

A

Learn more at https://tomassetti.me

About this Tutorial

Α

Α

Α



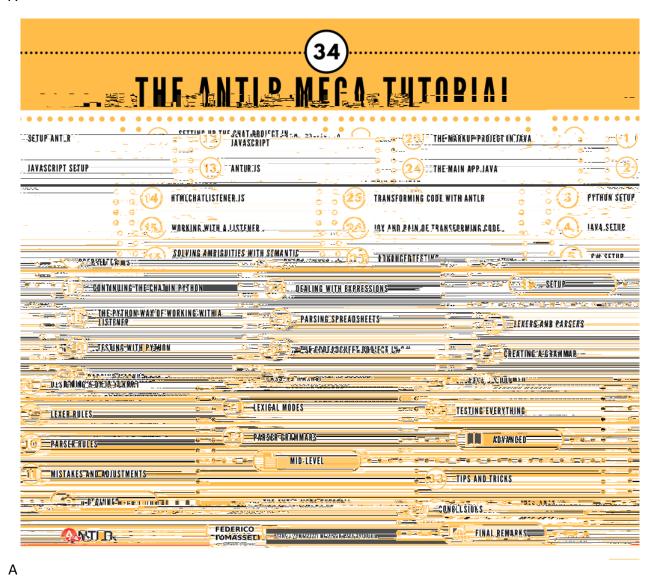
 $\mathsf{A} \quad \mathsf{A} \qquad \mathsf{A} \qquad \mathsf{A} \quad \mathsf{A} \quad$

Ahow to testA A A

Atons of examplesA

A A A fla A A

A Ac Ac A A A A A A A A B B A A A A

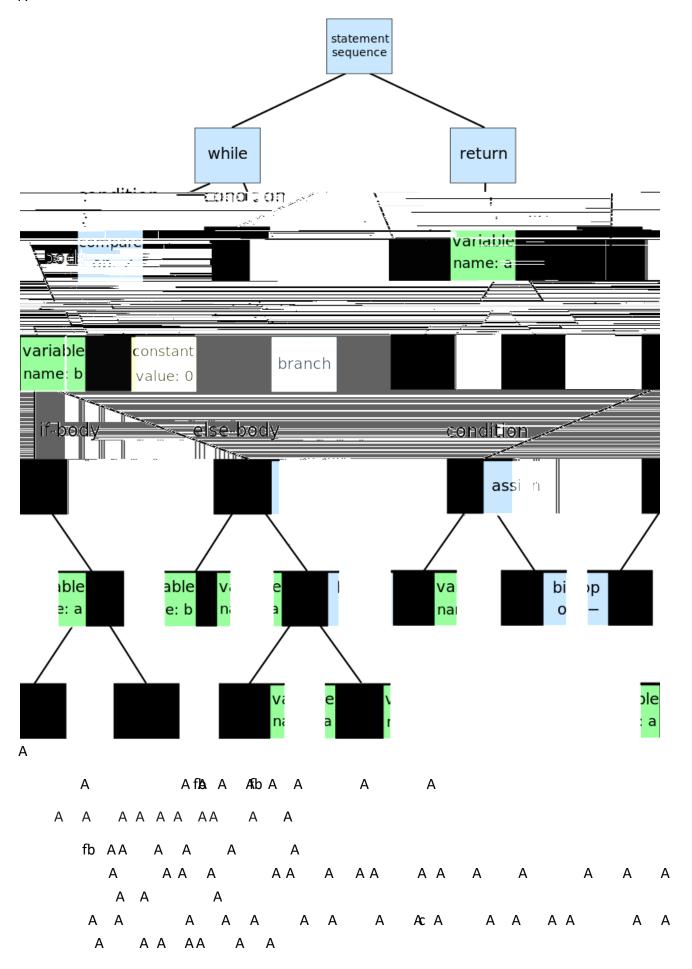


A AfA

What is ANTLR?

ÆΝ

Α



A A A A A ACAA A ACAA Afb AA AAA A A AAc A A A A Afb A A A Aparse tree A A AASTA A Ffb Α Α Α A A A A A A A A A A A A ACA Α AA AAACAC AAfAAA A A A A Α C AA A ACA AAA A A A A A C A fb Α A A A fb A A A Afb AAA A A BAA Α AA AA A $\mathsf{A} \quad \mathsf{A} \quad \mathsf{A} \quad \mathsf{A} \quad \mathsf{A} \quad \mathsf{A}$ Α A A A A A c A A A A A A A A A A Α Α

Aren't regular expressions enough?

