

Technical Communication for Computer Scientists

Summer 2013

**Final Report:
The Movie-Chain-Runner Project**

June 25, 2013

Submitted to
Thomas M. Keating
Assistant Teaching Professor
Computer Science Department
Carnegie Mellon University

Prepared and Submitted by

Shashank Singh
sss1@andrew.cmu.edu

Eugene Scanlon
escanlon@cmu.edu

Contents

1 Introduction **1**

 1.1 The Movie-Chain-Runner Problem 1

2 Progress **1**

 2.1 General Progress 1

 2.2 Status 2

 2.3 Projections 3

3 Recomendations **3**

4 Discussion **3**

Appendices **4**

A Our Longest Movie Chain **4**

1 Introduction

This report documents our progress on the Movie-Chain-Runner Problem (discussed in section 2) as of June 17, 2013, as well as our plans for how to proceed until the project deadline of June 27, 2013. We also discuss and justify some changes to the plan presented in our project proposal (submitted June 4, 2013).

1.1 The Movie-Chain-Runner Problem

The Movie-Chain-Runner Problem is to find the longest chain of overlapping titles in a list of movie titles, where two titles are said to overlap if some suffix of the first movie is identical to some prefix of the second movie. For example, in the list

- Day of the Dead
- Live and Let Die
- Dead Poets' Society
- Die Another Day
- The Last Samurai

the longest chain consists of 4 titles:

Live and Let Die Another Day of the Dead Poets' Society

By appropriately representing the movie list as a graph, the Movie-Chain-Runner Problem can be shown to be equivalent to the Longest Path Problem (finding the longest simple path in a directed graph). The Longest Path Problem is well known to be NP-Complete, meaning that no efficient algorithm exists to find longest paths in large graphs, including our movie title graph. A review of the literature reveals that good approximate longest paths also cannot be found efficiently in large graphs. Thus, our project is to study the movie title graph and innovate ways to find long paths in the graph.

2 Progress

We discuss our progress in three sections: General Progress, Status, and Projections.

2.1 General Progress

Our first step was to reconstruct the list of movie titles as a directed graph, reducing the problem to the well-known Longest Path Problem. We proceeded to implement a greedy brute force algorithm that simply followed as many paths as possible. This quickly constructed a path of 243 titles and then ceased to make appreciable progress.

The majority of our efforts since have been toward augmenting this brute force algorithm. We have run the algorithm “backwards” to extend the beginning of the chain. We have also attempted running the algorithm in parallel over several cluster computers, boosting our processing power and thus increasing the number of possible paths we can check. We are currently running the brute force algorithm on subsections of the current longest chain, in an attempt to lengthen the middle of the chain.

We also attempted a theoretically-inspired algorithm which computed acyclic subgraphs of our graph and then ran a polynomial time Longest Path algorithm known for directed acyclic graphs. However, the number of cycles proved too large to generate the subgraphs in a reasonable amount of time, and so we abandoned this effort.

2.2 Status

Our current longest chain (included in Appendix A) consists of 278 titles, 7 titles short of our June 17 goal of 285 titles. Since extending the movie chain has proven more difficult than we anticipated, we have decided to scale back our final goal from 300 titles to 285 titles. Since 285 titles is the necessary criterion for receiving an A on the project, the benefits of 300 titles are primarily cosmetic, and so we consider this change acceptable.

Figure 1 below shows the original Gantt chart presented in our project proposal. Figure 2 below shows our revised Gantt chart, as of June 17. The only changes are that the 300 title goal originally set for June 25 has been eliminated and the 285 title goal originally set for June 17 has been extended to June 25.

