ALF

Expressions régulière et Lexer

Bibliographie pour aujourd'hui

ALF
language crunch

Keith Cooper, Linda Torczon, *Engineering a Compiler*

- Chapitre 2
  - 2.4
  - 2.5
  - 2.6

Alfred V. Aho, Monica S. Lam, Ravi Sethi, Jeffrey D. Ullman, *Compilers: Principles, Techniques, and Tools (2nd Edition)*

- Chapitre 3
  - 3.1
  - 3.3

Terence Parr, *The Definitive ANTLR 4 Reference*, (2nd Edition)

Contenu

ALF
language crunch

- Expressions régulières
- Lexer
- ANTLR

Donald Knuth

- Américain
- TeX
- The Art of Computer Programming
- Stanford

Expressions régulières

ALF
language crunch

- Une façon de décrire la forme d'une string
- Les string sont reconnaissables par un automate fini
- Une string avec des caractères spéciaux

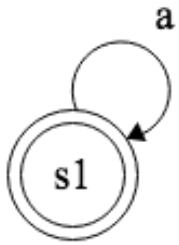
if / while / for

Syntaxe d'expression régulière

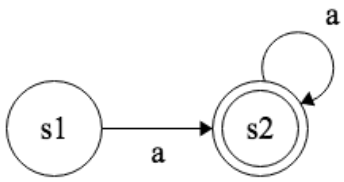
ALF
language crunch

Character	Description	Exemple	String
*	Zéro ou plusieurs fois	a*, (ab)*	aaaaaaaaa, abababababab
+	Une ou plusieurs fois	a+, (ab)+	aaaaaaa, abababababab
?	Zéro ou une fois	a?, (ab)?	a, ab
^	début de string	^ab*	abbbbbbb
$	fin de string	b*a$	bbbba
.	tout symbole	.	a, b, c
[ ]	Ensemble	[abc]	a, b
\s	Espace blanc	a\sb	a b
[^]	ensemble complémentaire	[^abc]	e, d
( )	groupe	(abc)+	Abcabcabc
|	Ou	a|b, (ab)|(ba)	a, ab

