

Teoría de Lenguajes - Trabajo práctico - Validación de archivos PGN

Segundo cuatrimestre de 2021

1. Introducción

Mediante un programa, se puede analizar sintácticamente un archivo en formato PGN (*Portable Game Notation*). Un archivo en este formato es un archivo Ascii que contiene una secuencia de partidas de ajedrez, con datos sobre las mismas como jugadores, eventos involucrados, jugadas y comentarios a lo largo de cada partida.

Sigue un ejemplo de archivo en formato PGN que contiene una sola partida:

```
[Event "Mannheim"]
[Site "Mannheim GER"]
[Date "1914.08.01"]
[EventDate "1914.07.20"]
[Round "11"]
[Result "1-0"]
[White "Alexander Alekhine"]
[Black "Hans Fahrni"]
[ECO "C13"]
[WhiteElo "?"]
[BlackElo "?"]
[PlyCount "45"]

1. e4 {Notes by Richard Reti} 1... e6 2. d4 d5 3. Nc3 Nf6 4. Bg5 Be7 5. e5 Nfd7
6. h4 {This ingenious method of play which has subsequently been adopted by all
modern masters is characteristic of Alekhine's style.} 6... Bxg5 7. hxg5 Qxg5
8. Nh3 {! The short-stepping knight is always brought as near as possible to
the actual battle field. Therefore White does not make the plausible move 8 Nf3
but 8 Nh3 so as to get the knight to f4.} 8... Qe7 9. Nf4 Nf8 10. Qg4 f5 {The
only move. Not only was 11 Qxg7 threatened but also Nxd5.} 11. exf6 gxf6 12.
0-0-0 {He again threatens Nxd5.} 12... c6 13. Re1 Kd8 14. Rh6 e5 15. Qh4 Nbd7
16. Bd3 e4 17. Qg3 Qf7 {Forced - the sacrifice of the knight at d5 was
threatened and after 17...Qd6 18 Bxe4 dxe4 19 Rxe4 and 20 Qg7 wins.} 18. Bxe4
dxe4 19. Nxe4 Rg8 20. Qa3 {Here, as so often happens, a surprising move and one
difficult to have foreseen, forms the kernel of an apparently simple Alekhine
combination.} 20... Qg7 {After 20.Qe7 21.Qa5+ b6 22.Qc3 would follow.} 21.
Nd6 Nb6 22. Ne8 Qf7 {White mates in three moves.} 23. Qd6+ 1-0
```

A los efectos de este TP, consideraremos un PGN simplificado y muy permisivo.

2. Formato

A continuación describimos el formato de cada partida del archivo.

Primero aparecen entre corchetes los descriptores del evento, como nombre, lugar, fecha, nombres de los jugadores y resultado de la partida, entre otros, de a un descriptor por línea. En nuestra simplificación, se puede asumir que los descriptores son cadenas arbitrarias y sus valores asociados son cadenas arbitrarias entre comillas.

Sigue luego la secuencia de jugadas de la partida, numeradas en forma creciente como se ve en el ejemplo. La última jugada de una partida podría ser del blanco o del negro, y sigue luego el resultado de dicha partida, que puede ser “1-0” si ganó el blanco, “0-1” si ganó el negro, o “1/2-1/2” si empataron (tablas).

A lo largo de las jugadas podría haber comentarios, que serán secuencias de caracteres encerradas entre llaves o paréntesis, los cuales podrían incluir jugadas (que representan líneas de juego alternativas). Estos paréntesis y llaves con comentarios pueden aparecer en forma anidada, debiendo respetarse el balanceo.

Cada jugada consiste en la inicial de la pieza que se mueve (P: peón, N: caballo, B: alfil, R: torre, Q: dama, K: rey) seguida de las coordenadas de la casilla a la cual se mueve (a1, a2, ... h8, con una letra de “a” a “h” para indicar la columna y un dígito del 1 al 8 para indicar la fila). Podría no especificarse la inicial de la pieza (en ese caso el que mueve es un peón). Existen asimismo jugadas especiales, como “O-O” y “O-O-O” (los enroques)¹. Tras la inicial de la pieza que mueve, opcionalmente podría haber una letra de “a” a “h” (identificando su columna de origen) y/o un número del 1 al 8 (identificando su fila de origen), por ej. “Ndf6”. Si la jugada es de captura, se inserta el símbolo x antes de la casilla destino (por ej., exf4 significa que un peón ubicado en la columna “e” captura a una pieza contraria ubicada en la casilla f4). Si la jugada es un jaque (i.e. produce una amenaza al rey contrario), va seguida del símbolo +. Si la jugada es un jaque mate (i.e. produce una amenaza al rey contrario de la que este no puede salir en la jugada siguiente), va seguida de ++.² También es posible que la jugada tenga una anotación en la forma de un signo de admiración (!) para indicar que se considera que fue buena, o un signo de interrogación (?) para indicar que fue mala. Una jugada podría estar precedida por un número y uno o tres puntos. Si los puntos son tres, el número deberá coincidir con el inmediato anterior.

