FACULTY OF ENGINEERING AND TECHNOLOGY
Department of Computer Engineering
01IT0701 – Advance Web Technology – Student Lab Manual
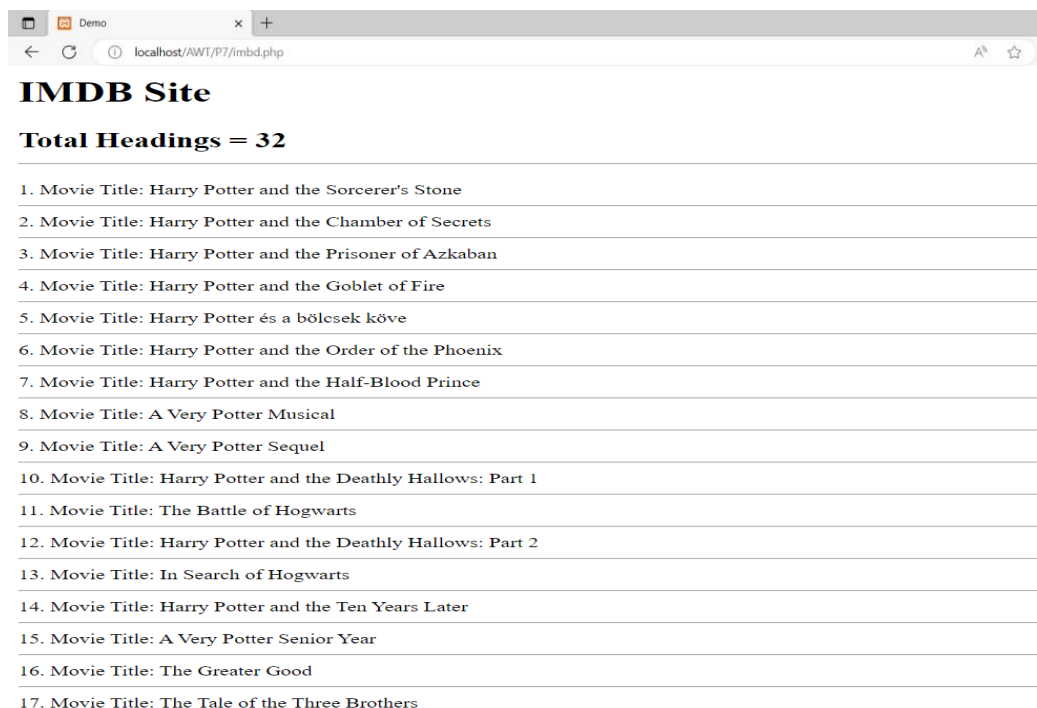
Lab #7(a)

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:

IMDB Site

Total Headings = 32

1. Movie Title: Harry Potter and the Sorcerer's Stone
2. Movie Title: Harry Potter and the Chamber of Secrets
3. Movie Title: Harry Potter and the Prisoner of Azkaban
4. Movie Title: Harry Potter and the Goblet of Fire
5. Movie Title: Harry Potter és a bölcsek köve
6. Movie Title: Harry Potter and the Order of the Phoenix
7. Movie Title: Harry Potter and the Half-Blood Prince
8. Movie Title: A Very Potter Musical
9. Movie Title: A Very Potter Sequel
10. Movie Title: Harry Potter and the Deathly Hallows: Part 1
11. Movie Title: The Battle of Hogwarts
12. Movie Title: Harry Potter and the Deathly Hallows: Part 2
13. Movie Title: In Search of Hogwarts
14. Movie Title: Harry Potter and the Ten Years Later
15. Movie Title: A Very Potter Senior Year
16. Movie Title: The Greater Good
17. Movie Title: The Tale of the Three Brothers

Code Snippet:

```php
<!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>
```