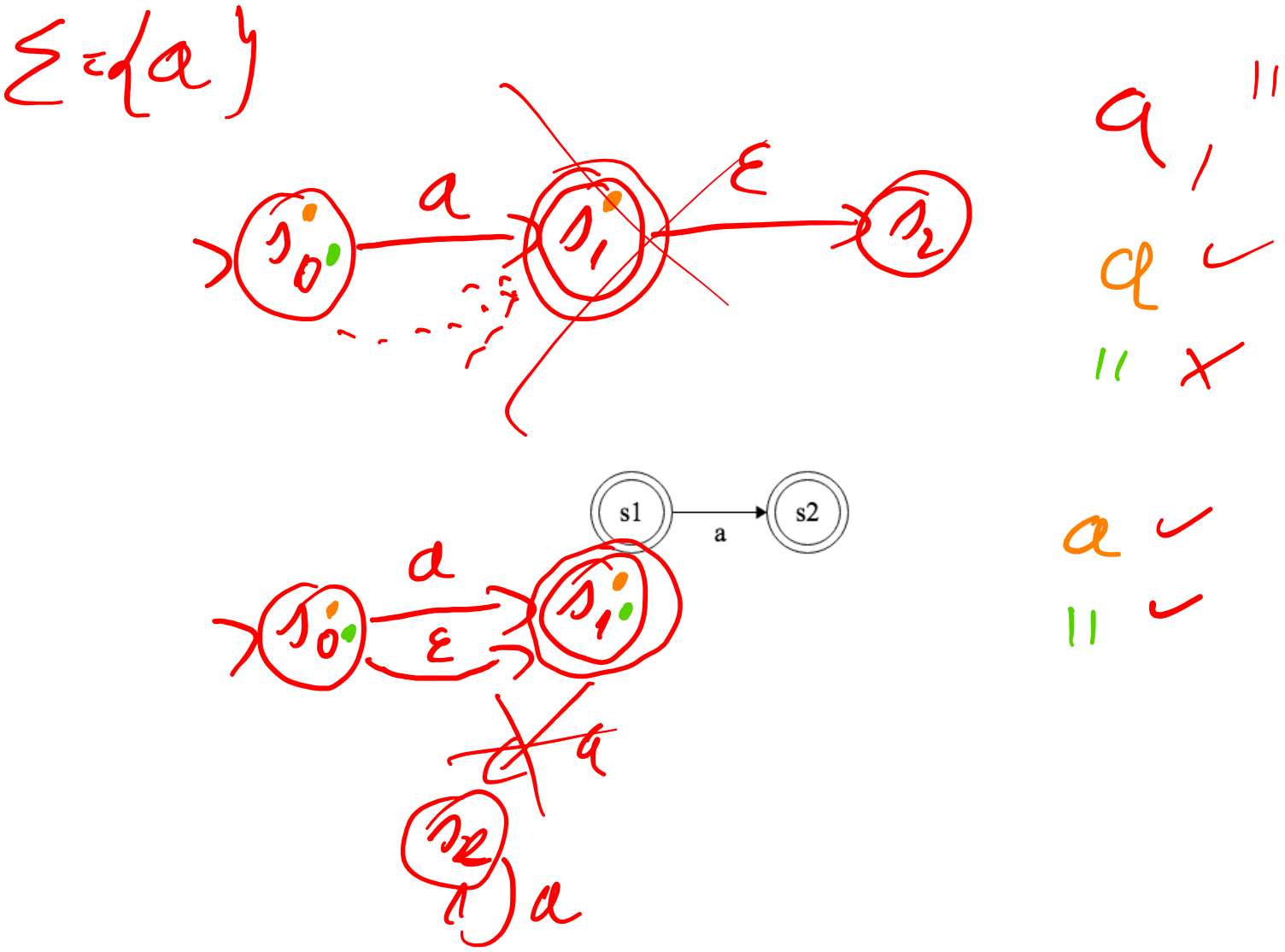
Automate fini pur a*



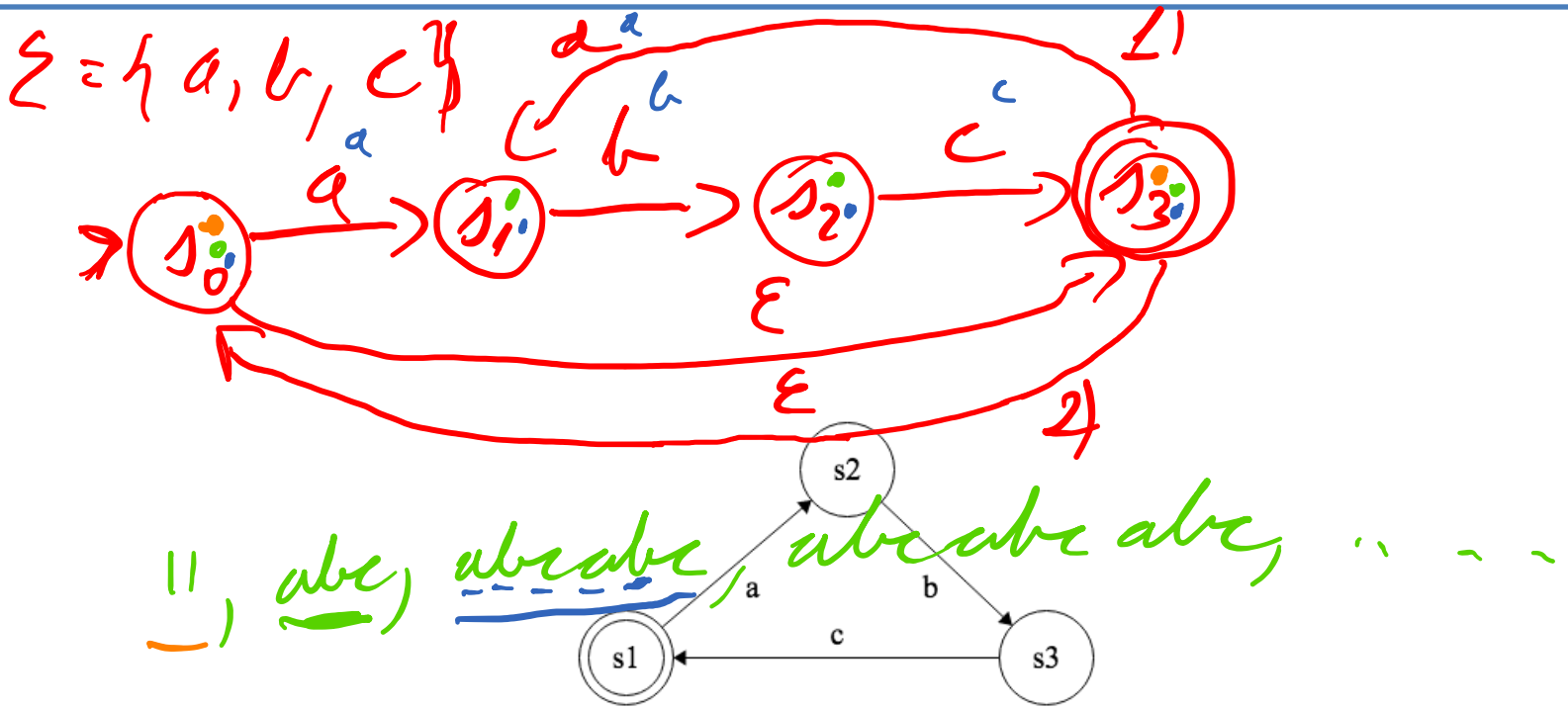
