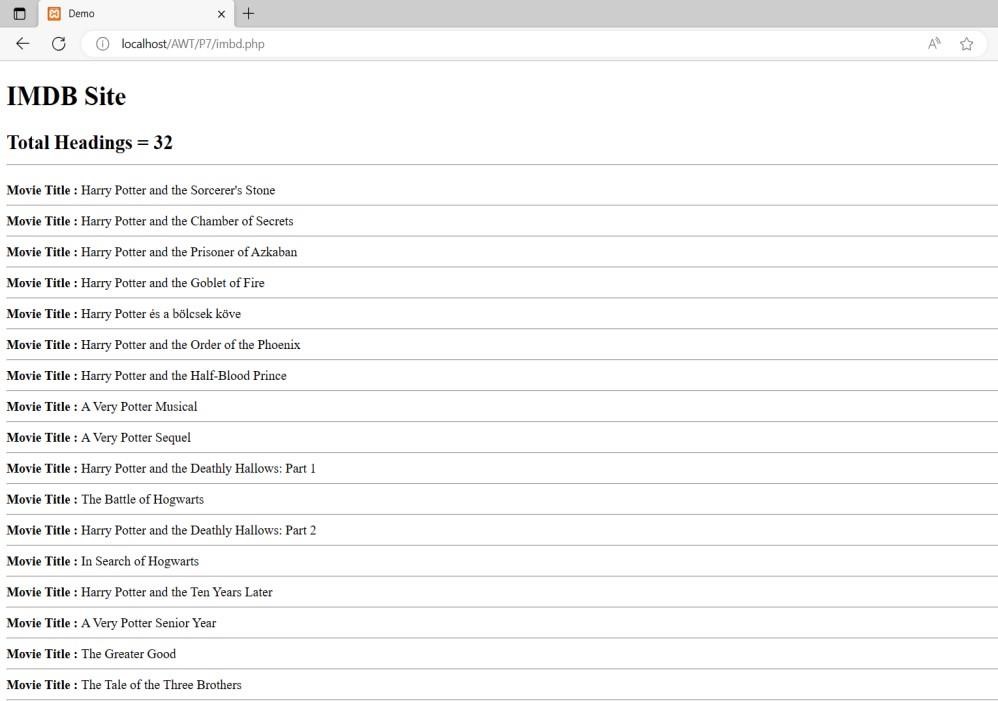
**Lab#7 Using PHP Curl,**

**(a) Implementing Web Scrapping using CURL to fetch the relevant data from IMDB Movie Website.**

|  |  |  |
| --- | --- | --- |
| **PROPERTY USED** | **ATTRIBUTES USED** | **DESCRIPTION** |
| <?php | -- | a server side scripting language that is embedded in HTML |
| curl\_init() |  | Initialize a curl session |
| curl\_setopt() | $curl,CURLOPT\_URL,$url | Set various options for the session |
| curl\_exec() | $curl | Execute and fetch/send data from/to server |
| errno() | $curl | Returns the error number or 0 (zero) if no error occurred |
| curl\_close() | $curl | Close the session |
| preg\_match\_all | pattern,input,matches  ,flags,offset | Returns the number of times the pattern was found in the string, which also may be 0 |

**Final Outcome:**



**Code Snippet:**

<!DOCTYPE html>

<html>

<head>

<title>Demo</title>

<head>

<body>

<h1>IMDB Site</h1>

<?php

$curl=curl\_init(); if(!$curl){

die("Cannot init curl.");

}

curl\_setopt($curl,CURLOPT\_URL,"https://www.imdb.com/list/ls072049366/");

curl\_setopt($curl,CURLOPT\_RETURNTRANSFER,true); curl\_setopt($curl,CURLOPT\_SSL\_VERIFYHOST,false); curl\_setopt($curl,CURLOPT\_SSL\_VERIFYPEER,false);

$result = curl\_exec($curl); if(curl\_errno($curl)){ echo "Error: ".curl\_error($curl); }

else {

$regex = '/<h3 class="lister\-item\-header">.\*?<a href=".\*?>(.\*?)<\/a>.\*?<\/h3>/s';

$match\_count=preg\_match\_all($regex,$result,$arr);

echo "<h2>Total Headings = ".$match\_count."<br/><hr/></h2>";

$mn = 1;

foreach ($arr[1] as $title) { echo "<div class='movie-title'>" . $mn . ". Movie Title: " . $title . "</div><hr>"; $mn++;

}

}

curl\_close($curl);

?>

</body>

</html>