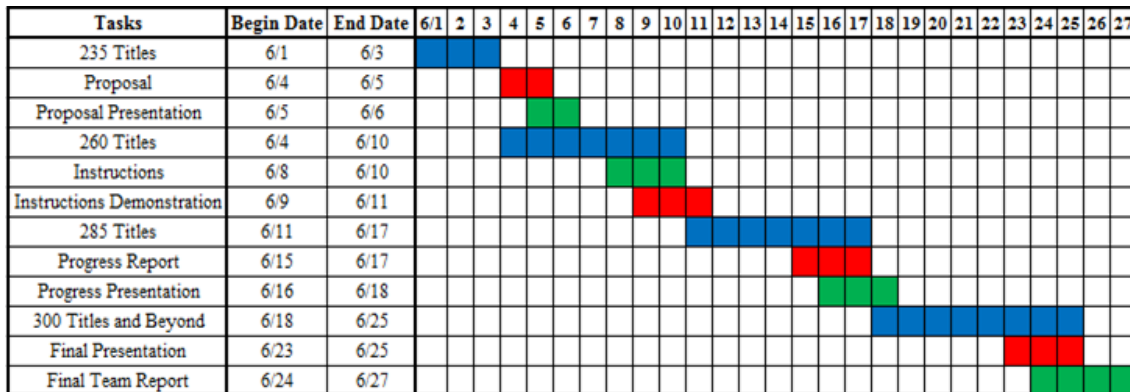


Figure 1: Our original Gantt chart.

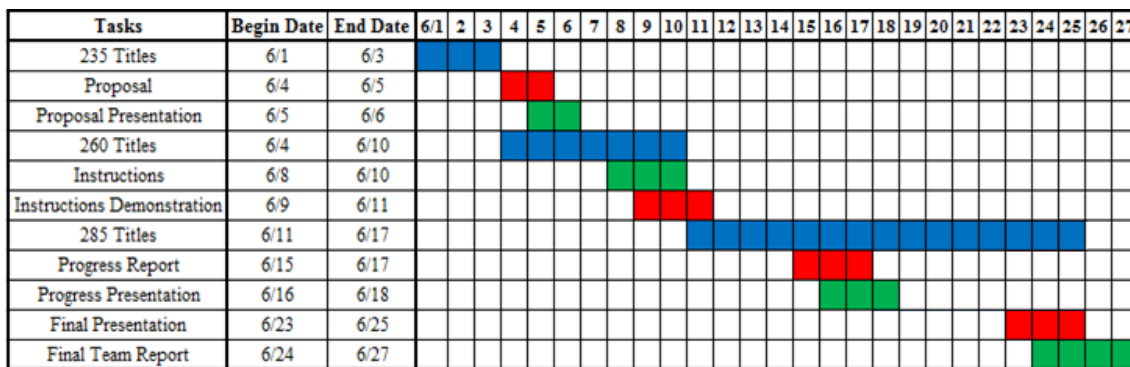


Figure 2: Our revised Gantt chart.

| | |
|-----------------|---|
| Key: | |
| Eugene/Sung | █ |
| Shashank/Yiming | █ |
| Everyone | █ |

2.3 Projections

We expect that we will accomplish our revised 285 title goal for June 25 and propose 2 simple approaches to this goal. Firstly, although we have devoted much computational power toward extending the longest chain at either end, we have not considered a potentially large number of alternative paths between nodes in the chain. Thus, we will attempt to compute, for adjacent or nearby nodes, alternative (longer) chains connecting the same nodes. Secondly, we are considering ways to “trim” the graph by removing nodes unlikely to be in the longest chain, thus reducing the number of possible paths that must be considered.

If simple methods prove insufficient, we may also implement a color-coding algorithm proposed by Alon et al.¹ or the genetic algorithms studied by Portugal et al.,² although, due to the implementation complexity of these algorithms, we hope this will be unnecessary.

3 Recomendations

Our only recomendation is the change of the final project goal from 300 titles to 285 titles, as discussed in the Progress section above.

4 Discussion

Our final deliverable will now be a chain of at least 285 movie titles rather than a chain of at least 300 movie titles. If, however, we achieve our 285 title goal before our deadline of June 25, we may spend the remaining time attempting to extend the chain further.

¹Noga Alon , Raphael Yuster , Uri Zwick, Color-coding, Journal of the ACM (JACM), v.42 n.4, p.844-856, July 1995. (accessed June 16, 2013 at <http://dl.acm.org/citation.cfm?id=210337>).

²D. Portugal, C. H. Antunes, R. Rocha, “A Study of Genetic Algorithms for Approximating the Longest Path in Generic Graphs,” Proc. of the IEEE SMC, pp. 2539-2544, 2010. (accessed June 16, 2013 at <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5641920& navigation=1>).

