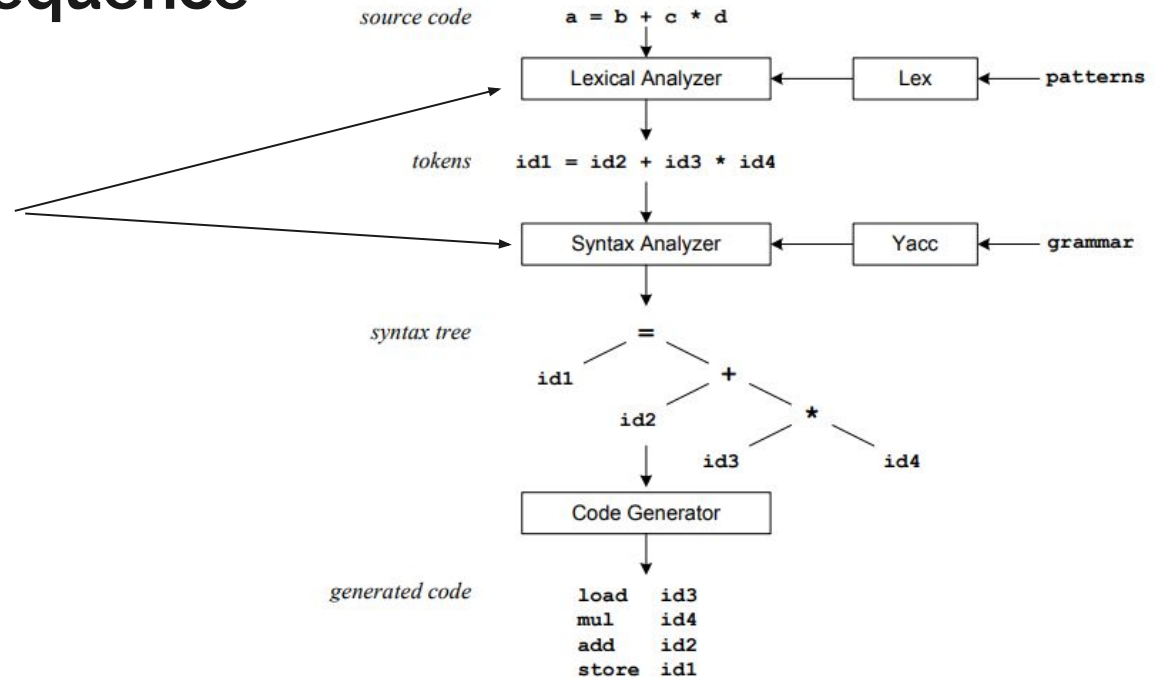




Lex & Yacc Introduction

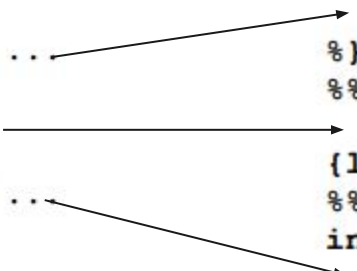
Compilation Sequence

Our focus in the next two assignments, To design Lex & Yacc to parse files



Lex structure

```
... definitions ...  
%%  
... rules ...  
%%  
... subroutines ...  
digit      [0-9]  
letter     [A-Za-z]  
%{  
    int count;  
}%  
%%  
    /* match identifier */  
    {letter}({letter}|{digit})*      count++;  
%%  
int main(void) {  
    yylex();  
    printf("number of identifiers = %d\n", count);  
    return 0;  
}
```





Variables in Lex

- **yyin**
 - yyin is a variable of the type FILE* and points to the input file. yyin is defined by LEX automatically. If the programmer assigns an input file to yyin in the auxiliary functions section, then yyin is set to point to that file. Otherwise LEX assigns yyin to stdin(console input).
- **yytext**
 - yytext is of type char* and it contains the lexeme currently found. A lexeme is a sequence of characters in the input stream that matches some pattern in the Rules Section.
- **yylen**
 - yylen is a variable of the type int and it stores the length of the lexeme pointed to by yytext.



Functions in Lex

- **yylex()**
 - `yylex()` is a function of return type `int`. LEX automatically defines `yylex()` in `lex.yy.c` but does not call it. The programmer must call `yylex()` in the auxiliary functions section of the LEX program. LEX generates code for the definition of `yylex()` according to the rules specified in the Rules section.
- **yywrap()**
 - LEX declares the function `yywrap()` of return-type `int` in the file `lex.yy.c`. LEX does not provide any definition for `yywrap()`. `yylex()` makes a call to `yywrap()` when it encounters the end of input.