Automate fini pur a+



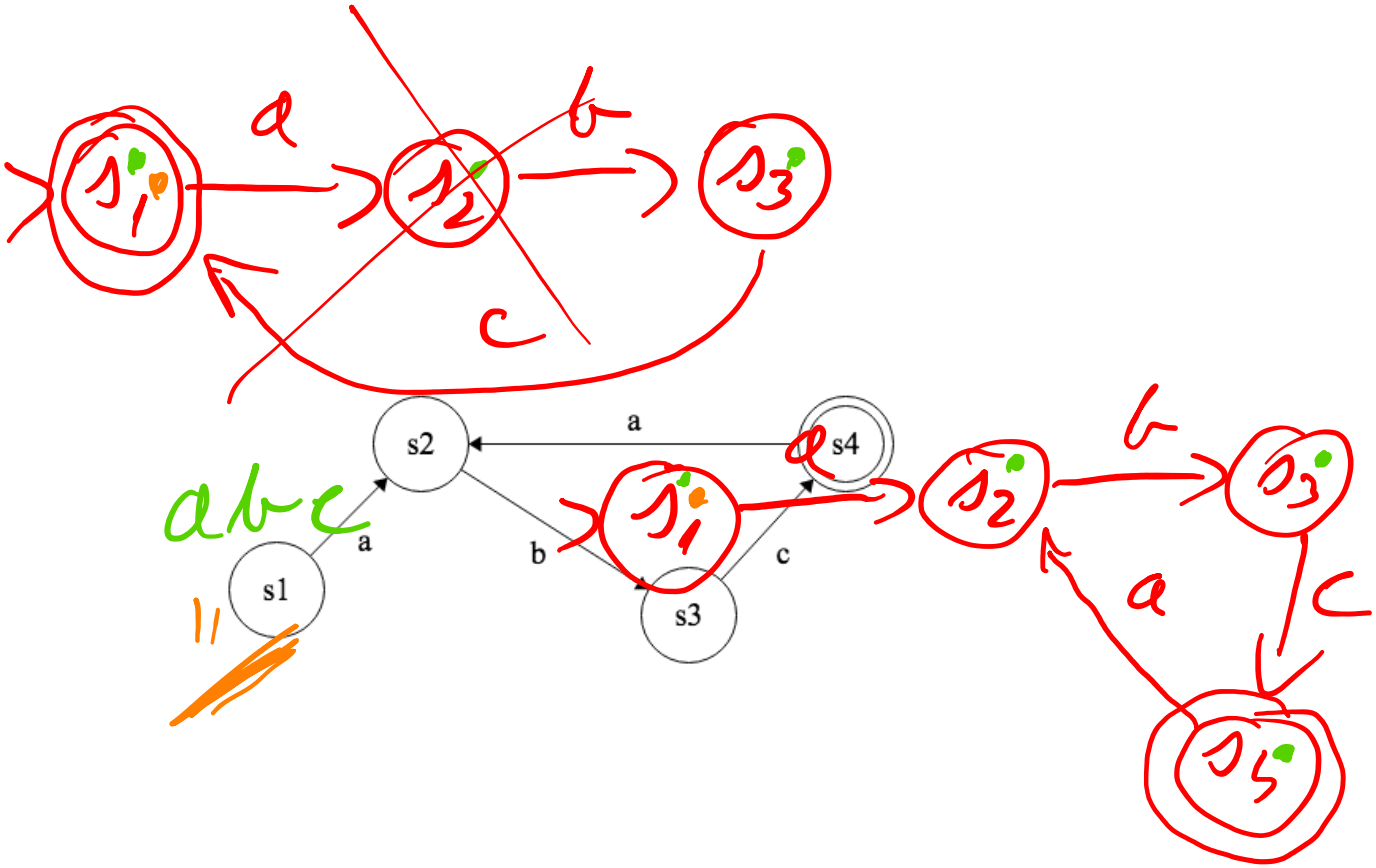
Automate fini pur a?



