktop/crawl_data_example/NbcNewsData/0efba320-bd09-4f7b-a636-2f812cdf694b.ht
        "og_image":["http://media4.s-nbcnews.com/j/newscms/2016_38/1722861/160925-baltimore-0403_34d9d9bcbe971e9
        "apple_itunes_app":["app-id=319740707, app-argument=http://www.nbcnews.com/news/us-news/baltimore-shooti
        "twitter_card":["summary_large_image"],
        "stream_content_type":["text/html"],
        "og_site_name":["NBC News"],
        "dc_date_issued":["2016-09-25T00:00:00Z"],
        "description":["Three gunmen shot and wounded eight people — including a 3-year-old girl — on an east Ba
        "title":["Baltimore Shooting: Gunmen Wound Eight, Including 3-Year-Old Girl - NBC News"],
        "og_description":["Three gunmen shot and wounded eight people — including a 3-year-old girl — on an east
        "twitter_creator":["NBCNews"],
        "twitter_image":["http://media4.s-nbcnews.com/j/newscms/2016_38/1722861/160925-baltimore-0403_34d9d9bcbe
        "twitter_site":["NBCNews"],
        "news_keywords":["News,U.S. news,U.S. news"],
        "dc_title":["Baltimore Shooting: Gunmen Wound Eight, Including 3-Year-Old Girl - NBC News"],
        "content_encoding":["UTF-8"],
        "content_type":["text/html; charset=UTF-8"],
        "stream_size":[99428],
        "x_parsed_by":["org.apache.tika.parser.DefaultParser",
          "org.apache.tika.parser.html.HtmlParser"],
        "og_type":["article"],
        "twitter_title":["Baltimore Shooting: Gunmen Wound Eight, Including 3-Year-Old Girl - NBC News"],
        "google_site_verification":["dMf4cxtcxKlWchKvPlULwDZi8XpI9-ARqe8j2mGmi4Y",
          "K8bowvEbzWfPQ76x6f9SCxMVwicVXrhiClRg80ak8ZA",
          "G4SGIUDh2TcgDVNjuleQ-Humwb3JAZaQD9Vg42Stfn8"],
        "og_title":["Eight people, including toddler, wounded in Baltimore shooting"],
```

Sublime Text 2 File Edit Selection Find View Goto Tools Project Window Help 75% Thu Nov 3 6:29:21 PM NAMAN AVASTHI

solrconfig.xml UNREGISTERED

managed-schema | solrconfig.xml

```xml
    <!-- SearchHandler

         http://wiki.apache.org/solr/SearchHandler

         For processing Search Queries, the primary Request Handler
         provided with Solr is "SearchHandler" It delegates to a sequent
         of SearchComponents (see below) and supports distributed
         queries across multiple shards
      -->
    <requestHandler name="/select" class="solr.SearchHandler">
      <!-- default values for query parameters can be specified, these
           will be overridden by parameters in the request
        -->
      <lst name="defaults">
        <str name="echoParams">explicit</str>
        <int name="rows">10</int>
        <str name="df">_text_</str>
        <str name="q.op">AND</str>
      </lst>


      <!-- In addition to defaults, "appends" params can be specified
           to identify values which should be appended to the list of
           multi-val params from the query (or the existing "defaults").
        -->
      <!-- In this example, the param "fq=instock:true" would be appended to
           any query time fq params the user may specify, as a mechanism for
           partitioning the index, independent of any user selected filtering
           that may also be desired (perhaps as a result of faceted searching).

           NOTE: there is *absolutely* nothing a client can do to prevent these
           "appends" values from being used, so don't use this mechanism
           unless you are sure you always want it.
        -->
```

Line 763, Column 33 Spaces: 2 XML

IR_trial.php

```php
<?php
    //  make    sure    browsers    see this    page    as  utf-8   encoded HTML
    header('Content-Type:   text/html;  charset=utf-8');
    error_reporting(0);
    $limit  = 10;
    $query  = isset($_REQUEST['q']) ?   $_REQUEST['q']  :   false;
    $results    = false;

$file1 = fopen("map_nbcnews.csv", "r");
$file2 = fopen("map_wsj.csv", "r");
$filetourl = array();
while(!feof($file1))
{
    $line = fgets($file1);
        $tokens = explode(",", $line);
    $filetourl[$tokens[0]] = $tokens[1];
}
while(!feof($file2))
{
    $line = fgets($file2);
    $tokens = explode(",", $line);
    $filetourl[$tokens[0]] = $tokens[1];
}



                // 'q'=>'&facet.field:og_url&facet.numTerms:true&facet.minCount:10'




    if ($query)
    {
        //  The Apache  Solr    Client  library should  be  on  the include path
        //  which   is  usually most    easily  accomplished    by  placing in  the
        //  same    directory   as  this    script  (   .   or  current directory   is  a   default
        //  php include path    entry   in  the php.ini)

        // $query = $query.'&facet.field=og_url&facet.minCount=10';
```

Line 249, Column 103 Spaces: 4 PHP

IR_trial.php

```php


if ($query)
{
    // The Apache Solr Client library should be on the include path
    // which is usually most easily accomplished by placing in the
    // same directory as this script ( . or current directory is a default
    // php include path entry in the php.ini)