Yacc

```
... definitions ...  
%%  
... rules ...  
%%  
... subroutines ...
```

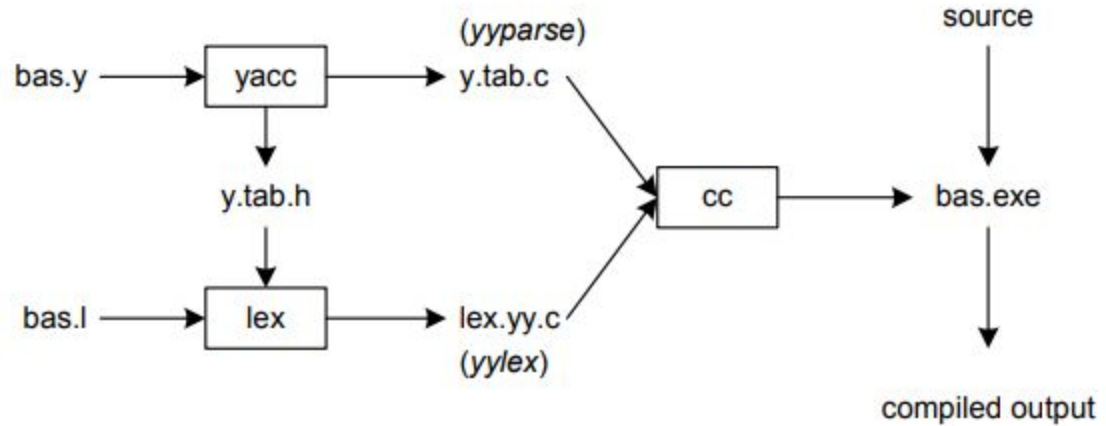
```
%{  
    #include <stdio.h>  
    int yylex(void);  
    void yyerror(char *);  
%}  
  
%token INTEGER  
  
%%  
  
program:  
    program expr '\n'          { printf("%d\n", $2); }  
    |  
    ;  
  
expr:  
    INTEGER                    { $$ = $1; }  
    | expr '+' expr            { $$ = $1 + $3; }  
    | expr '-' expr            { $$ = $1 - $3; }  
    ;  
  
%%  
  
void yyerror(char *s) {  
    fprintf(stderr, "%s\n", s);  
}  
  
int main(void) {  
    yyparse();  
    return 0;  
}
```



Yacc: Grammar Rules

- **Terminals (tokens) :**
 - `%token name1 name2 ...`
 - Any name not declared as a token in the declarations section is assumed to be a nonterminal.
- **Start symbol :**
 - may be declared, via: `%start name`
 - if not declared explicitly, defaults to the nonterminal on the LHS of the first grammar rule listed.
- **Productions :**
 - A grammar production $A \rightarrow B_1B_2 \dots B_n$ is written as `A : B1B2 ... Bn ;`
 - Left-recursion is preferred to right-recursion for efficiency reasons.

Whole flow

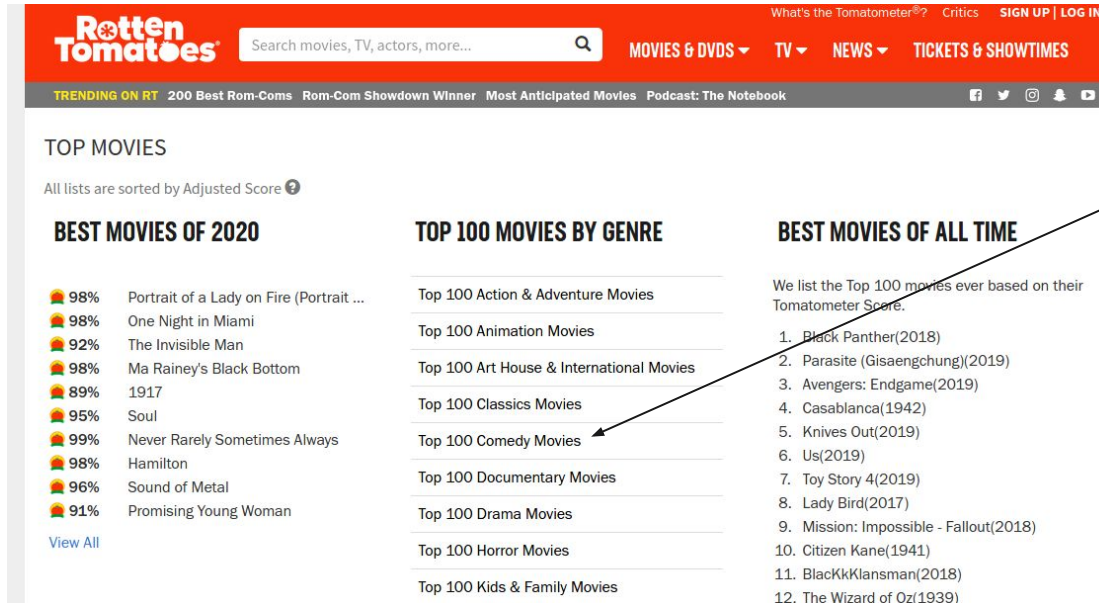


```
yacc -d bas.y          # create y.tab.h, y.tab.c
lex bas.l              # create lex.yy.c
cc lex.yy.c y.tab.c -obas.exe # compile/link
```