No se pide validar si la secuencia de jugadas respeta las reglas del ajedrez, pero sí que la numeración de las jugadas sea correcta. Por ende, algo como

1. e4 e4 3. e5 e5 4. e4 0-1

no será aceptable porque falta el movimiento 2, y tampoco algo como

1. e4 e5 { Nf3 1-0

porque falta cerrar una llave, mientras que

1. e4 e5 2. Nf3 { Nh3 se considera mala } 2... Nc6 1/2-1/2

sí es aceptable. En cada partida los números deben aparecer en forma creciente cada dos jugadas, comenzando con el 1 que antecede a la primera de la partida. Fuera de los comentarios, si una jugada del blanco va antecedita de número y punto, la que le sigue, del negro, no deberá tener número ni puntos o bien podrá tener el mismo número que la última del blanco seguido de tres puntos. No puede haber comentarios entre los números y los puntos, ni entre los puntos, ni entre los puntos y la jugada que le sigue. Adentro de cualquier comentario los números y puntos no serán relevantes.

La entrada será una cadena de caracteres a leer de la entrada estándar, y la salida debe ser la estándar. Si se produce algún error durante el análisis, se exhibirá con un mensaje claro (por ejemplo, casilla inválida, falta el resultado de la partida al final, llave o paréntesis no cierra, número de jugada inválido, etc.) Todo lo demás, como los descriptores que anteceden a cada partida, se puede asumir que son cadenas de caracteres arbitrarias.

¹Existen otras jugadas especiales como la coronación, que no hará falta tener en cuenta. A saber, un movimiento de peón hacia la fila 1 u 8 debe ser seguido del símbolo = y la inicial de una pieza: N, B, R, Q que es la que reemplazará al peón.

²Para este TP no se necesita verificar nada especial sobre las jugadas fuera de validar la sintaxis de cada una en forma independiente de las otras y que los números aparezcan en forma creciente. No hace falta verificar que cada + se corresponda efectivamente con un jaque, ni que cada x se corresponda efectivamente con una captura, ni que las jugadas sean reglamentarias.

3. Objetivo del trabajo

El objetivo del trabajo es, por un lado, verificar la sintaxis (superficial) de un archivo PGN informando si este es válido y, por otro, determinar el máximo nivel de comentarios en que no haya capturas, considerando todas las partidas.

3.1. Ejemplo 1

[a "b"]

1. e4 d5 {defensa escandinava (es común 2. exd5 Da5 {no es común 2... c6})} 1/2-1/2

Máximo nivel sin capturas: 3.

3.2. Ejemplo 2

[prueba "loca"]
[Nzscf5qWgtgÑVX "56BññQIeAhy"]
[gvk7dXkliRpR "2LAKQJGhz81"]
[ÑFS5lBHW4MmÑmJsP "e4ZhVulzl"]
[yZlPSI4r78KP "XwWzscEtUqkAuñNt7Hq5"]
[GARz0dNañITcsbF09ES "WUodxeqxI"]

1. 1d7 d7 2. Rf8 gg7 1-0

[2ujZrN6LTmOBGss5jzqw "pnAGk"]
[kRM "lihgftp5kLaiAvuNSub2"]
[vUPhgAiKx "a9C7is0kq0vyep7svNE"]
[Lbfh "j6htrD1wKNFryNGHdUcG"]

1. Pxe7 Pad3 2. Be6 b6 3. Bb2 h3 4. Qc1 Kh3 5. Pb1 Pb2 6. Ra1 a7! 7. Nce4 Qxa3 8. Kd4 Qh1 9. Kd8 fh6 10. Nc1 Pg1 11. Nh2 Bf4 12. ca6 Kd8 13. Nb4 d5 14. e1 Ra5 15. f8 K5xd3 16. b6 (Ka4 h 5a2) 16... g7 17. e5 Bg2 18. f3 Re2 19. Qc6 Kb4 20. Rg4 Rg3 21. g3 N2c7 22. Pb7 Rh7 23. Qd3 {bf8 B b3 K Nxf5 wuFzW c1 Pbd2 T c8 ka Pg3 K Ke6+ n Kg5 Pb7 f8 e2 0 1c5 Rxe4 Pxc5} 23... e6 24. d5 Nh7 25. Qxc6 Bxc3 26. Bf5 Qfa3 27. f4 Ka7 28. Nxb6 d6 29. Bg1 Qf2 30. Pb3 d6 31. Ba3 e7 32. Bab4 Kda1 33. Ph1 e2 34. cf4 Nxd7 35. Nc7 (gr Pfg2 iONp utl Pcd2 t (Pd3 Rc2 Bxd6 Kxg8 Kxh2 h6 Bxc8 zAx g2 qX Nh2 N c3 ñ Bxb5 D Nd1 h8 a1 c7 PVA0 Rg7 g5 Kb6 oGq) rt Pf8 Qp) 35... e2 36. Ra6 Bhb8 37. Q5a4 c5 38. Qg7 g1 39. g6 Pg7 40. Kd6 d1 41. h7+ Kcc5 42. Qd8 ce1 43. Qxh3 Pa6 44. Bxe2 e4 45. Bh7 d2+ 1/2-1/2

