Task 3 :

1. Install simplehtmldom from : [**https://excellmedia.dl.sourceforge.net/project/simplehtmldom/simplehtmldom/1.7/simplehtmldom\_1\_7.zip**](https://excellmedia.dl.sourceforge.net/project/simplehtmldom/simplehtmldom/1.7/simplehtmldom_1_7.zip)
2. **Create a** php file index.php
3. Add below coding to php script file

**<?php**

**include("simple\_html\_dom.php");**

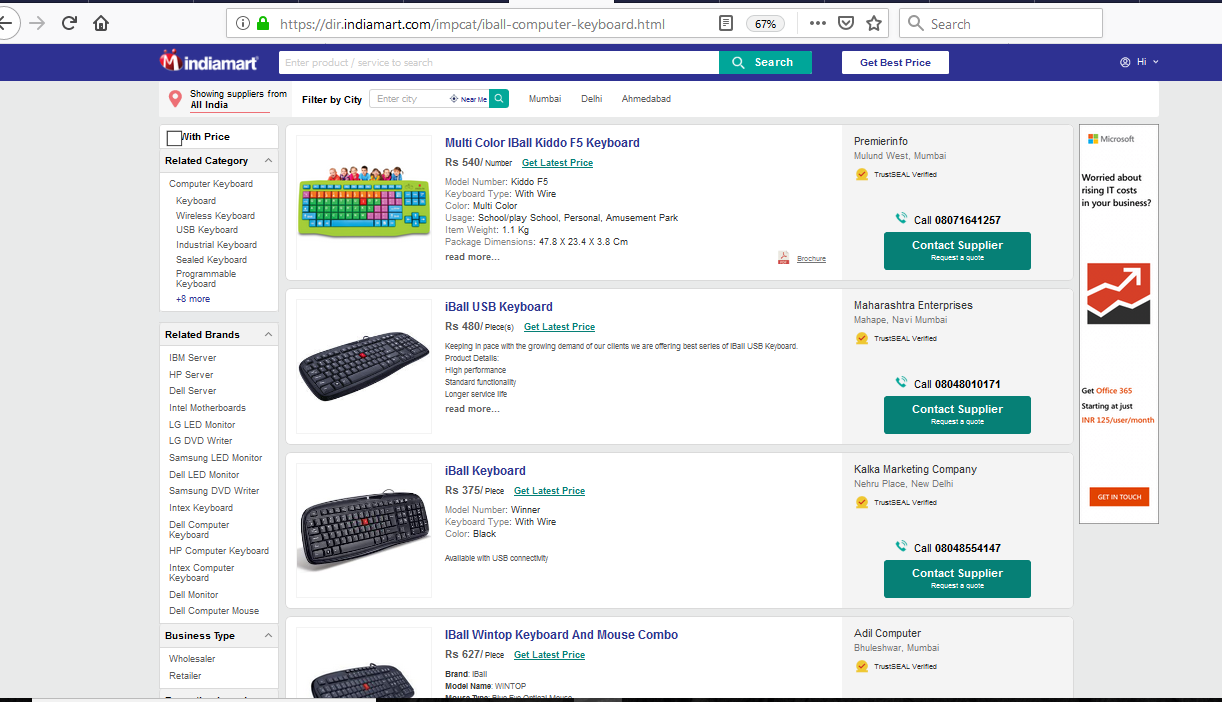
**$html = file\_get\_html('https://dir.indiamart.com/impcat/iball-computer-keyboard.html');**