    // $query = $query.'&facet.field=og_url&facet.minCount=10';

    // &facet.field=og_url&facet.minCount=10

    require_once('solr-php-client-master/Apache/Solr/Service.php');

    // create a new solr service instance - host, port, and corename
    // path (all defaults in this example)

    $solr = new Apache_Solr_Service('localhost', 8983, '/solr/IR_Assignment3_Core/');

    // if magic quotes is enabled then stripslashes will be needed
    if (get_magic_quotes_gpc() == 1)
    {
        $query = stripslashes($query);
    }

    // in production code you'll always want to use a try /catch for any
    // possible exceptions emitted by searching (i.e. connection
    // problems or a query parsing error)

    try
    {


        if($_GET['pageRankType'] == "pageRankAlgo") {
            $additionalParameters = array('fq'=>'og_url:[* TO *]','sort'=>'pageRankFile desc');
            // $additionalParameters = array('sort'=>'pageRankFile desc');
            $results = $solr->search($query, 0, $limit, $additionalParameters);
```

Line 249, Column 103 Spaces: 4 PHP

IR_trial.php

```
152
153         <h1>CSCI 572 : Information Retrieval and Web Seach Engines</h1>
154
155         <h3>Assignment #3 : Indexing the Web Using Solr</h3>
156
157         <div class="container">
158
159         <form accept-charset="utf-8" method="get" id="form_container">
160
161             <div class="form-group" id="search_container">
162                 <label for="q" id="searchlabel">Search : </label>
163                 <input type="text" name="q" class="form-control" id="q" aria-describedby="SearchFieldHelp" placeholder="Se
164                 <small id="SearchFieldHelp" class="form-text text-muted">Enter what you want to search here.</small>
165             </div>
166
167
168             <fieldset class="form-group">
169                 <legend id="legend_id">Type of Ranking Algorithm</legend>
170                 <div class="form-check">
171                     <label class="form-check-label">
172                         <input type="radio" class="form-check-input" name="pageRankType" id="optionsRadios1" value="lucene
173                     </label>
174                     <label class="form-check-label">
175                         <input type="radio" class="form-check-input" name="pageRankType" id="optionsRadios2" value="pageRa
176                     </label>
177                 </div>
178             </fieldset>
179
180
181             <button type="submit" class="btn btn-primary">Submit</button>
182
183         </form>
184
185         <?php
186             //  display results
187             if ($results)
188             {
189                 $total  = (int) $results->response->numFound;
190                 $start  = min(1,    $total);
```

Line 249, Column 103 Spaces: 4 PHP

IR_trial.php

```php
<?php
    //  display results
    if ($results)
    {
        $total  = (int) $results->response->numFound;
        $start  = min(1,    $total);
        $end    = min($limit,   $total);
?>

<?php

    //  iterate result  documents
        foreach ($results->response->docs   as $doc)
        {

?>
<div id="results">

<?php
            //  iterate document    fields  /   values
            foreach ($doc   as $field   => $value)
            {

                if ( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "title" ) {

                    $hreftitle = htmlspecialchars($value,ENT_NOQUOTES,'utf-8');
                    $titlebol = 1;
                }


                // if( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "og_url" ) {

                //     $hreflink = htmlspecialchars($value,ENT_NOQUOTES,'utf-8');

