#! /bin/bash

Exit="N"

echo "----------------------------------------"

echo "User Name: 이태경"

echo "Student Number: 12171683"

echo "[ MENU ]"

echo "1. Get the data of the movie identified by a specific 'movie id' from 'u.item'"

echo "2. Get the data of action genre movies from 'u.item'"

echo "3. Get the average 'rating’ of the movie identified by specific 'movie id' from 'u.data'"

echo "4. Delete the 'IMDb URL' from 'u.item'"

echo "5. Get the data about users from 'u.user'"

echo "6. Modify the format of 'release date' in 'u.item'"

echo "7. Get ::the data of movies rated by a specific 'user id' from 'u.data'"

echo "8. Get the average 'rating' of movies rated by users with'age' between 20 and 29 and 'occupation' as 'programmer'"

echo "9. Exit"

echo "----------------------------------------"

until [ "${Exit}" == "Y" ]; do

read -p "Enter your choice [ 1-9 ] " option

case ${option} in

**1)**

read -p "Please enter 'movie id'(1~1682):" movie\_id

echo ""

awk -F '|' '$1 == '$movie\_id' { print $0 }' ./u.item

;;

=> movie\_id를 입력받아 u.item에서 첫번째 열에서 같은 movie\_id인 행 전부를 프린트

**2)**

read -p "Do you want to get the data of ‘action’ genre movies from 'u.item’?(y/n):" c

echo ""

if [ "${c}" = "y" ]; then

temp\_file=$(mktemp)

awk -F '|' '$7 == 1 { print $1, $2 }' u.item > "$temp\_file"

head -n 10 "$temp\_file"

rm "$temp\_file"

fi

;;

=> file을 만들어 u.item의 $7(Action)이 1인 행의 1,2번째 원소를 대입 후 상위 10개만 출력. 이후에 file 삭제

**3)**

read -p "Please enter the 'movie id’(1~1682):" movie\_id

echo ""

total\_rating=0

num\_user=0

while IFS=$'\t' read -r user\_id movie\_ids rating timestamp; do

if [ "$movie\_id" == "$movie\_ids" ]; then

total\_rating=$((total\_rating + rating))

num\_user=$((num\_user + 1))

fi

done < u.data

echo $movie\_id $total\_rating $num\_user | awk '{printf "average rating of %s: %.5f", $1, ($2 / $3 + 0.000005)}'

;;

movie\_id랑 같으면 total\_rating에 rating을 더하고, num\_users에는 1씩 더한 후에 total\_rating / num\_users로 평균 rating출력. .5f는 버림이므로 반올림을 위해서 뒤에 0.000005를 더한 후 반올림 진행.

**4)**

read -p "Do you want to delete the ‘IMDb URL’ from ‘u.item’?(y/n):" c

echo ""

if [ "${c}" = "y" ]; then

sed -e 's/http[^|]\*|//g' u.item | head -n 10

fi

;;

=> u.item에 http로 시작하는 원소가 있으면 다음 |가 나오기 전까지 그 부분은 삭제 후 상위 10개 행 출력

**5)**

read -p "Do you want to get the data about users from ‘u.user’?(y/n):" c

echo ""

if [ "${c}" = "y" ]; then

head -n 10 u.user | while IFS='|' read -r user\_id age gender occupation zip\_code; do

if [ "$gender" == "M" ]; then

gender="male"

elif [ "$gender" == "F" ]; then

gender="female"

fi

result="user $user\_id is $age years old $gender $occupation"

echo "$result"

done

fi

;;

=> 각각의 원소들을 입력 받아 gender 유무에 따라서 male, female 둘중에 어떤 형태로 출력할 것인지 정하고 result에다가 정보를 저장후 result 출력

**6)**

read -p "Do you want to Modify the format of 'release data' in 'u.item'?(y/n):" c

echo ""

convert\_month() {

case $1 in

"Jan") echo "01";;

"Feb") echo "02";;

"Mar") echo "03";;

Apr) echo "04";;

May) echo "05";;

Jun) echo "06";;

Jul) echo "07";;

Aug) echo "08";;

Sep) echo "09";;

Oct) echo "10";;

Nov) echo "11";;

Dec) echo "12";;

\*) echo "00";;

esac

}

if [ "${c}" = "y" ]; then

sed -E 's/([0-9]{2})-([A-Za-z]{3})-([0-9]{4})/\3'$(convert\_month "$\2")'\1/' u.item | tail -n 10

fi

;;

=> Jan, Feb등을 01, 02….로 만들어주는 함수를 만들고, release\_date 부분을 해당 함수를 이용해 치환 한 후 나머지 부분 전부 출력

**7)**

read -p "Please enter the ‘user id’(1~943):" user\_id

echo ""

movie\_list=$(awk -v user\_id="$user\_id" -F '\t' '$1 == user\_id { print $2 }' u.data)

sort\_movie\_list=$(echo "$movie\_list" | sort -n)

for movie\_id in $sort\_movie\_list; do

echo -n "$movie\_id |"

done

echo ""

echo ""

sorted\_movie\_list=$(echo "$sort\_movie\_list" | head -n 10 )

for movie\_id in $sorted\_movie\_list; do

movie\_name=$(awk -v movie\_id="$movie\_id" -F '|' '$1 == movie\_id { print $2 }' u.item)

echo "$movie\_id | $movie\_name"

done

;;

=> u.data에서 user\_id랑 매치되는 행을 전부 찾고, 그 행들의 두번째 원소(movie\_id)를 전부 탐색해서 sort\_movie\_list에 대입 후 전부 출력. 이후 sorted\_movie\_list에 movie\_id 기준 오름차순을 하여 10개의 데이터를 대입 한 후 출력

**8)**

read -p "Do you want to get the average 'rating' of movies rated by users with 'age' between 20 and 29 and 'occupation' as 'programmer'?(y/n):" c

echo ""

if [ "${c}" = "y" ]; then

user\_ids=$(awk -F '|' '$2 >= 20 && $2 <= 29 && $4 == "programmer" { print $1 }' u.user)

awk -F '\t' -v user\_ids="$user\_ids" '

BEGIN { split(user\_ids, users, " "); }

{

if ($1 in users) {

movie\_ratings[$2] += $3;

movie\_counts[$2] += 1;

}

}

END {

for (movie\_id in movie\_ratings) {

if (movie\_counts[movie\_id] > 0) {

avg\_rating = movie\_ratings[movie\_id] / movie\_counts[movie\_id];

printf("%d %.5g \n", movie\_id, avg\_rating);

}

}

}' u.data

fi

;;

=> u.user에서 20대이면서 programmer인 사람들의 user\_id 추출. 이 user\_id들을 split을 이용해서 users 라는 배열을 만들고 영화별 rating 합산과 인원수를 count를 진행. 마지막에 rating 평균을 계산하고 출력하는데 문제에 소수점 5자리까지 출력하지만 정수일땐 정수만 출력되게 해야하므로 %.5g 이용

**9)**

echo "Bye!"

Exit="Y"

;;

esac

done