**foreach($html->find('.lg') as $element)**

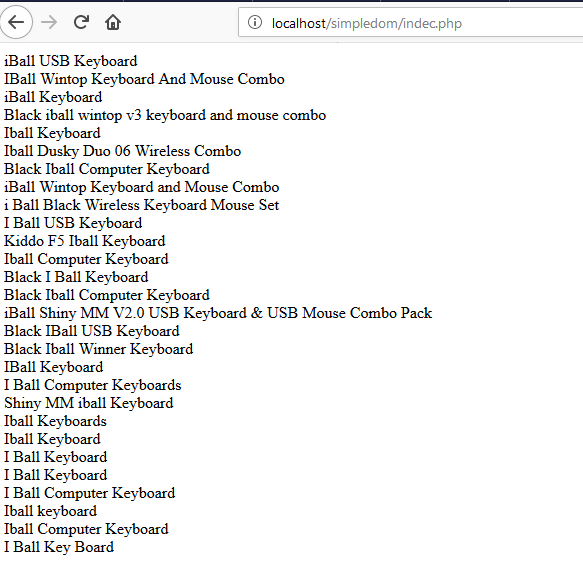
**echo $element->plaintext . '<br>';**

**?>**

**Website look like-- https://dir.indiamart.com/impcat/iball-computer-keyboard.html**



1. **After running the script all the products of particular vendor Iball in above URL is shown in the page**



**Task 2:**

1. Install simplehtmldom from : [**https://excellmedia.dl.sourceforge.net/project/simplehtmldom/simplehtmldom/1.7/simplehtmldom\_1\_7.zip**](https://excellmedia.dl.sourceforge.net/project/simplehtmldom/simplehtmldom/1.7/simplehtmldom_1_7.zip)
2. **Create a 2** php file wordpress.php and wordpressimage.php
3. A) wordpress.php- This script fetch all post , content of the website and header file tag and various other element**.**

**<?php**

**include("simple\_html\_dom.php");**

**$html = file\_get\_html('https://wordpress.com/create-blog/');**

**$scriptDocument = new DOMDocument();**

**//disable libxml errors**

**libxml\_use\_internal\_errors(TRUE);**

**//check if any html is actually returned**

**if(!empty($html)){**

**//loadHTML**

**$scriptDocument->loadHTML($html);**

**//clear errors for yucky html**

**libxml\_clear\_errors();**

**//init DOMXPath**

**$scriptDOMXPath = new DOMXPath($scriptDocument);**

**//get all the h2's with an id**

**$scriptRow = $scriptDOMXPath->query('//h2[@id]');**

**//check**

**if($scriptRow->length > 0){**

**foreach($scriptRow as $row){**

**echo $row->nodeValue . "<br/>";**

**}**

**}**

**//get all the h1's**

**$scriptRow = $scriptDOMXPath->query('//h1');**

**//check**

**if($scriptRow->length > 0){**

**foreach($scriptRow as $row){**

**echo $row->nodeValue . "<br/>";**

**}**

**}**

**//get all the unorder list**

**$scriptRow = $scriptDOMXPath->query('//ul');**

**//check**

**if($scriptRow->length > 0){**

**foreach($scriptRow as $row){**

**echo $row->nodeValue . "<br/>";**

**}**

**}**

**$scriptRow = $scriptDOMXPath->query('//div');**

**//check**

**if($scriptRow->length > 0){**

**foreach($scriptRow as $row){**

**echo $row->nodeValue . "<br/>";**

**}**

**}**

**}**

**?**

B) **wordpressimage.php- This script fetch all image and other content from the website**

<?php

include("simple\_html\_dom.php");

//get data

$html = file\_get\_html('https://wordpress.com/create-blog/');

//init

$matches = array();

//Here’s the code for grabbing the images

$output = preg\_match\_all('/<img.+src=[\'"]([^\'"]+)[\'"].\*>/i', $html, $matches,PREG\_SET\_ORDER );

//display

echo "<pre>";

print\_r($matches);