A A A A flawhy can't luse a regular expression A A fA A A A A AA A fbA AA A A Afb AA c AAA A fbA Ac AA A A A Α A A A A A Ac A Afb A A A A Ac Α CAAACAAAAA CAfBAAA AA A AA AAA AA AAfb A A A Ac A A A AA A A A A A Α A A A A N A AA A A AA c A Afb A A A A

ANTLR vs writing your own parser by hand

A fa c A A A AA AfAA Α AA AfA AA AAAA AfbA AAAA A A A A A A A A A A A A A A A A A AAA A AAACA A A AA A AA AC AA A A AC A AAA AffaA AAAAAAAAA A A A A

Table of Contents

WHAT IS ANTLR?	3 A
AREN'T REGULAR EXPRESSIONS ENOUGH?	5 A
ANTLR VS WRITING YOUR OWN PARSER BY HAND	6 A
TABLE OF CONTENTS	7 A
SETUP	9 A
1. SETUP ANTLR	9 A
A	Α
2. JAVASCRIPT SETUP	11 A
3. PYTHON SETUP	12 A
4. JAVA SETUP	12 A
5. C# SETUP	15 A
ALTERNATIVES IF YOU ARE NOT USING VISUAL STUDIO CODE	15 A
PICKING THE RIGHT RUNTIME	15 A
BEGINNER	16 A
6. Lexers and Parsers	16 A
7. CREATING A GRAMMAR	18 A
A A	Α
C N A A	A
8. DESIGNING A DATA F	

A N A C A
C N A A
A
34. CONCLUSIONS
A

A A A A A A A <u>A</u>

Setup

1. Setup ANTLR

A A

Instructions

Executing the instructions on Linux/Mac OS

Α	Α	Α		A A		c			А	
Α	Α	fbA	Α	A fla	Α	АА	Α	Α	A A	
Α	А		Α					Α		

Executing the instructions on Windows

А	A A
Α	AA AG CAA AA fb A
Α	AA AA A CAA Af a A
Α	A A A A A A A A A A A A A
Α	A A A CAA AV CA
Α	A A AG c A
Α	AA A
Α	A A C AA
Α	A A AA
Α	A A C AA
Α	A A A
Α	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Α

Typical Workflow

A A Ac A A Agrammar A Afb A A A A A A A Α A A АА A Afb A Α Α A A A A A AΑ A A A A A A A A A A A Α



Α A AA A f A A A fbA Α A A G A A A fbA A A A AAA A A A Α A A AΑ A ffb Æfb Α Α AA A fb A A A Α A A A AΑ A A AAf**l**a A A A AΑ Α $\mathsf{A} \quad \mathsf{A} \quad \mathsf{A}$ A A Α ΑА C A fb A A A AAA A AA A A A A AΑ Α A A Afla A A A A A A A Α A A A Α A A Ac A A A fb Ac Α Α Α

fb A Α Α A A A A A A AA A A Α Α AAAAAAAA A AΑ Α fb Α Afb A A Af**l**a Α АА A A Ac A Ac A AA ffbo A ΑА Α Α Α AA fbA A A Afb A fbA fbA fbA A A c A A A A A A ΑА AfA A A 2. Javascript Setup AAAA Afb AAA Afb A Afb A A A A A A A A BA A A A A A A A BA A $\mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A}$ A A A ACA fb A A A A A A A Α Α Α A A A A A A A AfMa A AA A A AA AAf**la** AA A A A ACA ACA A A A A AC AGA AC A A A A c AA Afla A A A A A A A A A A A A ffb A AfbA A A A A A A A A C A A A АА Α Afb A A A A A A A A Afb A Α Α A A

3. Python Setup

```
Α
Α
  dependencies {A
      antlr "org. antlr: antlr4: 4. 9. 1"A
Α
      compile "org. antlr: antlr4-runtime: 4.9.1" A
Α
      testImplementation(platform('org.junit:junit-bom: 5.7.0'))A
Α
      testImplementation('org.junit.jupiter:junit-jupiter')A
Α
Α
  }A
Α
  Α
Α
  generateGrammarSource {A
      maxHeapSize = "128m"A
      arguments += ['-package',
  'me. tomassetti.examples.MarkupParser', '-visitor', '-no-listener']A
A }A
  compileJava.dependsOn generateGrammarSourceA
Α
Α
  sourceSets {A
Α
      generated {A
          java.srcDir 'generated-src/antlr/main/'A
Α
Α
      }A
Α
  compileJava. source sourceSets. generated. java, sourceSets. main. java
Α
  cl ean{
Α
      delete "generated-src"
Α
  }
Α
Α
 idea {
Α
Α
      module {
          sourceDirs += file("generated-src/antlr/main/")
```