                //     // $shorturl = substr($hreflink, 0, 45) . " ...";
                // }

                if( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "og_description" ) {
```

Line 249, Column 103 Spaces: 4 PHP

IR_trial.php

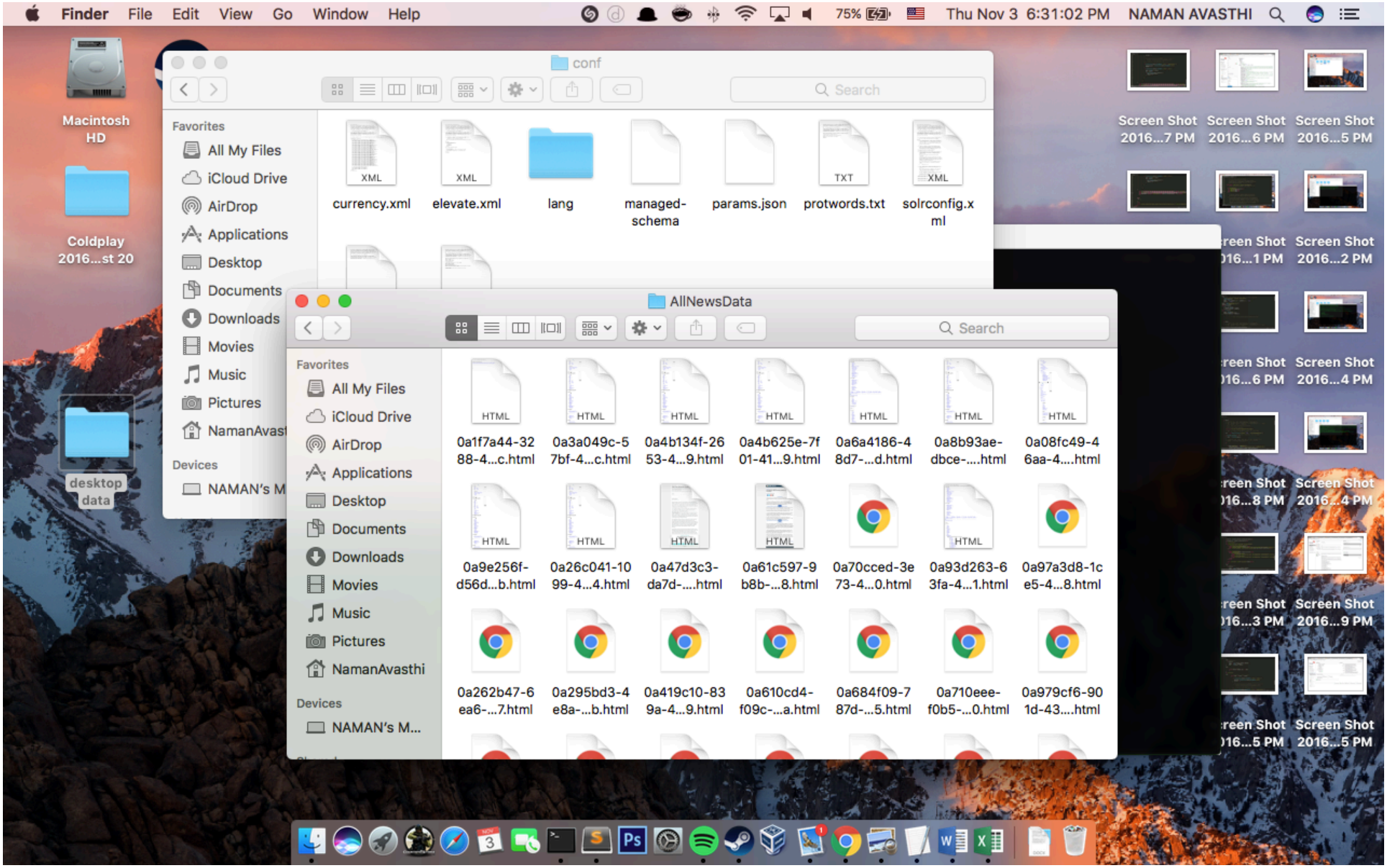
```
211                                 $titlebol = 1;
212                             }
213
214
215                             // if( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "og_url" ) {
216
217                             //     $hreflink = htmlspecialchars($value,ENT_NOQUOTES,'utf-8');
218
219                             //     // $shorturl = substr($hreflink, 0, 45) . " ...";
220                             // }
221
222                             if( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "og_description" ) {
223
224                                 $link_description = htmlspecialchars($value,ENT_NOQUOTES,'utf-8');
225                                 $descbol = 1;