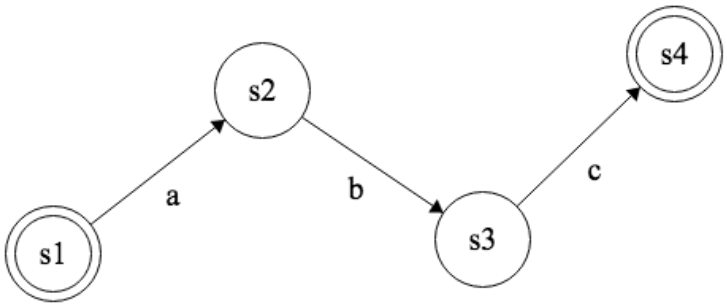
Automate fini pour (abc)*



Automate fini pour (abc)+



Automate fini pour (abc)?



Exemples

- Groupe
- Téléphone
- Email
- Nom de variable
- Numéro entier
- Numéro avec virgule

12201FA|B|E
12[0123FCABE]

FAGS/(1|21...|(12)[1234]

[012345]

nom@nom.nom.....

[ABC....Z]
[ABC...Z]?

[A-Z]

[a-z]

[A-Za-z]

[A-Za-z0-9_\.\-]+@
[A-Za-z0-9_\-]+\.
[A-Za-z0-9_\-]+)+

JavaScript RegEx

ALF
language crunch

let
```
var regex = new RegEx ('...');
```

let
```
var regex = /.../;
```

Documentation

https://developer.mozilla.org/en/docs/Web/JavaScript/Guide/Regular_Expressions

Fonctions

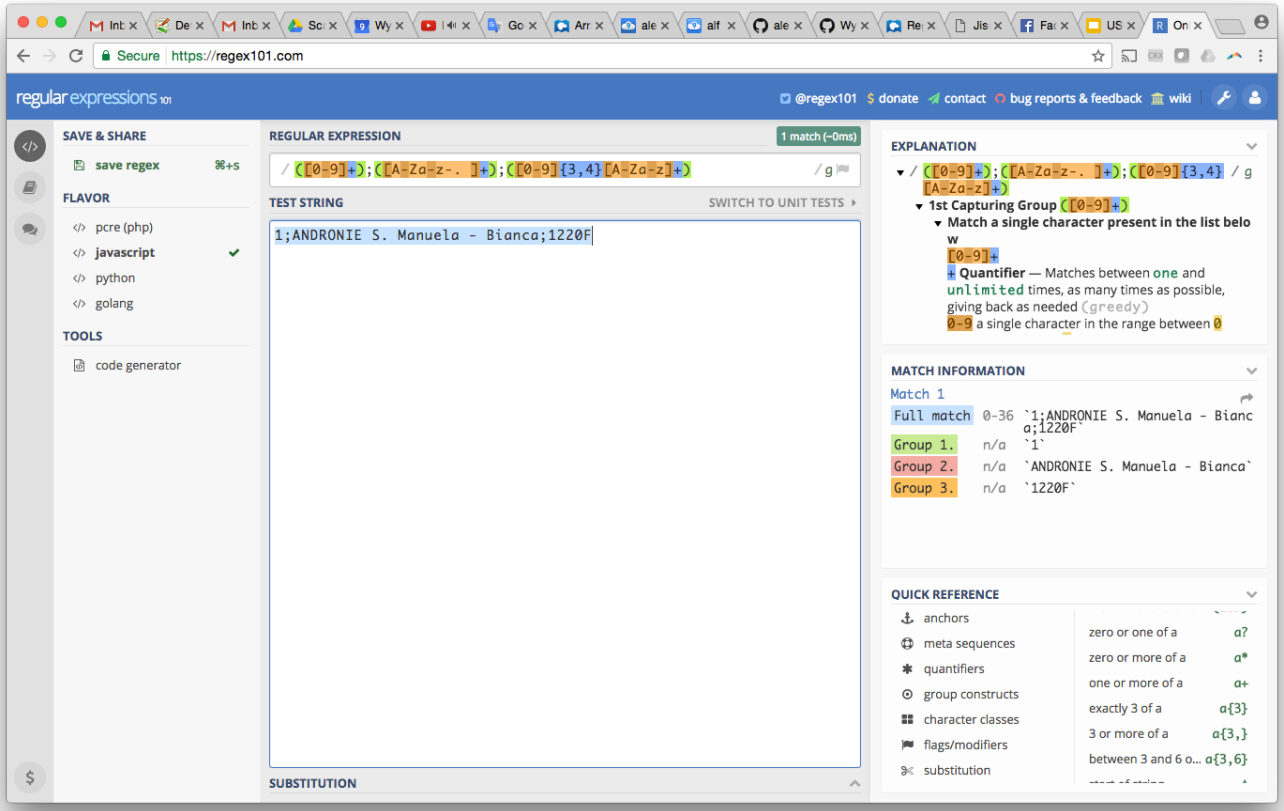
ALF
language crunch

- RegEx
  - regex.exec (string)
    - Array
  - regex.test (string)
    - Boolean
- String
  - str.match (regex)
    - Array
  - str.search (regex)
    - Index of match

s.split /\r?\n/

RegEx101

- https://regex101.com (vérificateur de regex)