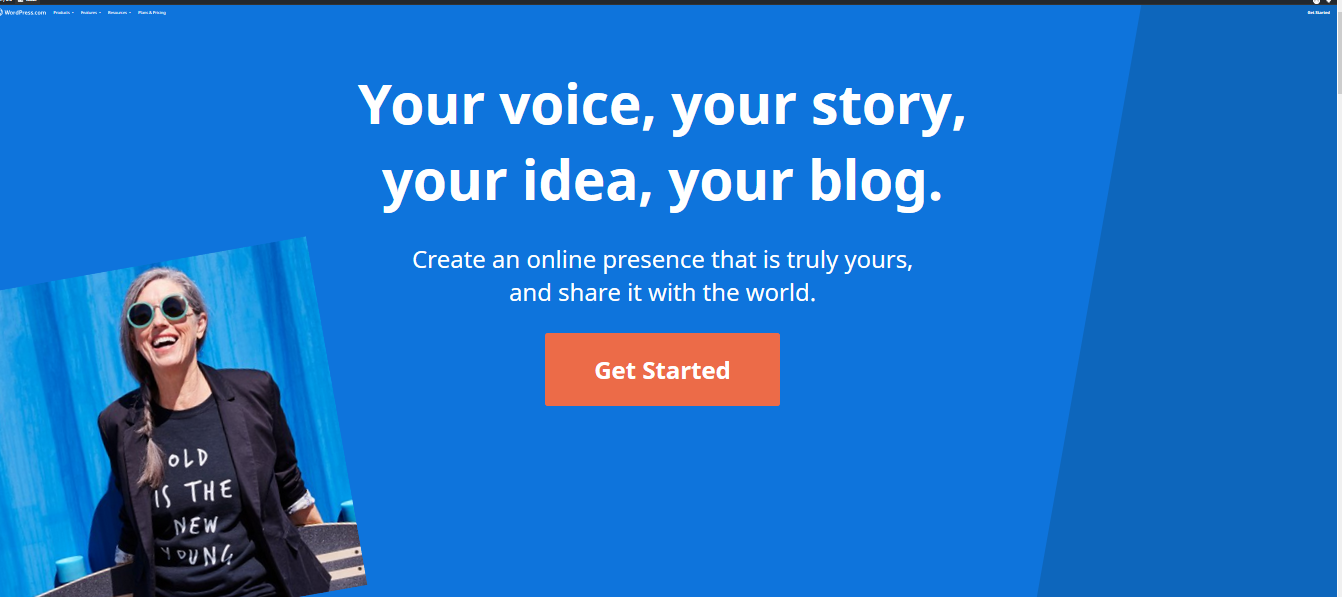
echo "</pre>";