226                             }
227
228                             if ( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "id" ) {
229
230                                 $line = explode("/", $value);
231                                 $fil = $line[6];
232
233                                 if(array_key_exists($fil, $filetourl)) {
234                                     $url = $filetourl[$fil];
235                                     $urlbol = 1;
236                                 }
237
238
239                             }
240
241
242
243                         }
244
245         ?>
246
247             <span>TITLE               : </span><a href="<?php  ec
248             <div class="subtext"><span>FILE                &n
249             <span>URL                 : </span><a class="
```

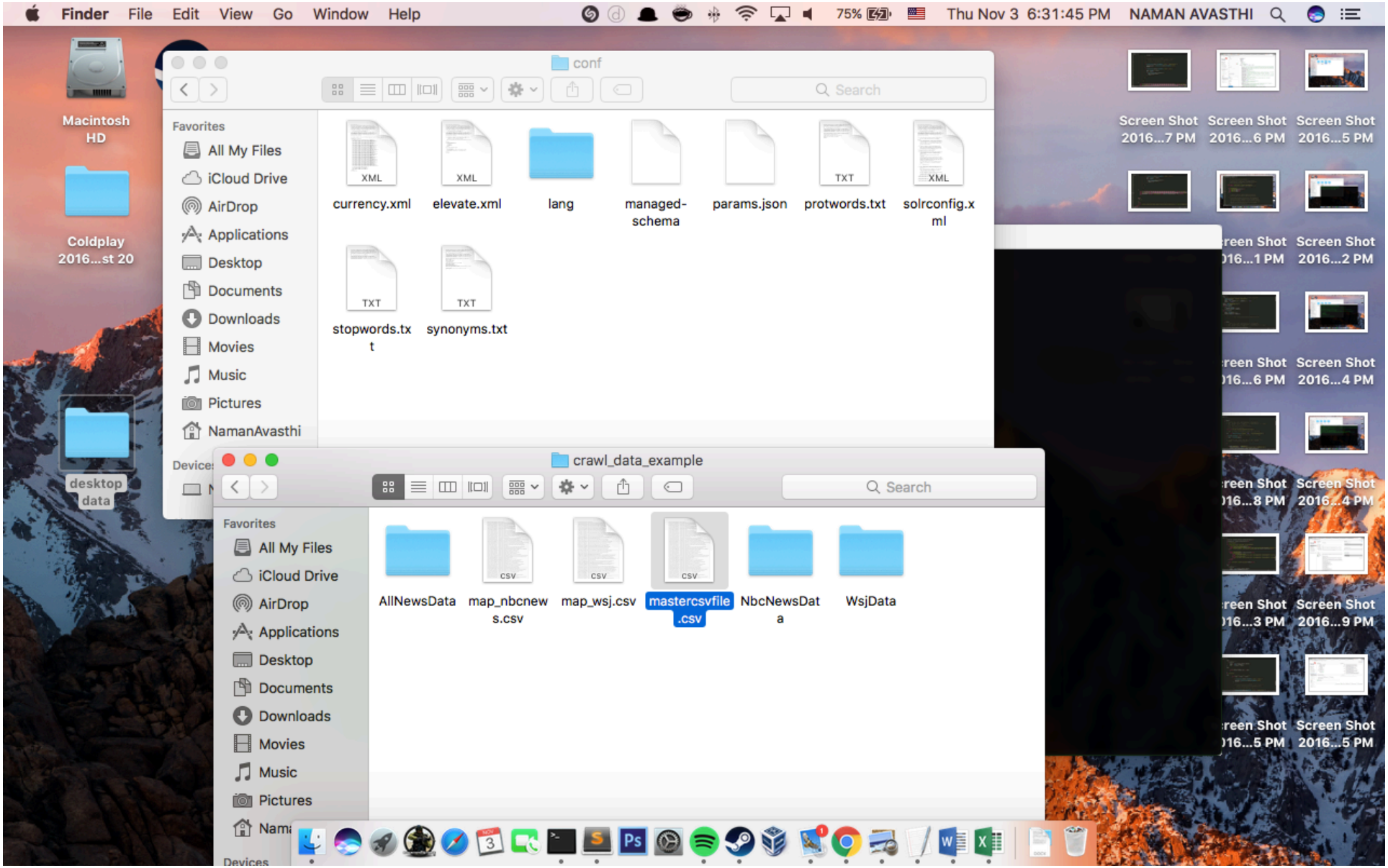
Line 249, Column 103 Spaces: 4 PHP

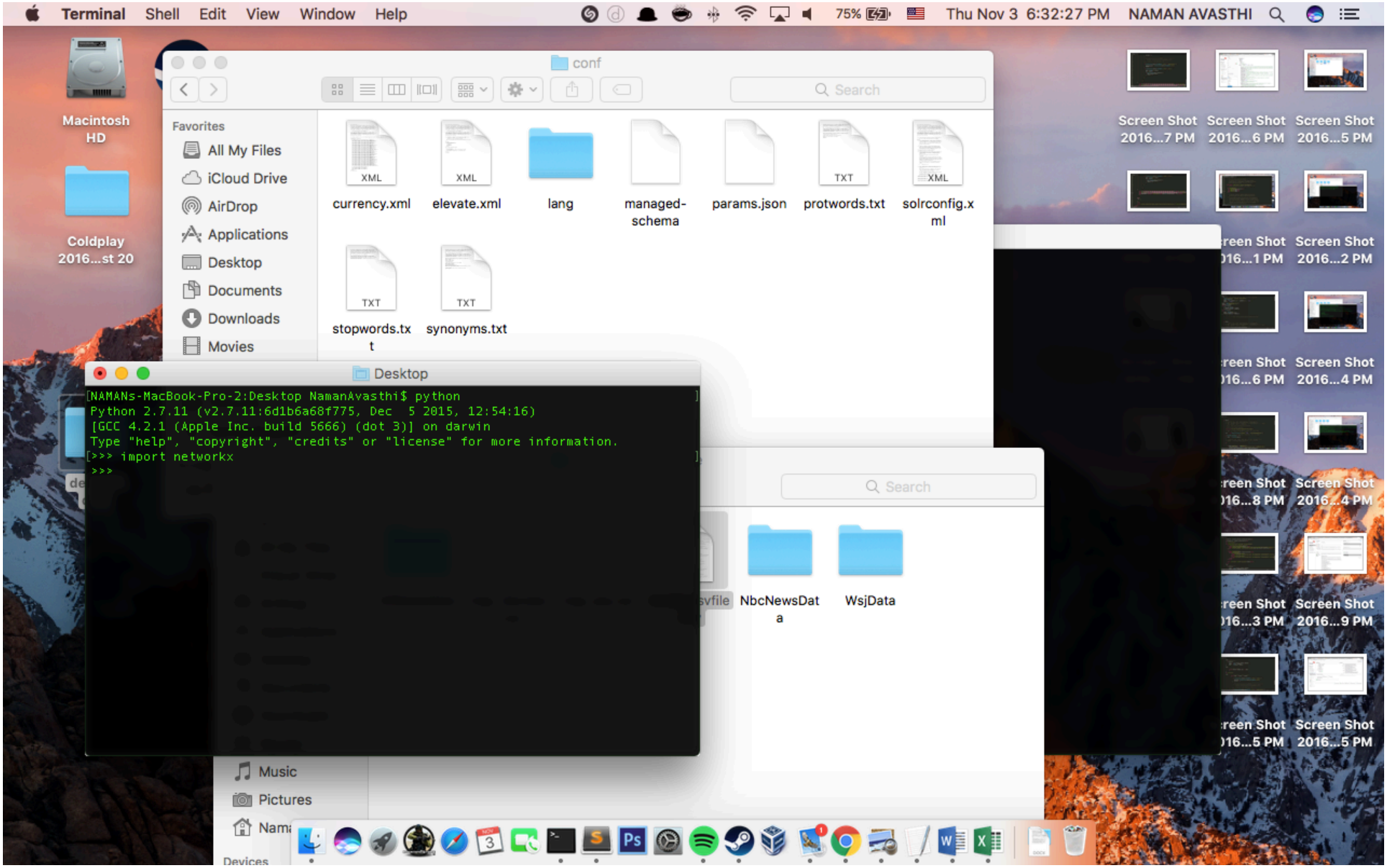
IR_trial.php