Α	}
Α	}
Α	
Α	test {
Α	useJUni tPI atform()
Α	testLoggi ng {
Α	events "passed", "ski pped", "failed"
Α	}
Α	}

Α Α fb A A Α c A abstract syntax tree tokens expr PARSER LEXER PLUS NUM ↓ sum -437⁻ → 734 Α

A AA A AC A A A A A

AAAA A AA

/*A Α Α Ж AAA NCAA NCAA Α Ж Α Α Α Α Α Ж N C AA AAA A

A ACAAA AAAAAA A A AAAAA AACAA AAAfbA $\mathsf{A} \quad \mathsf{A} \quad \mathsf{A} \quad \mathsf{A}$ A f/A AAA AfbAA A A A A A A A A A A A A A A A A C A Alexer rules A A A A Aare analyzed in the order that they appear A A A A C A CA AA A fb AA A AA AcA A AfBA A Α Ac AAA AAfb AAfbc Ac A A A AclassA с А A A AclassA AfunctionA A fb Afb A A A Afb A A A A fb A fb A A A A

AC A A fla A A A A A Athere is a name, a colon, the definition of the rule and a terminating semicolon A

Af**énumber**A AA Affà AAAA cAA A AfBA A A A A C A A A fB<u>A</u>A A A A A A A A A A B A AWHITESPACEA A A Α A A A A A A A A A A AAfbA A Affa A A A A A A A A ACAfBA AAAAA A A Ac Α A A A A NC A A A A NC A A A A $\mathsf{A} \quad \mathsf{A} \quad \mathsf{A}$ $\mathsf{A} \quad \mathsf{A} \quad \mathsf{A}$ Α A fb AA A A A A A A A A A A Α AA 7. Creating a Grammar A A A A A A A A A A A A A A A A A Blb A A Ac A A fb A A Top-down approach A AA AfbAA A AfAAfbA AA A AAAA AfBAAT AAA AAA AAA GA AA AfbA AcA AA A Α A A Α • A fb A A ACAAAAAA AAfb AAAA AΑ AfbAAA ccAA A fb AcA A AA Ac A A A A fb A A A A A A A A A Ac A fb A A A A A A A A Afb A A A AAA AAA BAAAA ccA A A АА **Bottom-up approach** Ac A A A Afb A A A Afb A fb A A A Α A A Afb A A A A fb A A A

A AfbAA AfbAAc AAc A AA A AA AA Α AAfb AAA A AfBA A Ac A Afb A A A A A A Α A A A A A A A A A A A A AG A A A A A A A A AAAAAAA AAfbAA A A A A A C A A A A A fb A Α AANAAFb AAAAA AAAA AAA ABb A AAΑ A A A Afb A A A A A A A AfNAAC A AA fbAfb AACA AAAAAAAAAAA A Afb Afla A A A A A A A A A A

8. Designing a Data Format

A A AAc A A fla Ac A

9. Lexer Rules

Α	/ *A	
Α	A A A	
Α	А	
Α	Æ	
А	fb A AMM	AA
Α	fb A AWWWWA	AA

Α	fb A AWWAAA AA
Α	fb A AMMMAA AA
Α	fb A AMMMAA AA
Α	fb A AMMMAA AA
А	fb A AWWWAA AA
Α	AA .
Α	fb A AAA AA
Α	fb A AAA AA
Α	AA .
А	/**********A A A A A A
Α	AA .
Α	#######AAAAAAAA
Α	AA .
Α	<i>***********</i> A AA AA
Α	AA .
Α	ANNONNA AA A AA
Α	AA .
Α	####### A A A A AA AA
Α	AA .
Α	Y ///////AAA AA

Α	fb	Α	AA	AA					
А	N C	/XXXX A	А	Α	Α	AA			

Α

10. Parser Rules

Α	/ *A
Α	A A A
Α	A
Α	AA .
Α	##########A A G4A
Α	AA .
Α	///////AAAAAAAA
Α	AA .
Α	ANNONNA A AA AA AA AA
Α	AA .
Α	########A A A
Α	AA .
Α	ANNONNA A AA AA AA
Α	/*************************************
Α	#######A A A A
А	ANNOW A A A A A
А	AMMMMMMAA
Α	AA .
А	////////AAAYAAA
Α	AA .
Α	////////A A A A A A A A A A A A A A A A
Α	AA .
Α	AXXXXXXXX A A AA