Appendices

A Our Longest Movie Chain

Here, we include our current longest movie chain, with 278 titles and 885 words:

| | |
|----------------------------|---------------------------------|
| THE RESCUERS | END OF DAYS |
| THE RESCUERS DOWN UNDER | DAYS OF HEAVEN |
| UNDER CAPRICORN | HEAVEN CAN WAIT |
| CAPRICORN ONE | WAIT UNTIL DARK |
| ONE NIGHT STAND | DARK CITY |
| STAND IN | CITY BY THE SEA |
| IN OLD CALIFORNIA | SEA OF LOVE |
| CALIFORNIA SPLIT | LOVE AND DEATH |
| SPLIT SECOND | DEATH BECOMES HER |
| SECOND BEST | HER MAJESTY MRS BROWN |
| BEST OF THE BEST | BROWN SUGAR |
| THE BEST OF EVERYTHING | SUGAR AND SPICE |
| EVERYTHING RELATIVE | SPICE WORLD |
| RELATIVE FEAR | WORLD TRADE CENTER |
| FEAR STRIKES OUT | CENTER STAGE |
| OUT OF THE PAST | STAGE FRIGHT |
| PAST MIDNIGHT | FRIGHT NIGHT |
| MIDNIGHT RUN | NIGHT AND THE CITY |
| RUN SILENT RUN DEEP | CITY OF JOY |
| DEEP BLUE | JOY RIDE |
| BLUE CAR | RIDE THE HIGH COUNTRY |
| CAR 54 WHERE ARE YOU | COUNTRY LIFE |
| YOU CANT TAKE IT WITH YOU | LIFE IS BEAUTIFUL |
| YOU LIGHT UP MY LIFE | BEAUTIFUL GIRLS |
| MY LIFE WITHOUT ME | GIRLS GIRLS GIRLS |
| ME MYSELF I | GIRLS JUST WANT TO HAVE FUN |
| I SPY | FUN AND FANCY FREE |
| SPY HARD | FREE WILLY |
| HARD TIMES | FREE WILLY 2 THE ADVENTURE HOME |
| TIMES SQUARE | HOME ALONE |
| SQUARE DANCE | ALONE IN THE DARK |
| DANCE WITH A STRANGER | THE DARK HALF |
| STRANGER IN THE HOUSE | HALF LIGHT |
| HOUSE OF DRACULA | LIGHT OF DAY |
| DRACULA DEAD AND LOVING IT | DAY FOR NIGHT |
| IT TAKES TWO | NIGHT OF THE LIVING DEAD |
| 24 7 TWENTY FOUR SEVEN | DEAD HEAT |
| SEVEN YEARS IN TIBET | HEAT AND DUST |
| TIBET CRY OF THE SNOW LION | DUST TO GLORY |
| LION OF THE DESERT | GLORY ROAD |
| DESERT BLUE | ROAD GAMES |
| BLUE STEEL | GAMES PEOPLE PLAY NEW YORK |
| STEEL DAWN | NEW YORK NEW YORK |
| DAWN OF THE DEAD | NEW YORK COP |
| DEAD BANG | COP LAND |
| BANG BANG YOURE DEAD | LAND OF THE DEAD |
| DEAD END | DEAD MAN |

DEAD MAN ON CAMPUS
CAMPUS MAN
MAN OF THE HOUSE
HOUSE OF FRANKENSTEIN
FRANKENSTEIN AND THE MONSTER FROM HELL
HELL NIGHT
NIGHT FALLS ON MANHATTAN
MANHATTAN MURDER MYSTERY
MYSTERY ALASKA
ALASKA SPIRIT OF THE WILD
THE WILD ANGELS
ANGELS WITH DIRTY FACES
FACES OF DEATH
DEATH SHIP
SHIP OF FOOLS
FOOLS RUSH IN
IN COLD BLOOD
BLOOD BEACH
BEACH PARTY
PARTY GIRL
GIRL IN THE CADILLAC
CADILLAC MAN
MAN ON FIRE
FIRE IN THE SKY
SKY HIGH
HIGH CRIMES
CRIMES OF PASSION
PASSION IN THE DESERT
DESERT HEARTS
HEARTS OF DARKNESS A FILMMAKERS APOCALYPSE
APOCALYPSE NOW
NOW YOU SEE HIM NOW YOU DONT
DONT BOTHER TO KNOCK
KNOCK OFF
OFF THE BLACK
BLACK AND WHITE
WHITE LIGHTNING
LIGHTNING IN A BOTTLE
BOTTLE ROCKET
ROCKET MAN
MAN TROUBLE
TROUBLE EVERY DAY
DAY OF THE DEAD
DEAD OF NIGHT
NIGHT MOTHER
MOTHER JUGS AND SPEED
SPEED 2 CRUISE CONTROL
CONTROL ROOM
ROOM AT THE TOP
TOP GUN
GUN CRAZY
CRAZY AS HELL
HELL UP IN HARLEM
HARLEM RIVER DRIVE
DRIVE ME CRAZY

