LiChess Piece Counter

Andy Vo and Daniel Schultz

***Abstract*— Chess is a popular strategy-based board game between two players and is derived from the ancient Indian game, Chaturanga. It consists of 32 pieces, 16 pieces for each player: one king, one queen, two rooks, two knights, two bishops, and eight pawns. The board is a square chessboard with an eight-by-eight grid of 64 squares. The objective of the game is to "checkmate" the opponent's king, where the king is under attack and can not move any pieces for its escape. Organized competitive chess began in the 19th century, where the matches and the moves of each player were recorded. We intend to analyze official chess matches among high-ranked players using LiChess's API and database to record data such as how many pieces they won, how many pieces they survived with, and possibly other statistics if achievable within the time frame.**

I. OBJECTIVES

Our main objective is to analyze official chess matches among high-ranked chess players using LiChess's API and database to gather data on statistics such as how many pieces they won and how many pieces they survived with. Additional statistics that we would like to research for the project, but cannot guarantee, include win rate percentages based on the opening, the number of moves in a game before a piece was traded, and the number of sacrifices used per match to gain an advantage.

II. MOTIVATION

After the introduction of computers, the world of chess was thoroughly explored through many individuals studying chess theory. People were able to develop engines like Stockfish that can analyze any moment in a game of chess to compute the next best set of possible moves for the player and the opponent's best response. Through these engines, many people were able to analyze important statistics that contribute to the progression of the game such as openings, strategies, and sequences of moves. However, our group was more interested in statistics that hold no meaningful value, yet satisfied the curiosity of the mind. In their journey to become Grandmaster in the realm of chess, how many pieces did they win, how many pieces did they sacrifice, how many pieces did they survived with? We wanted to find the answer for many high-ranked chess players, so our goal is to uncover this information.

III. DISCUSSION OF DATA

Portable Game Notation (PGN) is a standard plain text format for chess games that contain information such as moves and the players' respective colors. This format can be seen in many chess databases, as such, we will be analyzing matches that are in this format. The data is ordered by the following structure: Event, Site, Date, Round, White, Black, Result, and Movetext. Event is the name of the tournament or match event. Site has the location of the match and is in **City, Region COUNTRY** format. Date is the starting date of the game and is in **YYYY.MM.DD** format. Round represents the specific round in terms of the organization of the event. For example, if the semi-finals of a chess tournament were the 29th match, the round would contain the number 29. White is the player who went first and represents the white side. Black is the player who went second and represents the black side. Result is the result of the match in the format: White's score, dash, and Black's score. It can also have a * to represent an ongoing match. Movetext records the actual sequence of moves in the chess match. The moves are recorded by the move number and are in Standard Algebraic Notation. SAN is the standard for recording chess moves and consists of the abbreviation for a piece, an **x** if a capture took place, and the name of the final square the piece moved to in two-character algebraic notation. There are also optional tags people can use for clarity's sake. An example of a PGN file can be seen below:

```
[Event "Hoogovens Group A"]
[Site "Wijk aan Zee NED"]
[Date "1999.01.20"]
[EventDate "1999.01.16"]
[Round "4"]
[Result "1-0"]
[White "Garry Kasparov"]
[Black "Veselin Topalov"]
[ECO "B07"]
[WhiteElo "2812"]
[BlackElo "2700"]
[PlyCount "87"]
1. e4 d6 2. d4 Nf6 3. Nc3 g6 4. Be3 Bg7 5. Qd2 c6
6. f3 b5 7. Nge2 Nbd7 8. Bh6 Bxh6 9. Qxh6 Bb7
10. a3 e5 11. O-O-O Qe7 12. Kb1 a6 13. Nc1 O-
O-O 14. Nb3 exd4 15. Rxd4 c5 16. Rd1 Nb6 17. g3
Kb8 18. Na5 Ba8 19. Bh3 d5 20. Qf4+ Ka7 21. Rhe1
d4 22. Nd5 Nbxd5 23. exd5 Qd6 24. Rxd4 cxd4 25.
Re7+ Kb6 26. Qxd4+ Kxa5 27. b4+ Ka4 28. Qc3
Qxd5 29. Ra7 Bb7 30. Rxb7 Qc4 31. Qxf6 Kxa3 32.
Qxa6+ Kxb4 33. c3+ Kxc3 34. Qa1+ Kd2 35. Qb2+
Kd1 36. Bf1 Rd2 37. Rd7 Rxd7 38. Bxc4 bxc4 39.
Qxh8 Rd3 40. Qa8 c3 41. Qa4+ Ke1 42. f4 f5 43.
Kc1 Rd2 44. Qa7 1-0
```

Fig. 1. Kasparov vs. Topalov, Wijk ann Zee 1999

In terms of the scope of the project, it depends on what is available on the internet. There are a variety of chess

databases that we can utilize, and the scope of the project depends on what we can legally use. We would like to analyze the matches of all official World Champions when their rating was International Master (2400) or higher if possible. We would also like to analyze the recorded matches of all Grandmasters in the past 20 years. As a last resort, LiChess has its own database of chess matches played on their site, and we can use their matches of players on their site that are International Masters or higher.

IV. RESPONSIBILITIES AND ROLES OF EACH MEMBER

A. Andy Vo

- Find different chess databases to analyze the matches
- Develop algorithms and data structures for the data
- Create a visualization of results from the data

B. Daniel Schultz

- Clean and parse the PGN's data
- Assist in developing algorithms and data structures
- Assist in visualization of results from the data

Both members will be responsible for the final presentation and research paper.

V. TIMELINE OF MILESTONES

The milestones we aim to achieve for this project are to research chess databases, clean and parse the data, develop the code to analyze the data, and analyze and visualize the data. We also have some optional statistics that we may explore, but ultimately, it depends on the data we manage to receive.

According to the approximate time frame of the final project, we have six weeks to complete our project. The following table is a timeline for our project:

TABLE I
TIMELINE OF MILESTONES

Week	Milestones
Week 1:	Research Chess Databases
Week 2:	Clean and Parse the Data
Week 3:	Develop Data Structures and Algorithms to Analyze the Data
Week 4:	Optional Additional Statistical Research
Week 5:	Analyze and Visualize the Data
Week 6:	Final Report and Presentation

VI. EXPECTED OUTCOME

After obtaining and analyzing our data, we expect to obtain various statistics that symbolize the path of a Grandmaster such as how many pieces they won, how many pieces they sacrificed, and how many pieces they had left. Considering the skills of a Grandmaster, we anticipate the average of these numbers to be lower compared to the average player. We expect this data to cater to a niche audience who would be interested in non-meaningful yet intriguing data concerning top chess players. Also, we hope this data can assist in achieving a higher level of understanding about the game of Chess.