AAA AAA AAA AfAA AAA AAA A A A A A Ac A A A A A A A Α A A AAfBAAA AAAA AAAAAA AAAA A A A A A A A ACA AA

А	A A Ac AA A A
А	AA A A A A
Α	AA A Afb A A A
Α	A A A
Α	A A A
А	A AA A A AA A AA AA AA AA AA AA AA AA A
Α	A A A A
А	A A A A A A A
А	A A A A A A
Α	A A
А	A A A A A A A A A A A A A A A A A A A

Α

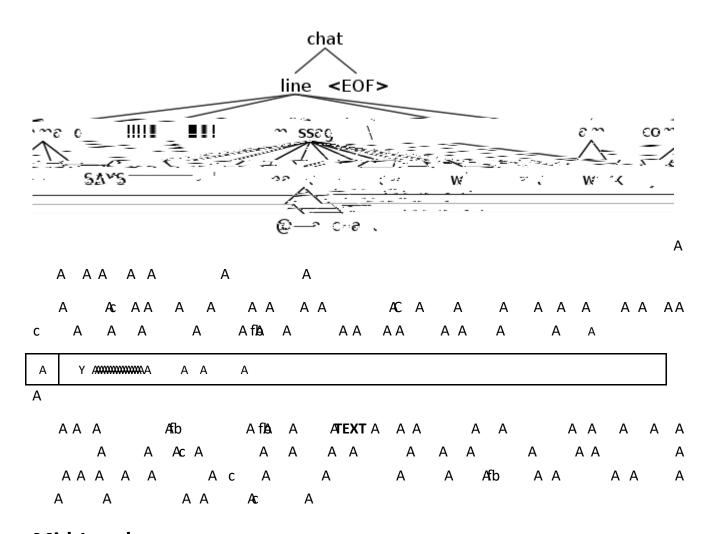
А	Α	Α	Α	Α	Α												
Α	Α	Α	Α	Α		Α	Α	Α Α	١	Α							
Α	Α			Α	Α	Α		Α	Α	Α	Α	Υ	Α				
Α	Α			G	G		Α										

Α

Α	A
Α	AA .
Α	/////////AA Y A Y AA
Α	AA .

	Α															
Ж																
	Y ///////	*******	A	Α	Α		Α									
4	Α		Α	Α	С		Α	АА	А	A	A A	A	Α		АА	A
	Α	Α	Α	Α		Α	Α	Α	Α	Ac		A A	Α	Α	Α	A f
	Α	А		Ac		AC	AAA	Afb A	Α	A A	Α	Α	A fb	Α	Α	
Α	Af	A A		Α												
	A	Y 2000000	A A A A A A	A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A	A A A A A A A A A A A A A A A A A A A

A A A N AfbAAA A **A** A N AAAA A AfbAA



Mid-Level

12. Setting Up the Chat Project with JavaScript

Α		Α	Α	
Α	А			
Α	Α			
А	Α	Α	Α	

Α	А								
Α	Æ								
Α	/***********A	A GAA							
Α	Æ								
Α	////// A	А	Α	Α	AA				
Α	AA								
А	AWWWWA	А	AA AA	A AA	АА	АА	AA		
Α	AA								
А	A XXXXXXXXX A	Α	,	A					
А	Æ								
А	/XXXXXXXX	A AA	А	Α	AA				
А	/*************************************	*********							
А	AXXXXXXXXX	AA AA							
Α	//////A A	АА							
А	///////A								
Α	Æ								
Α	AA A A	fb A	Α	АА	Α	А А	<i>А</i> А А	A A	AC GA

Α	fb A AMMMAA AA
Α	fb A AWWWAA AA
Α	fb A AWWWAA AA
Α	AA .
Α	fb A AAA AA
Α	fb A AAA AA
Α	A
Α	////////A A A A AA
Α	AA .
Α	AWWWWAAA A A A AAA
Α	A
Α	<i>,,,,,,,,,,,,,,</i> ,,,,,,,,,,,,,,,,,,,,,,
Α	AA.
Α	ANNONNA AA A AA
Α	AA.
Α	###### A A A A A AA
Α	AA.
Α	Y //////AAA A A A
Α	
A	
	A A A