Assignment - 8

Rotten Tomatoes



The screenshot shows the Rotten Tomatoes homepage. At the top is a red navigation bar with the Rotten Tomatoes logo, a search bar, and links for MOVIES & DVDS, TV, NEWS, and TICKETS & SHOWTIMES. Below the navigation bar is a grey bar with trending topics like '200 Best Rom-Coms' and 'Rom-Com Showdown Winner'. The main content area is titled 'TOP MOVIES' and includes a note that all lists are sorted by Adjusted Score. There are three main sections: 'BEST MOVIES OF 2020', 'TOP 100 MOVIES BY GENRE', and 'BEST MOVIES OF ALL TIME'. An arrow points from a yellow text box to the 'Top 100 Comedy Movies' link in the 'TOP 100 MOVIES BY GENRE' section.

TOP MOVIES
All lists are sorted by Adjusted Score

BEST MOVIES OF 2020

98%	Portrait of a Lady on Fire (Portrait ...
98%	One Night in Miami
92%	The Invisible Man
98%	Ma Rainey's Black Bottom
89%	1917
95%	Soul
99%	Never Rarely Sometimes Always
98%	Hamilton
96%	Sound of Metal
91%	Promising Young Woman

[View All](#)

TOP 100 MOVIES BY GENRE

- Top 100 Action & Adventure Movies
- Top 100 Animation Movies
- Top 100 Art House & International Movies
- Top 100 Classics Movies
- Top 100 Comedy Movies
- Top 100 Documentary Movies
- Top 100 Drama Movies
- Top 100 Horror Movies
- Top 100 Kids & Family Movies

BEST MOVIES OF ALL TIME

We list the Top 100 movies ever based on their Tomatometer Score.

1. Black Panther(2018)
2. Parasite (Gisaengchung)(2019)
3. Avengers: Endgame(2019)
4. Casablanca(1942)
5. Knives Out(2019)
6. Us(2019)
7. Toy Story 4(2019)
8. Lady Bird(2017)
9. Mission: Impossible - Fallout(2018)
10. Citizen Kane(1941)
11. BlackKkKlansman(2018)
12. The Wizard of Oz(1939)

You will be provided with some of the Top 100 movies by genre URL

Contd.

movie genre link:

1. Action &Adventure:
https://www.rottentomatoes.com/top/bestofrt/top_100_action_adventure_movies/
2. Animation:
https://www.rottentomatoes.com/top/bestofrt/top_100_animation_movies/
3. Drama:
https://www.rottentomatoes.com/top/bestofrt/top_100_drama_movies/
4. Comedy:
https://www.rottentomatoes.com/top/bestofrt/top_100_comedy_movies/
5. Mystery & Suspense:
https://www.rottentomatoes.com/top/bestofrt/top_100_mystery_suspense_movies/
6. Horror:
https://www.rottentomatoes.com/top/bestofrt/top_100_horror_movies/
7. Sci-Fi:
https://www.rottentomatoes.com/top/bestofrt/top_100_science_fiction_fantasy_movie_s/
8. Documentary:
https://www.rottentomatoes.com/top/bestofrt/top_100_documentary_movies/
9. Romance:
https://www.rottentomatoes.com/top/bestofrt/top_100_romance_movies/
10. Classics:
https://www.rottentomatoes.com/top/bestofrt/top_100_classics_movies/

Click on
the link

TOP 100 ACTION & ADVENTURE MOVIES



BEST OF ROTTEN TOMATOES

Movies with 40 or more critic reviews vie for their place in history at Rotten Tomatoes. Eligible movies are ranked based on their Adjusted Scores.

