

# Teoría de Lenguajes - Trabajo práctico

Primer cuatrimestre de 2020

## Obtención de información a partir de archivos PGN

### 1. Introduction

**Objetivo:** Mediante un programa, analizar sintácticamente un archivo en formato PGN (Portable Game Notation). Un archivo PGN es una colección de partidas de ajedrez con datos sobre las mismas, como los jugadores, eventos involucrados y comentarios a lo largo de las jugadas. A los efectos de este TP, consideraremos un PGN simplificado.

Sigue un ejemplo de archivo en formato PGN que contiene 1 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
```

Para cada partida aparecen primero entre corchetes los descriptores del evento, como nombre, lugar, fecha, nombres de los jugadores y resultado de la partida, entre otros. En esta 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, eventualmente numeradas, como se ve en el ejemplo. La última jugada de una partida podría ser del blanco o del negro, para terminar con 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 juga-

das que representan líneas de juego alternativas. Estos paréntesis y llaves con comentarios pueden aparecer en forma anidada, debiendo respetarse el balanceo.

### 1.1. Formato de cada jugada

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)<sup>1</sup>. 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 del símbolo #.<sup>2</sup> Una jugada podría estar precedida por un número y uno o tres puntos.

## 2. Objetivo del trabajo

El objetivo del trabajo es, por un lado, verificar la sintaxis (superficial) de un archivo PGN, informando si es válido. Y, por otro, resolver lo siguiente:

- Determinar cuál es la jugada que más se repite como primera jugada del jugador blanco a lo largo de todo el archivo (principal, fuera de comentarios).<sup>3</sup>
- Determinar, a lo largo del archivo, el máximo nivel de anidamiento de comentarios que contengan al menos una jugada.

Por ej.,

```
e4 {acá un comentario {y otro que dice que es la más común,  
a diferencia de la poco usual e3 {que es mala} y la respuesta  
e5 es muy común} }
```

tiene nivel de anidamiento 2, porque el comentario más interno no incluye ninguna jugada.

No se pide validar si la secuencia de jugadas respeta las reglas del ajedrez ni si la numeración de las jugadas es correcta. Por ende, algo como

```
1. e4 3. e5 e5 3. e4 1-0
```

será aceptable, no así algo como

```
1. e4 e5 { Nf3 1-0
```

porque falta cerrar una llave.

La entrada será una cadena de caracteres a ser leída de la entrada estándar y la salida debe ser la estándar. Si se produce algún error durante el análisis,

<sup>1</sup>Existen 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.

<sup>2</sup>Para 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. 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.

<sup>3</sup>Asumir que cada posible primera jugada aparecerá escrita siempre de igual manera a lo largo de todo el archivo.

se exhibirá un mensaje claro señalando el mismo (por ejemplo, casilla inválida, falta el resultado de la partida al final, paréntesis no cierra, etc.) Todo lo demás, como los nombres de los jugadores y del evento, se podrá asumir que son simples cadenas de caracteres. Recordar que puede haber jugadas dentro de los comentarios anidados.

### 3. 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.

### 4. A presentar

- Una descripción de la solución adoptada, el método utilizado, decisiones tomadas, tipo de gramática y/o parser, con los comentarios que se crean 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, relevancia de los temas vistos en la materia, eficiencia, posibilidades adicionales 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](mailto:tptleng@gmail.com) hasta el día 6/8, en un archivo .rar o .zip.

*Subject:* entrega de TP.

### 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://portablegamenotation.com/FIDE.html>
  - <http://www.saremba.de/chessgml/standards/pgn/pgn-complete.htm>
  - <http://www.pgnmentor.com>