```
228                         if ( htmlspecialchars($field, ENT_NOQUOTES, 'utf-8') == "id" ) {
229
230                             $line = explode("/", $value);
231                             $fil = $line[6];
232
233                             if(array_key_exists($fil, $filetourl)) {
234                                 $url = $filetourl[$fil];
235                                 $urlbol = 1;
236                             }
237
238
239                         }
240
241
242
243                     }
244
245         ?>
246
247             <span>TITLE                : </span><a href="<?php  ec
248             <div class="subtext"><span>FILE                &nb
249             <span>URL                 : </span><a class="
250             <span>DESCRIPTION : </span><div class="subtext"><span><?php echo $link_description    ?> </span></div>
251             <br>
252
253
254
255         </div>
256         <?php
257                 }
258             }
259         ?>
260
261         </div>
262
263         <script src="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0-alpha.5/js/bootstrap.min.js" integrity="sha384-BLiI7
264
265     </body>
266 </html>
```

Line 249, Column 103 Spaces: 4 PHP


Finder File Edit View Go Window Help
75%
Thu Nov 3 6:31:02 PM
NAMAN AVASTHI
Macintosh HD
Coldplay 2016...st 20
desktop data
conf
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvast
Devices
NAMAN's M
currency.xml
elevate.xml
lang
managed-schema
params.json
protwords.txt
solrconfig.xml
AllNewsData
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
0a1f7a44-3288-4...c.html
0a3a049c-57bf-4...c.html
0a4b134f-2653-4...9.html
0a4b625e-7f01-41...9.html
0a6a4186-48d7-...d.html
0a8b93ae-dbce-....html
0a08fc49-46aa-4....html
0a9e256f-d56d...b.html
0a26c041-1099-4...4.html
0a47d3c3-da7d-....html
0a61c597-9b8b-...8.html
0a70cced-3e73-4...0.html
0a93d263-63fa-4...1.html
0a97a3d8-1ce5-4...8.html
0a262b47-6ea6-...7.html
0a295bd3-4e8a-...b.html
0a419c10-839a-4...9.html
0a610cd4-f09c-...a.html
0a684f09-787d-...5.html
0a710eee-f0b5-...0.html
0a979cf6-901d-43....html
Screen Shot 2016...7 PM
Screen Shot 2016...6 PM
Screen Shot 2016...5 PM
Screen Shot 2016...2 PM
Screen Shot 2016...4 PM
Screen Shot 2016...4 PM
Screen Shot 2016...9 PM
Screen Shot 2016...5 PM


Finder File Edit View Go Window Help
75%
Thu Nov 3 6:31:45 PM
NAMAN AVASTHI
Macintosh HD
Coldplay 2016...st 20
desktop data
conf
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
currency.xml
elevate.xml
lang
managed-schema
params.json
protwords.txt
solrconfig.xml
stopwords.txt
synonyms.txt
crawl_data_example
AllNewsData
map_nbcnews.csv
map_wsj.csv
mastercsvfile.csv
NbcNewsData
WsjData
Screen Shot 2016...7 PM
Screen Shot 2016...6 PM
Screen Shot 2016...5 PM
Screen Shot 2016...2 PM
Screen Shot 2016...4 PM
Screen Shot 2016...9 PM


Terminal Shell Edit View Window Help
75%
Thu Nov 3 6:32:27 PM
NAMAN AVASTHI
Macintosh HD
Coldplay 2016...st 20
conf
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
currency.xml
elevate.xml
lang
managed-schema
params.json
protwords.txt
solrconfig.xml
stopwords.txt
synonyms.txt
Desktop
[NAMANs-MacBook-Pro-2:Desktop NamanAvasthi$ python
Python 2.7.11 (v2.7.11:6d1b6a68f775, Dec 5 2015, 12:54:16)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
[>>> import networkx
>>>
NbcNewsData
WsjData
Music
Pictures
Devices
Screen Shot 2016...7 PM
Screen Shot 2016...6 PM
Screen Shot 2016...5 PM
Screen Shot 2016...2 PM
Screen Shot 2016...4 PM
Screen Shot 2016...9 PM

IR_trial.php × networkx_python_code.py ×