[BYsR4sFynE2Rñ2 "lNyKA4WPhvh2kKPTjfaD"]
[loCuhQ "WDOTwZIISKjDikñ6"]

1. Pb7 ec6 2. f4 Nxf7 3. Kxe5 d7 4. Kc7 Rd6 5. e7 Qc7 6. h1 {s Rxb7 iI d5 Nh7} 6... e4+ 7. 7d6 Nf7a3 8. Pa7 e2 9. h8 h4 10. g8 Kxf6 11. 2b3 Bxg3 12. Q5h1 Pb6 13. fd4 e5 14. 5c6 f3 15. cb6 {1 Pb6} 15... Pb1 16. Nb3 h7 17. b2 Nb7 18. Ra1 N8h6 19. f2 c8 20. P5g1 e1 21. Pg3 c3 22. f5 Kh1 23. Pd7 Nf3 24. Nd1 Rd1 25. Kd5 Ng8 26. e3 Nxg6 27. Rh8 fc3 28. Kxf1 Ncc4 29. c1+ e3 30. cc6 (a Kdxa3 e7 JQ Qe2 VLzeFq c7 FU f4 N Qxb7 Nd8++ Tf Nd2 yN ba2 me 1b6 hu Kd7 R Pxe6 Nxg7 DaR Ba8 Kxd5 Nd7 ttxI fe7 GK a2 m f6 Skp) 30... h8 31. Qh1 a1 32. fb7 b7 0-1

[4e "ws20"]

1. Bxd1 Pa6 0-1

Máximo nivel sin capturas: 1.

4. Herramientas

Se puede utilizar una herramienta de generación de analizadores léxicos como LEX, y una herramienta de generación de parsers como YACC, ANTLR o PLY.

5. A presentar

- Una descripción de la solución adoptada, el método utilizado, decisiones tomadas, tipo de gramática y/o parser, con los comentarios necesarios.
- Código de la solución en C++, C#, Java, Python o Go, con comentarios relevantes, más el código ingresado a la herramienta de parsing en caso de usar alguna. Si desean utilizar otro lenguaje, u otra herramienta, consulten con el JTP.
- Ejemplos de corrida, con entradas y salidas acordes, preferentemente mediante texto capturado.
- Conclusiones: dificultades encontradas, eficiencia, posibilidades adicionales, relevancia de los temas vistos en la materia y todo lo que consideren relevante. Se espera algo técnico sobre el contenido, los problemas específicos que encontraron y las decisiones concretas tomadas.

Grupos: de 3 personas.

Modo de entrega: por e-mail a tptleng@gmail.com hasta el día 29/11, en un archivo .rar o .zip., con *subject:* entrega de TP grupo...

Referencias

- [1] Alfred V. Aho, Ravi Sethi and Jeffrey D. Ullman. Compilers, Principles, Techniques and Tools (2nd edition). Addison-Wesley, Reading, Massachusetts, 2006.
- [2] Stephen C. Johnson. Yacc: Yet Another Compiler Compiler. Computing Science Technical Report No. 32, Bell Laboratories, Murray Hill, New Jersey, 1975.
- [3] Jim Gardner, Chris Retterath and Eric Gisin. MKS Lex & Yacc. Mortice Kern Systems Inc., Waterloo, Ontario, Canada, 1988.
- [4] M. E. Lesk, and E. Schmidt. Lex - A Lexical Analyzer Generator. Computing Science Technical Report No. 39, Bell Laboratories, Murray Hill, New Jersey, 1975.
- [5] John R. Levine, Tony Mason and Doug Brown. Lex & Yacc. O'Reilly & Associates, Inc. Sebastopol, California, 1992.
- [6] Sobre el formato PGN:
 - <http://www.saremba.de/chessgml/standards/pgn/pgn-complete.htm>
 - <http://www.pgnmentor.com>
 - [https://en.wikipedia.org/wiki/Algebraic_notation_\(chess\)](https://en.wikipedia.org/wiki/Algebraic_notation_(chess))