Syntaxe d'expression régulière (JS)

ALF
language crunch

Character	Description	Exemple	String
{n}	n fois	a{3}	aaa
{n,m}	au moins n, au plus m	a{3,7}	aaa, aaaa
\w	alphanumérique et _	\w	a, ab
\t	TAB	^a\ta*	a »TAB »a
\n	fin de linge	a\nb	a b
\r	retour chariot	a\rb	a\rb
a(?!b)	a seulement si non suivi par b	a(?!b)	a, aa
a(?=b)	a seulement si suivi par b	a(?=b)	ab
( )	group	a(ab)a	aaba

JavaScript Example

ALF
language crunch

```
var regex = new RegEx ('([0-9]+);([A-Za-z-\. ]+);([0-9]{3,4}[A-Za-z]+)');

var regex = /([0-9]+);([A-Za-z-\. ]+);([0-9]{3,4}[A-Za-z]+)/;

var match = regex.exec ('1;ANDRONIE S. Manuela - Bianca;1220F');
```

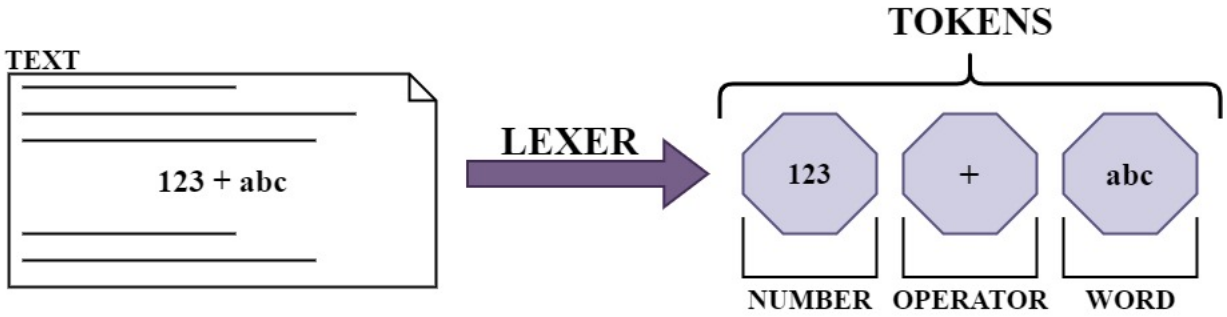
ALF
language crunch

```
([0-9]+);([A-Za-z-\. ]+);([0-9]{3,4}[A-Za-z]+)
```

match:

```
 [ '1;ANDRONIE S. Manuela - Bianca;1220F',
   '1',
   'ANDRONIE S. Manuela - Bianca',
   '1220F',
   index: 0,
   input: '1;ANDRONIE S. Manuela - Bianca;1220F fdsfs' ]
```