```python
import networkx as nx
import os

G=nx.read_edgelist('edgeList.txt')
pr = nx.pagerank(G, alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None)
f = open('external_pageRankFile.txt', 'w+')
for u in G:
    f.write('/Users/NamanAvasthi/Desktop/crawl_data_example/'+u+'='+str(pr[u])+'\n')
```

Line 1, Column 1 Tab Size: 4 Python

managed-schema × solrconfig.xml ×

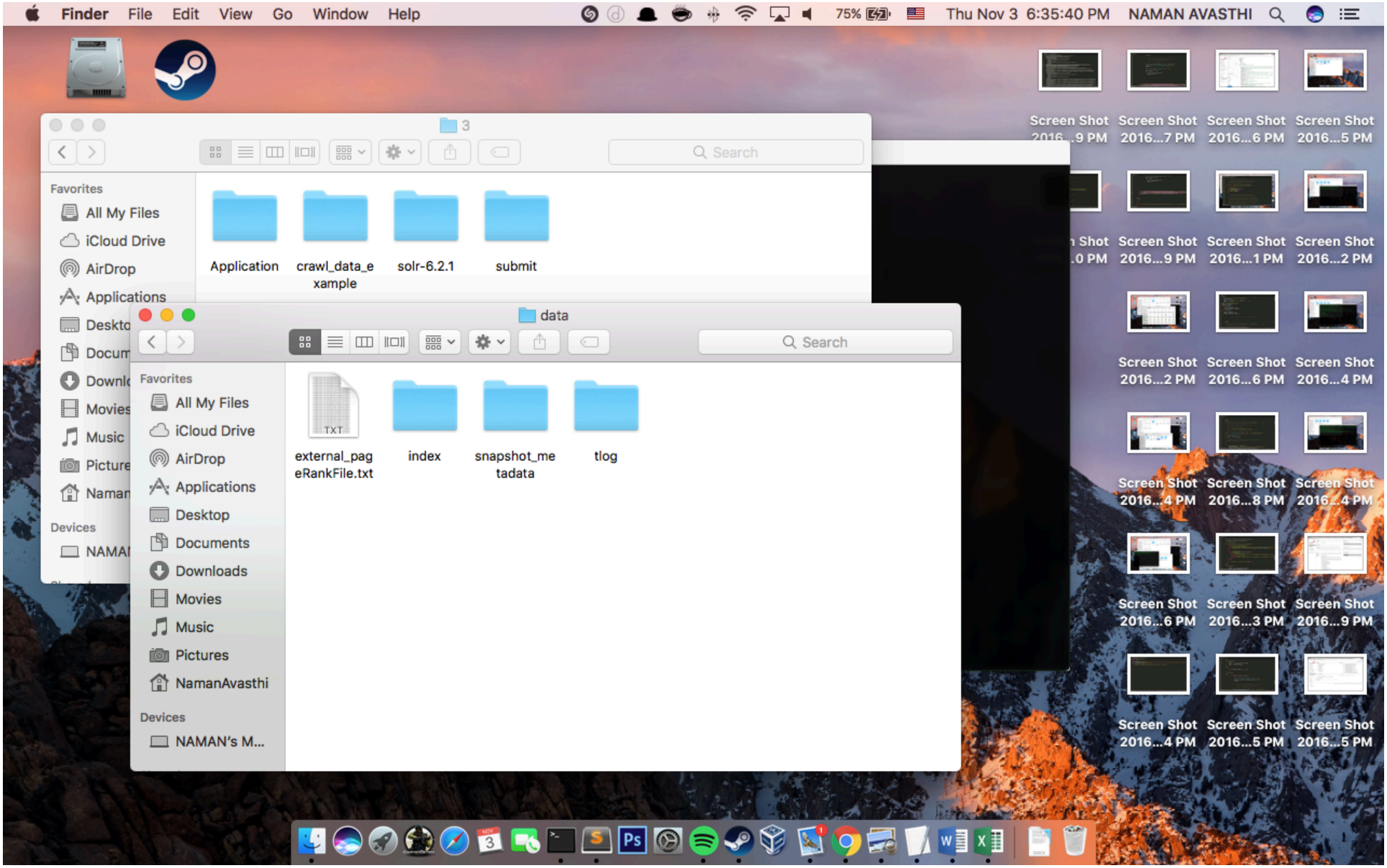
```xml
<?xml version="1.0" encoding="UTF-8"?>
<!-- Solr managed schema - automatically generated - DO NOT EDIT -->
<schema name="example-data-driven-schema" version="1.6">
  <uniqueKey>id</uniqueKey>
  <fieldType name="ancestor_path" class="solr.TextField">
    <analyzer type="index">
      <tokenizer class="solr.KeywordTokenizerFactory"/>
    </analyzer>
    <analyzer type="query">
      <tokenizer class="solr.PathHierarchyTokenizerFactory" delimiter="/"/>
    </analyzer>
  </fieldType>
  <fieldType name="external" keyField="id" defVal="0" class="solr.ExternalFileField" valType="pfloat" />
  <field name="pageRankFile" type="external" stored="false" indexed="false" />
  <fieldType name="binary" class="solr.BinaryField"/>
  <fieldType name="boolean" class="solr.BoolField" sortMissingLast="true"/>
  <fieldType name="booleans" class="solr.BoolField" sortMissingLast="true" multiValued="true"/>
  <fieldType name="currency" class="solr.CurrencyField" currencyConfig="currency.xml" defaultCurrency="USD"
  precisionStep="8"/>
  <fieldType name="date" class="solr.TrieDateField" positionIncrementGap="0" docValues="true" precisionStep="0"/>
  <fieldType name="dates" class="solr.TrieDateField" positionIncrementGap="0" docValues="true" multiValued="true"
  precisionStep="0"/>
  <fieldType name="descendent_path" class="solr.TextField">
    <analyzer type="index">
      <tokenizer class="solr.PathHierarchyTokenizerFactory" delimiter="/"/>
    </analyzer>
    <analyzer type="query">
      <tokenizer class="solr.KeywordTokenizerFactory"/>
    </analyzer>
  </fieldType>
  <fieldType name="double" class="solr.TrieDoubleField" positionIncrementGap="0" docValues="true" precisionStep="0"/>
  <fieldType name="doubles" class="solr.TrieDoubleField" positionIncrementGap="0" docValues="true" multiValued="true"
  precisionStep="0"/>
  <fieldType name="float" class="solr.TrieFloatField" positionIncrementGap="0" docValues="true" precisionStep="0"/>
  <fieldType name="floats" class="solr.TrieFloatField" positionIncrementGap="0" docValues="true" multiValued="true"
  precisionStep="0"/>
  <fieldType name="ignored" class="solr.StrField" indexed="false" stored="false" docValues="false" multiValued="true"/>
  <fieldType name="int" class="solr.TrieIntField" positionIncrementGap="0" docValues="true" precisionStep="0"/>
  <fieldType name="ints" class="solr.TrieIntField" positionIncrementGap="0" docValues="true" multiValued="true"
```

2 lines, 181 characters selected Spaces: 2 Plain Text

managed-schema | solrconfig.xml