Genre: Action & Adventure

Sorted by Adjusted Score

Rank	Rating	Title	No. of Reviews
1.	96%	Black Panther (2018)	518
2.	94%	Avengers: Endgame (2019)	537
3.	97%	Mission: Impossible - Fallout (2018)	432
4.	97%	Mad Max: Fury Road (2015)	426
5.	97%	Spider-Man: Into the Spider-Verse (2018)	388
6.	93%	Wonder Woman (2017)	462
7.	92%	Dunkirk (2017)	457
8.	97%	Coco (2017)	347
9.	93%	Thor: Ragnarok (2017)	429
10.	93%	Logan (2017)	418



Task 1

1. For this assignment your task is to wait for user input for a genre (any of 10 genre URL provided).
2. Then given the input, you need to list all the 100 movies on that genre. (To get that you need to crawl the HTML file for that genre)
3. Then you need to prompt the user to input any of the movies from that list and given the input you need to download and save that particular movie HTML page from rottentomatoes.com
4. Study the syntax of the HTML file.

In the Movie page you can find such infos...



BLACK PANTHER

2018, Fantasy/Action, 2h 14m



96%

TOMATOMETER
518 Reviews



79%

AUDIENCE SCORE
50,000+ Ratings

CAST & CREW



Chadwick
Boseman
T'Challa, Black
Panther



Michael B.
Jordan
Erik Killmonger



Lupita Nyong'o
Nakia



Danai Gurira
Okoye



Martin
Freeman
Everett K. Ross



Daniel
Kaluuya
W'Kabi

MOVIE INFO

After the death of his father, T'Challa returns home to the African nation of Wakanda to take his rightful place as king. When a powerful enemy suddenly reappears, T'Challa's mettle as king – and as Black Panther – gets tested when he's drawn into a conflict that puts the fate of Wakanda and the entire world at risk. Faced with treachery and danger, the young king must rally his allies and release the full power of Black Panther to defeat his foes and secure the safety of his people.

Genre: Fantasy, Action, Adventure

Original Language: English

Director: [Ryan Coogler](#)

Producer: [Kevin Feige](#)

Writer: [Ryan Coogler](#), [Joe Robert Cole](#)

Release Date (Theaters): Feb 16, 2018 Wide

Release Date (Streaming): May 2, 2018

Box Office (Gross USA): \$700.2M

Runtime: 2h 14m

Production Co: Walt Disney Pictures

Sound Mix: DTS, Dolby Atmos

Aspect Ratio: Scope (2.35:1)

View the collection: [Marvel Cinematic Universe](#)



Task 2

1. You need to write a grammar to parse below fields from the HTML file.
 - Movie Name
 - Director
 - Writers
 - Producer
 - Original Language
 - Cast with the character name
 - Storyline
 - Box Office Collection
 - Runtime

Parse HTML to get info

```
<li class="meta-row clearfix" data-qa="movie-info-item">
  <div class="meta-label subtle" data-qa="movie-info-item-label">Director:</div>
  <div class="meta-value" data-qa="movie-info-item-value">
    <a href="/celebrity/ryan_coogler" data-qa="movie-info-director">Ryan Coogler</a>
  </div>
</li>
```

Can be used to extract
Director info

```
<li class="meta-row clearfix" data-qa="movie-info-item">
  <div class="meta-label subtle" data-qa="movie-info-item-label">Runtime:</div>
  <div class="meta-value" data-qa="movie-info-item-value">
    <time datetime="P2h 14mM">
      2h 14m
    </time>
  </div>
</li>
```

can be used for Runtime

```
<title>Black Panther (2018) - Rotten Tomatoes</title>
<meta name="description" content="After the death of his father, T'Challa returns home to the African nation of Wakanda to take his rightful place as king. When a powerful enemy suddenly reappears, T'Challa's mettle as king -- and as Black Panther -- gets tested when he's drawn into a conflict that puts the fate of Wakanda and the entire world at risk. Faced with treachery and danger, the young king must rally his allies and release the full power of Black Panther to defeat his foes and secure the safety of his people.">
```

Can be used for
story line

```
<li class="meta-row clearfix" data-qa="movie-info-item">
  <div class="meta-label subtle" data-qa="movie-info-item-label">Box Office (Gross USA):</div>
  <div class="meta-value" data-qa="movie-info-item-value">$858.4M</div>
</li>
```

