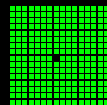


# Analizar, no validar (Parse, don't validate)

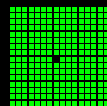
Ramón Soto **Mathiesen**, SPISE MISU ApS

2025-03-07 @ OIFem Final (Barcelona | UPC-CFIS | Edición virtual)



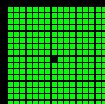
# Programa

- Sobre mi ... y que hago
- Analizar, no validar
- Combinador de analizadores sintácticos
- Demostración de parser combinators
  - Valores booleanos en C vs F#



# Sobre mi ...

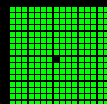
- **Datólogo** (estudioso de datos) y **matemático** de las **Universidades** de **Copenhague** (Dinamarca) y **Pisa** (Italia)
- **Enseñaba a niñ@s** de entre 7-17 años a **programar** en Coding Pirates en **Copenhague** (Dinamarca)
- Inicios (Amstrad CPC 464) y **programar sin estructura** (BASIC y GOTO)
- Programación **imperativa** y **orientado a objetos**: (C/C++/C#, Java, JavaScript, Python, TypeScript, ...)
- Programación **Funcional** (Lisp, OCaml, F#, Elm, ...)
- **Teoría de categorías** (Haskell, ...)
  - «Las Matemáticas, de las Matemáticas» frase de Eugenia Cheng en su libro: The Joy of Abstraction, ISBN: 9781108477222
- SPISE MISU ApS utiliza este tipo de tecnologías (I, OO pero sobre todo PF y TC)



# ... y que hago (OMXC25 → IBEX35)

The image shows a video player interface. The main content is a presentation slide with the following text: "HOW TO TRANSFORM AVRO (IDL) DATA TO MULTIPLE PARQUET FILES" and "PANDORA" with a crown icon above the 'O'. To the right of the slide is a yellow circle with a black 'F' inside. Below the slide is a yellow banner with the text "FOSS NORTH 2023" and "How to transform AVRO (IDL) data to multiple PARQUET files" by "Ramon Soto Mathiesen". The video player controls at the bottom show a play button, a progress bar at 0:00 / 57:56, and various icons for volume, subtitles, and full screen. To the right of the video player is a social media post from "Ramón Soto Mathiesen" with the text "Hmmm, that code looks familiar" and a link "https://lnkd.in/dn7CcKYd". Below the post is a video thumbnail with the text "BUT WHAT IS IT LIKE TO WORK AT PANDORA?". The social media interface shows 2 comments, 27 reactions, and 3,547 impressions.

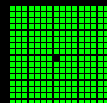
Código abierto para la Joyería más grande del Mundo



# Día Internacional de la Mujer (mañana)

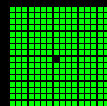


Se aprobó en la 2 Conf. Int. de Mujeres CPH en 1910



# Analizar, no validar (Parse, don't validate)

- Es una frase que utilizamos para explicar que, una vez que los **datos que ingresan al sistema** (input), **se analizan al tipo correcto** (parse), **el resto del sistema** puede **usar esos datos de manera segura y sin tener que hacer suposiciones**.

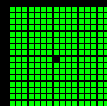


# Combinador de analizadores sintácticos (parser combinators)

- Un **combinador de analizadores sintácticos** es una **función de orden superior** (\*) que acepta **varios analizadores sintácticos como entrada y devuelve un nuevo como salida**.
- Por ejemplo, un **analizador sintáctico** podría ser una **función que aceptase cadenas** (`string / char[]`) como entrada y devolviese alguna **estructura** como salida.