```
343          env - environment variables to set.  (default is none)
344        -->
345      <!-- This example shows how RunExecutableListener could be used
346          with the script based replication...
347          http://wiki.apache.org/solr/CollectionDistribution
348        -->
349      <!--
350        <listener event="postCommit" class="solr.RunExecutableListener">
351          <str name="exe">solr/bin/snapshooter</str>
352          <str name="dir">.</str>
353          <bool name="wait">true</bool>
354          <arr name="args"> <str>arg1</str> <str>arg2</str> </arr>
355          <arr name="env"> <str>MYVAR=val1</str> </arr>
356        </listener>
357        -->
358
359      <listener event="newSearcher" class="org.apache.solr.schema.ExternalFileFieldReloader" />
360      <listener event="firstSearcher" class="org.apache.solr.schema.ExternalFileFieldReloader" />
361
362
363    </updateHandler>
364
365    <!-- IndexReaderFactory
366
367         Use the following format to specify a custom IndexReaderFactory,
368         which allows for alternate IndexReader implementations.
369
370         ** Experimental Feature **
371
372         Please note - Using a custom IndexReaderFactory may prevent
373         certain other features from working. The API to
374         IndexReaderFactory may change without warning or may even be
375         removed from future releases if the problems cannot be
376         resolved.
377
378
379         ** Features that may not work with custom IndexReaderFactory **
380
381         The ReplicationHandler assumes a disk-resident index. Using a
```

Line 358, Column 1 | Spaces: 2 | XML


Finder File Edit View Go Window Help
75%
Thu Nov 3 6:35:40 PM
NAMAN AVASTHI
3
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
Application
crawl_data_example
solr-6.2.1
submit
data
Search
Favorites
All My Files
iCloud Drive
AirDrop
Applications
Desktop
Documents
Downloads
Movies
Music
Pictures
NamanAvasthi
Devices
NAMAN's M...
TXT
external_pageRankFile.txt
index
snapshot_metadata
tlog
Screen Shot 2016...7 PM
Screen Shot 2016...6 PM
Screen Shot 2016...5 PM
Screen Shot 2016...9 PM
Screen Shot 2016...1 PM
Screen Shot 2016...2 PM
Screen Shot 2016...2 PM
Screen Shot 2016...6 PM
Screen Shot 2016...4 PM
Screen Shot 2016...4 PM
Screen Shot 2016...8 PM
Screen Shot 2016...4 PM
Screen Shot 2016...6 PM
Screen Shot 2016...3 PM
Screen Shot 2016...9 PM
Screen Shot 2016...4 PM
Screen Shot 2016...5 PM
Screen Shot 2016...5 PM

Solr Admin | PHP Solr Form | PHP Solr Form | Homework3.pdf | CSCI 572 Home Pa | GradingGuidelines | Solr Admin

localhost/IR_trial.php?q=Dow+Jones&pageRankType=lucene

CSCI 572 : Information Retrieval and Web Seach Engines

Assignment #3 : Indexing the Web Using Solr

Search :

Dow Jones

Enter what you want to search here.

Type of Ranking Algorithm

Lucene PageRank

Submit

TITLE : Dow Jones Names Its Finance Chief - WSJ
FILE : 99d77518-c571-4126-a43c-74e2c68d913c.html
URL : http://www.wsj.com/articles/SB120365584684285449
DESCRIPTION :
Dow Jones & Co. said it named Stephen Daintith chief financial officer and executive vice president. Mr. Daintith, 43 years old, succeeds Bill Plummer, who resigned in December when News Corp. closed on its purchase of Dow Jones, publisher of The Wall Street Journal.

TITLE : Stand-Alone REIT Grouping Seen Unlikely to Whet Some Funds' Appetites - WSJ
FILE : 59942085-efea-4b41-a7d3-20f5ff8f951c.html
URL : http://www.wsj.com/amp/articles/stand-alone-reit-grouping-seen-unlikely-to-whet-some-funds-appetites-1474659953
DESCRIPTION :
The move by S&P Dow Jones Indices and MSCI Inc. to split off most listed real estate investment trusts from the financials to give them their own grouping, isn't seen likely to spark a round of buying in the sector.