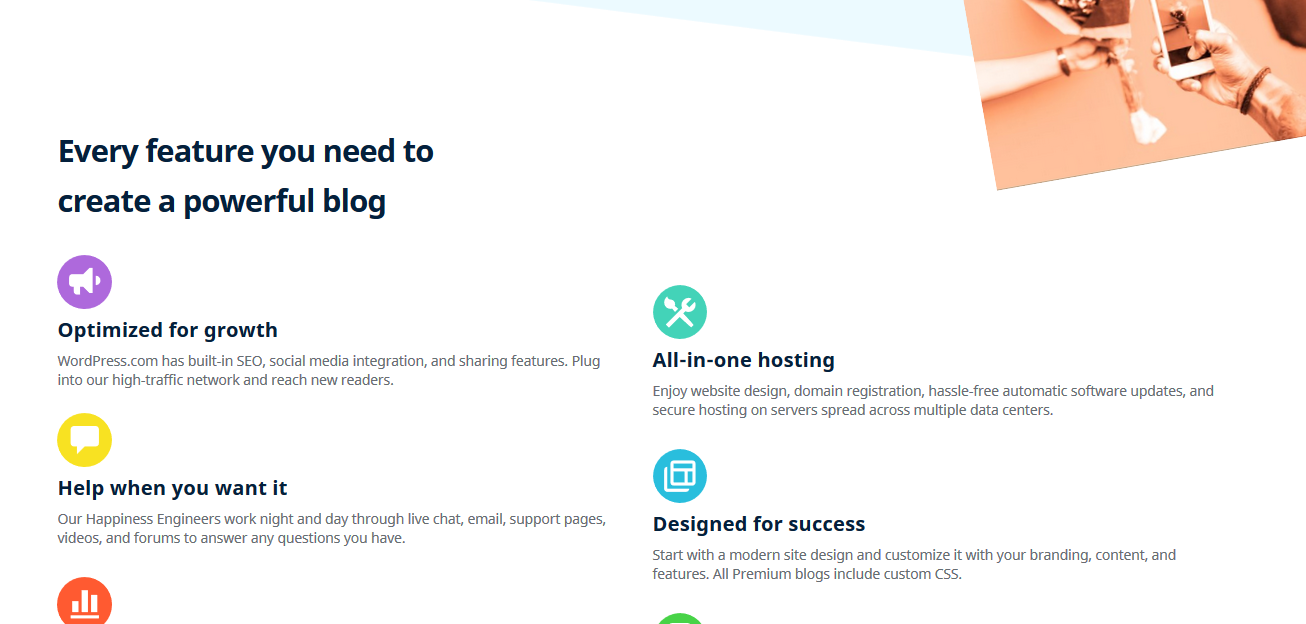
//display

echo "<p>" . count($matches) . " matches found</p>";

?>

**Website look like----https://wordpress.com/create-blog/**

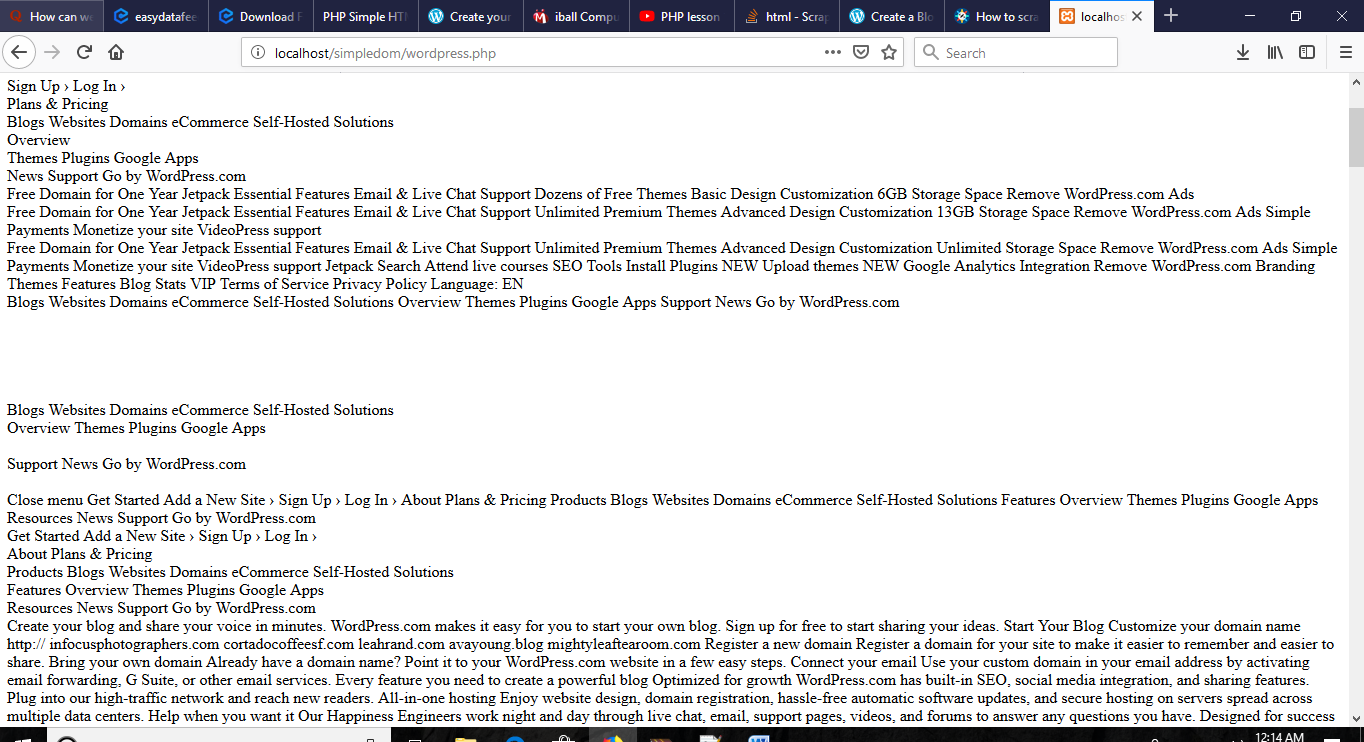






1. **After running the script output is**

**Wordpress.php output**



**Wordpressimage.php output**