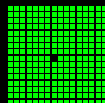
(\*) - Son **funciones que cumplen al menos** una de las siguientes condiciones:

- Tomar una o más funciones como entrada
- Devolver una función como salida



# Demostración de parser combinators

```
File Edit Options Buffers Tools YASnippet Complete In/Out Signals Help
1 #!usr/bin/env -S dotnet fsi --langversion:8.0 --optimize --warnaserror+:25,26
2
3 #time "on"
4
5 #r "nuget: SpiseMisu.ParserCombinator, 00.11.14"
6
7 open System
8 open SpiseMisu.Parser
9
10 let trueP : bool parser =
11   ( fun _ -> true
12     )
13   <|> choiceP [ charP '1'; stringP "true" ]
14
15 let falseP : bool parser =
16   ( fun _ -> false
17     )
18   <|> choiceP [ charP '0'; stringP "false" ]
19
20 let boolP : bool parser =
21   choiceP [ trueP; falseP ]
22
23 let boolsP : bool seq parser =
24   let auxP =
25     sepByP
26       ( boolP )
27       ( skipSpacesP *> charP ',' <* skipSpacesP )
28     charP '[' *> skipSpacesP *> auxP <* skipSpacesP <* charP ']'
29
30   let _ =
31
32     "1" // 1 -> Verdadero
33     |> runP boolP
34     |> printfn "%A"
35
36     "0" // 0 -> Falso
37     |> runP boolP
38     |> printfn "%A"
39
40     "true" // Verdadero
41     |> runP boolP
42     |> printfn "%A"
43
44     "false" // Falso
45     |> runP boolP
46     |> printfn "%A"
47
48     "-1" // ¿Ehhh?
49     |> runP boolP
50     |> printfn "%A"
51
52     "[ 1, 0,1,0,0,      true,false, false ,0,1,  true      ]"
53     |> runP boolsP
54     |> Result.map (Seq.toArray)
55     |> printfn "%A"
56
57     00
220 Ok true
221 Ok false
222 Ok true
223 Ok false
224 Error
225 "Parser error"
226 * Message.....: choiceP
227 * Offset.....: 0
228 * Unparsed string...: "-1"
229 "
230 Ok [|true; false; true; false; false; true; false; false; false; true; true|]
231 Real: 00:00:00.002, CPU: 00:00:00.000, GC gen0: 0, gen1: 0, gen2: 0
232 val trueP: bool parser = Parser <fun:fmapP@77>
233 val falseP: bool parser = Parser <fun:fmapP@77>
234 val boolP: bool parser = Parser <fun:choiceP@930>
235 val boolsP: bool seq parser = Parser <fun:lift2AP@95>
236
237 > |
UUU:*** F1 *inferior-fsharp* Bot (237,2) (Inferior fsharp:run yas Compilation) -----
1 #include <stdio.h>
2 #include <stdbool.h>
3
4 int main(int argc, char* argv[])
5 {
6     bool t = true;
7     bool f = false;
8
9     printf("True..... %d\n", t); // 1 -> Verdadero
10    printf("False.... %d\n", f); // 0 -> Falso
11
12    bool q = (bool) -1; // Verdadero o Falso o ...
13
14    printf("( -1).... %d\n", q); // 1 -> Verdadero ... ¿Ehhh?
15
16    return 0;
17 }
UUU:--- F1 bool.c All (14,57) (C*1 yas Abbrev) -----
18 [plg5:~/priv/spiseMisu/sponsorships/2025/oifem/src]$ ./build.sh && ./bool
19 True.... 1
20 False.... 0
21 (-1).... 1
22 [plg5:~/priv/spiseMisu/sponsorships/2025/oifem/src]$
UUU:--- F1 bool.fsx All (58,0) (fsharp dotnet yas Flymake 0 0 1) company EIDoc) eglot src -----
UUU:*** F1 *shell* Bot (18,53) (Shell:run yas) -----
```





# Resumen

- Analizar, no validar con librerías de combinador de analizadores sintácticos. Nos ayudará a evitar:
  - «Basura que entra, basura que sale» / «garbage in, garbage out»
- Inteligencia Artificial (IA):
  - Intentar adquirir la mayor cantidad de conocimientos posible y utilizar la IA con plantillas que debe completar (✓ en lo que es muy bueno) en vez de utilizar la IA para que te de las soluciones (✗ en lo que no es tan bueno, por ahora, a excepción de algunos casos como AlphaFold2 en CASP14 con una puntuación +90 en GDT)
- ¡¡¡ Finalmente, mucha suerte y sobre todo, pasarlo bien !!!
- Y feliz **Día Internacional de la Mujer** mañana 🎉🎈👏👏