CRAZY PEOPLE
PEOPLE I KNOW
I KNOW WHAT YOU DID LAST SUMMER
SUMMER CATCH
CATCH A FIRE
FIRE ON THE MOUNTAIN
THE MOUNTAIN MEN
MEN CRY BULLETS
BULLETS OVER BROADWAY
BROADWAY DANNY ROSE
ROSE RED
RED EYE
EYE FOR AN EYE
AN EYE FOR AN EYE
EYE OF GOD
GOD IS GREAT IM NOT
NOT OF THIS EARTH
EARTH GIRLS ARE EASY
EASY MONEY
MONEY FOR NOTHING
NOTHING BUT TROUBLE
TROUBLE IN PARADISE
PARADISE ROAD
ROAD HOUSE
HOUSE PARTY
PARTY MONSTER
MONSTER IN A BOX
BOX OF MOON LIGHT
LIGHT IT UP
UP CLOSE AND PERSONAL
PERSONAL BEST
BEST MEN
MEN WITH GUNS
GUNS OF THE MAGNIFICENT SEVEN
THE MAGNIFICENT SEVEN RIDE
RIDE WITH THE DEVIL
THE DEVIL RIDES OUT
OUT COLD
COLD FEVER
FEVER PITCH
PITCH BLACK
BLACK HAWK DOWN
DOWN WITH LOVE
LOVE LIFE
LIFE OR SOMETHING LIKE IT
IT HAPPENED AT THE WORLDS FAIR
FAIR GAME
GAME OF DEATH
DEATH WISH V THE FACE OF DEATH
DEATH WISH
WISH UPON A STAR
A STAR IS BORN
BORN AMERICAN
AMERICAN HISTORY X
X THE MAN WITH THE X RAY EYES

EYES OF AN ANGEL
ANGEL BABY
BABY SECRET OF THE LOST LEGEND
LEGEND OF THE LOST
THE LOST BOYS
BOYS AND GIRLS
GIRLS WILL BE GIRLS
GIRLS OF SUMMER
SUMMER LOVERS
LOVERS AND OTHER STRANGERS
STRANGERS WHEN WE MEET
MEET JOE BLACK
BLACK LIKE ME
ME WITHOUT YOU
YOU ONLY LIVE ONCE
ONCE AROUND
AROUND THE BEND
BEND OF THE RIVER
THE RIVER WILD
WILD THINGS
THINGS TO COME
COME AND GET IT
IT HAPPENED ONE NIGHT
ONE NIGHT WITH THE KING
THE KING AND I
I WANT TO LIVE
LIVE AND LET DIE
DIE MOMMIE DIE
DIE MONSTER DIE
DIE HARD
HARD EIGHT
EIGHT AND A HALF WOMEN
WOMEN IN LOVE
IN LOVE AND WAR
WAR OF THE WORLDS
THE WORLDS FASTEST INDIAN
INDIAN SUMMER

SUMMER SCHOOL
SCHOOL OF ROCK
ROCK STAR
STAR TREK THE MOTION PICTURE
PICTURE BRIDE
BRIDE OF THE WIND
THE WIND AND THE LION
THE LION KING
KING OF THE JUNGLE
JUNGLE 2 JUNGLE
JUNGLE BOOK
BOOK OF LOVE
LOVE WALKED IN
IN GODS HANDS
HANDS ON A HARD BODY
BODY DOUBLE
DOUBLE TEAM
TEAM AMERICA WORLD POLICE
POLICE ACADEMY
POLICE ACADEMY 3 BACK IN TRAINING
TRAINING DAY
DAY OF THE WOMAN
THE WOMAN IN RED
RED RIVER
RIVER OF NO RETURN
RETURN TO HORROR HIGH
HIGH SCHOOL HIGH
HIGH SPIRITS
SPIRITS OF THE DEAD
DEAD MAN WALKING
WALKING AND TALKING
TALKING ABOUT SEX
SEX AND THE OTHER MAN
MAN OF THE YEAR
YEAR OF THE DRAGON
DRAGON SEED
SEED OF CHUCKY