HTML
snippet for
Box office
collection

```
%option noyywrap
%{
#include "y.tab.h"
%}
%%

"<div class=\"meta-value\" data-qa=\"movie-info-item-value\">\".*\"</div>" { yylval.sval=strdup(yytext);
return ATTR_VALUE;}
"<div class=\"meta-label subtle\" data-qa=\"movie-info-item-label\">\"Box Office\".+\"</div>" { yylval.sval=strdup(yytext);
return ATTR_NAME;}
[\\n]      {;}
. {;}
%% |
```

Sample Lex
code

```
%token <sval> ATTR_NAME
%token <sval> ATTR_VALUE
%start START|
%%
START : ATTR_NAME ATTR_VALUE { process($1) ; process($2) ; exit(0);}
%%
```

Sample Yacc code



Contd.

2. Then you need to show that those field can be queried for the film, and wait for user input for any particular field.
3. Given the input,you need to show corresponding field for the film and also store your result in a log file.(Format is mentioned in the assignment pdf.)



Quick demo of Lex & Yacc



Assignment - 9

Assignment 9 -Task 1

1. You need to add two more fields for a movie with the fields you have already extracted in the last assignment.
 - a. YOU MIGHT ALSO LIKE
 - b. WHERE TO WATCH

YOU MIGHT ALSO LIKE



72% 86%
Interstellar



87% 75%
Ant-Man and The Wasp



94% 87%
Mission: Impossible
Rogue Nation



93% 83%
Wonder Woman



92% 86%
Skyfall

WHERE TO WATCH



Rent/buy



Rent/buy



Rent/buy



Buy



Subscription



Rent/buy



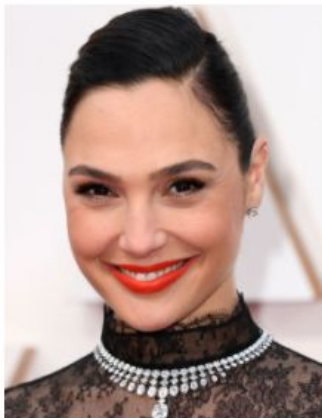
Contd.

2. Here you need to write the code for recursive crawling.

Ex. Movie1 → YOU MIGHT ALSO LIKE[list of movies] → user select one of them → you need to show all the possible fields that can be extracted for the movie → again user may ask for YOU MIGHT ALSO LIKE, so basically it can go on until user wants to exit.

Task -2

1. If the user want to see all the cast members for a movie, you need to list them out and also prompt user to see specific info for an actor/actress such as -
 - a. Highest Rated film
 - b. Lowest Rated film
 - c. Birthday
 - d. His/Her other movies



Gal Gadot

Highest Rated: 🌟 93% [Wonder Woman \(2017\)](#)

Lowest Rated: 🌟 19% [Keeping Up With the Joneses \(2016\)](#)

Birthday: Apr 30, 1985

Birthplace: Rosh Ha'ayin, Israel











Gal Gadot managed to successfully combine both modeling and acting after being cast in a series of blockbuster movies. Born in Rosh HaAyin, Israel, she was crowned Miss Israel in 2004. Due to conscription, Israeli citizens over the age of 18 are required to spend time in the army. Gadot spent her time split between modeling and the military, drawing...

[Read more](#)

Contd.

2. If user wants to see all the movies of his/her, ask for additional year by which you can filter the movies on or after that year.

Movies

TOMATOMETER®	AUDIENCE SCORE	TITLE	CREDIT	BOX OFFICE	YEAR ↕
No Score Yet	No Score Yet	Death on the Nile	Linnet Ridgeway Doyle (Character)	-	2021
No Score Yet	No Score Yet	Zack Snyder's Justice League	Wonder Woman/ Diana Prince (Character)	-	2021
 60%	 74%	Wonder Woman 1984	Diana Prince (Character), Producer	-	2020
No Score Yet	No Score Yet	Red Notice	Unknown (Character)	-	2020
 74%	 43%	Between Two Ferns: The Movie	Self	-	2019
 88%	 65%	Ralph Breaks the Internet	Shank (Voice)	\$201.1M	2018
 93%	 83%	Wonder Woman	Diana (Character)	\$412.8M	2017
 40%	 71%	Justice League	Wonder Woman/ Diana Prince (Character)	\$227M	2017