Lexer

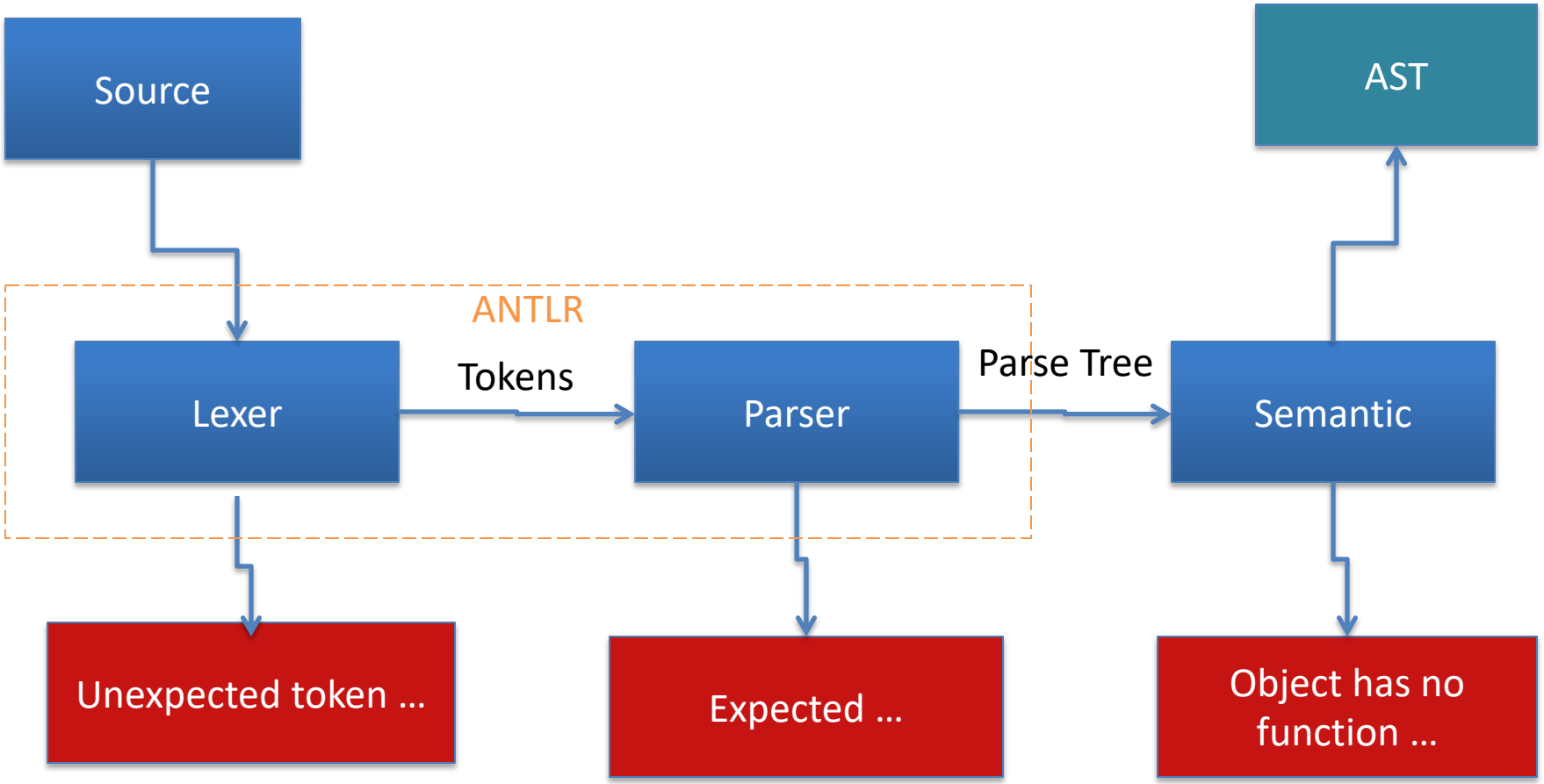


ANTLR

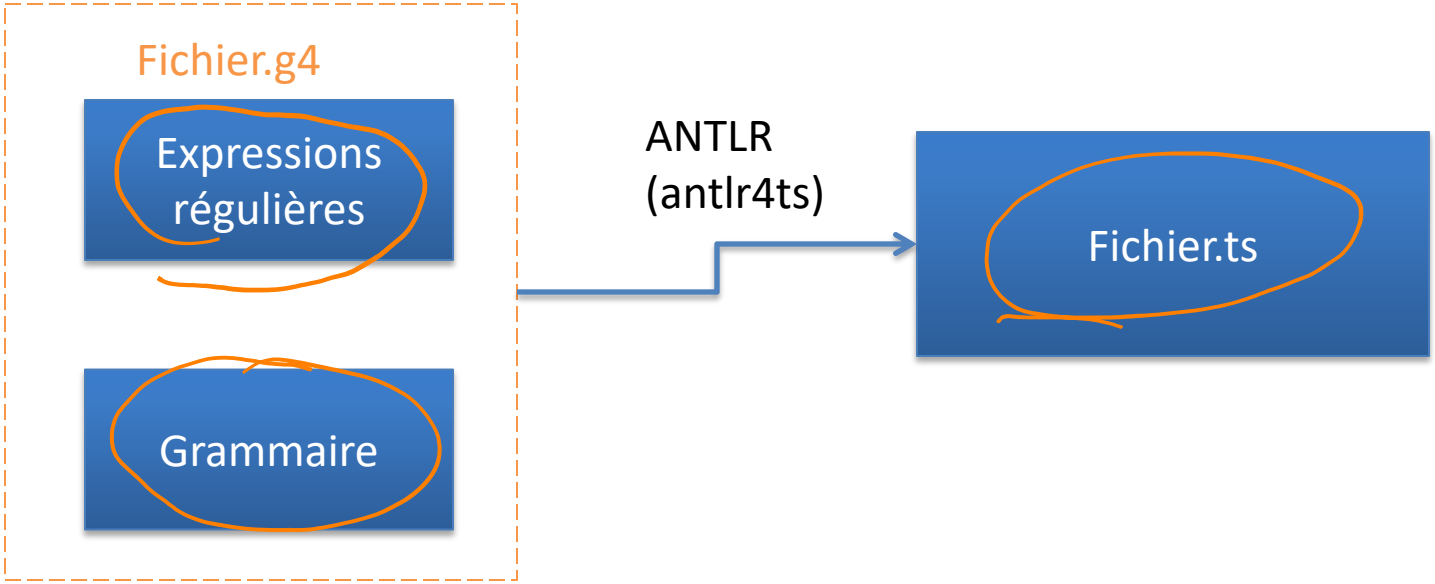
ALF
language crunch

Frontent

ALF
language crunch



Générateur de lexer/parseur



Fichiers G4

ALF
language crunch

- Lexer
  - Expressions réguliers
- Parser
  - Grammaire

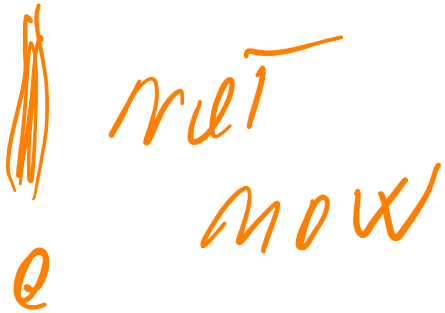
Fichier.l (lex ou flex)

ALF
language crunch

```
grammar Expr;

NEWLINE : [\r\n]+ ;
INT : [0-9]+ ;

prog: (expr NEWLINE)* ;
expr: expr ('*'|'/') expr
    | expr ('+'|'-') expr
    | INT
    | '(' expr ')'
    ;
```



Fichier.g4

```
grammar Expr;
```

Nom

```
NEWLINE : [\r\n]+ ;
INT : [0-9]+ ;
```

Expressions Régulières

```
prog: (expr NEWLINE)* ;
expr: expr ('*'|'/') expr
    | expr ('+'|'-') expr
    | INT
    | '(' expr ')'
    ;
```

Règles de Grammaire

Nom

Expr.g4

ALF
language crunch

```
grammar Expr;
```

Nom

Expression Régulières

Règles de Grammaire

Expressions Régulières

ALF
language crunch

Nom

```
NEWLINE : [\r\n]+ ;
INT : [0-9]+ ;
```

Expressions Régulières

Règles de Grammaire

Règles de Grammaire

ALF
language crunch

Nom

Expressions Régulières

```
prog: (expr NEWLINE)* ;
expr: expr ('*'|'/') expr
    | expr ('+'|'-') expr
    | INT
    | '(' expr ')'
    ;
```

Règles de Grammaire

Expr.g4 (vide)

ALF
language crunch

```
grammar Expr;

prog: ;
```

Sujets

ALF
language crunch

- Expressions régulières
  - Mathématique
  - JavaScript
- Lexer
- ANTLR

Questions

ALF
language crunch