TITLE : Warren Buffett Could Lose an $8-Per-Second Windfall on His Dow Chemical Stock - WSJ
FILE : 45a4c92b-dc40-44db-89a4-81d1bbd9de45.html
URL : http://www.wsj.com/amp/articles/dows-buffett-ceiling-stock-refuses-to-cross-berkshires-8-a-second-threshold-1472062768
DESCRIPTION :
The Dow Jones Industrial Average's recent rise has lifted many stocks in Berkshire Hathaway's portfolio, but the rally in Dow Chemical has put the chemical firm's dividends to Warren Buffett's company at risk.

TITLE : Brexit Fallout: Dow Continues Decline Down More Than 250 Points - NBC News
FILE : 8c73e46c-19ae-42d5-b9a2-9aa5c0de5180.html
URL : http://www.nbcnews.com/storyline/brexit-referendum/amp/brexit-fallout-stocks-set-continue-decline

Solr Admin | PHP Solr Form | PHP Solr Form | Homework3.pdf | CSCI 572 Home Pa | GradingGuidelines | Solr Admin

localhost/IR_trial.php?q=Dow+Jones&pageRankType=pageRankAlgo

CSCI 572 : Information Retrieval and Web Seach Engines

Assignment #3 : Indexing the Web Using Solr

Search :

Dow Jones

Enter what you want to search here.

Type of Ranking Algorithm

Lucene PageRank

Submit

TITLE : Mansions For Sale: Oil's Rough Ride Hits Houston's Millionaires - NBC News
FILE : 0b8b8291-3d91-4926-8dc7-3679dac02c40.html
URL : http://www.nbcnews.com/business/economy/mansions-sale-oil-s-rough-ride-hits-houston-s-millionaires-n525486
DESCRIPTION :
Prices for mansions in Houston's swankiest neighborhood have tumbled in step with crude prices.

TITLE : Dow Drops 100 Points on Surprise Rise in Crude Inventories - NBC News
FILE : 42571a84-eb89-4cd0-bdda-4378a2910efd.html
URL : http://www.nbcnews.com/business/markets/dow-drops-100-points-surprise-rise-crude-inventories-n640721
DESCRIPTION :
Stocks traded lower Wednesday, with energy lagging, as investors analyzed employment data and looked ahead to Friday's jobs report.

TITLE : BoE Stimulus Sledgehammer Lifts World Stocks, UK Shares Hit 1-Year High - NBC News
FILE : 43470142-24da-4fee-b33b-67913b16b01e.html
URL : http://www.nbcnews.com/business/markets/amp/boe-stimulus-sledgehammer-lifts-world-stocks-uk-shares-hit-1-n623441
DESCRIPTION :
Acting on its chief economist's wish to use "a sledgehammer to crack a nut", the BoE cut interest rates and unleashed billions of pounds of stimulus.

TITLE : Brexit Wiped $2 Trillion Off Markets: Standard & Poors - NBC News
FILE : 4ac176a4-6fc4-470f-a457-4f110401017f.html
URL : http://www.nbcnews.com/storyline/brexit-referendum/brexit-wiped-2-trillion-wiped-markets-standard-poors-n599411

BUSINESS
FEB 25 2016, 7:23 AM ET

Mansions For Sale: Oil's Rough Ride Hits Houston's Millionaires

by REUTERS

SHARE
Share
Tweet
Share
Email
Print

Prices for mansions in Houston's swankiest neighborhood have tumbled in lock step with crude prices. The Houston Opera has offered free season tickets to patrons who lost their jobs in the oil bust. A fancy restaurant offers cut-price dinners.

Twenty months into the worst oil price crash since the 1980s, well-heeled residents of the world's oil capital are among the hardest hit largely because tanking energy firm shares make up much of oil and gas executives' compensation.

In River Oaks, a neighborhood of palatial mansions and lush gardens, the average sales price of a home has tumbled to $1.3 million from $2 million in the middle of 2014 when oil began its more than 70 percent slide, according to data from the Houston Association of Realtors and Keller Williams. Median property prices in the area have already fallen further in this downturn, which is not yet over, than the 16 percent drop in the previous oil slump in 2008 and 2009.

"When oil does well, River Oaks does well. When oil does bad River Oaks does bad," said Paige Martin, a Keller Williams broker who specializes in the