Α

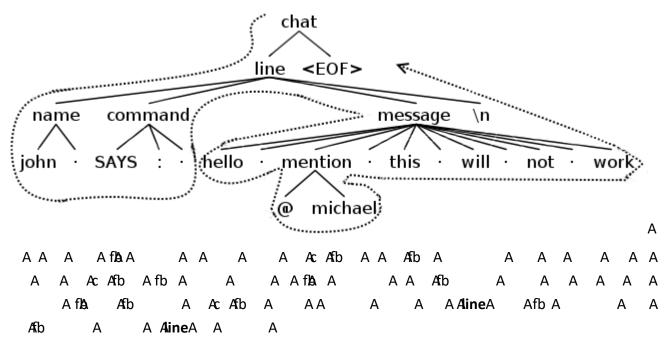
A A A A A fla A

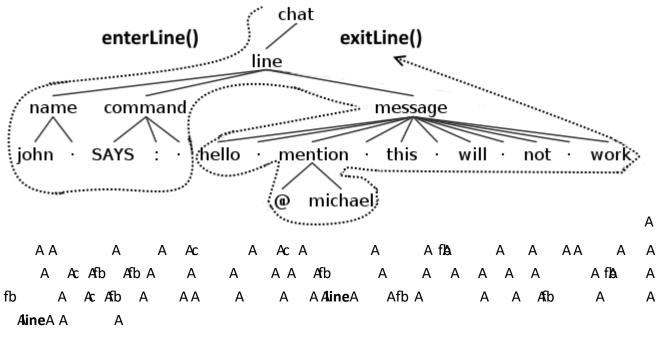
AATb AAAA AAAA AAAAAA Α fbA Ac

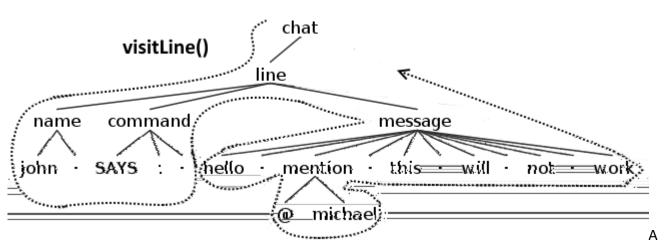
A AA A A A Afb A A Afb A A A A A Afla A A A A A A Afb A A fb A

Α	import antlr4 from 'antlr4';A						
А	A						
А	// This class defines a complete listener for a parse tree produced by ChatParser. A						
А	export default class ChatListener extends antlr4.tree.ParseTreeListener {A						
Α	A						
Α	// Enter a parse tree produced by ChatParser#chat. A						
Α	enterChat(ctx) {A						
Α	}A						
Α	A						
Α	// Exit a parse tree produced by ChatParser#chat. A						

A exit.





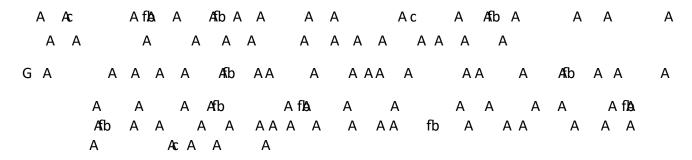


13. Antlr.js

AATO A AA A A A A A

Α	<pre>import { createServer } from 'http'; A</pre>
Α	import antlr4 from 'antlr4'; A
Α	<pre>const { CommonTokenStream, InputStream } = antIr4; A</pre>
А	<pre>import ChatLexer from './ChatLexer.js'; A</pre>
Α	<pre>import ChatParser from './ChatParser.js';A</pre>
А	<pre>import Html ChatListener from './Html ChatListener.js'; A</pre>
Α	A
Α	createServer((req, res) => {A
Α	A
Α	res.writeHead(200, {A
А	'Content-Type': 'text/html', A
Α	});A
Α	A
Α	res.write(' <html><head><meta charset="utf-8"/></head><body>');A</body></html>
Α	A
Α	var input = "john SHOUTS: hello @michael /pink/this will work/:-) \n"; A
Α	var chars = new InputStream(input, true) A
Α	var lexer = new ChatLexer(chars); A
Α	var tokens = new CommonTokenStream(lexer); A
Α	var parser = new ChatParser(tokens); A
Α	A
Α	parser.buildParseTrees = true; A
Α	var tree = parser.chat(); A
Α	var html Chat = new Html ChatListener(res); A
Α	antlr4.tree.ParseTreeWalker.DEFAULT.walk(htmlChat, tree); A

А	A									
А	res.write('');A									
А	res. end(); A									
Α	A									
Α	}).listen(1337); A									





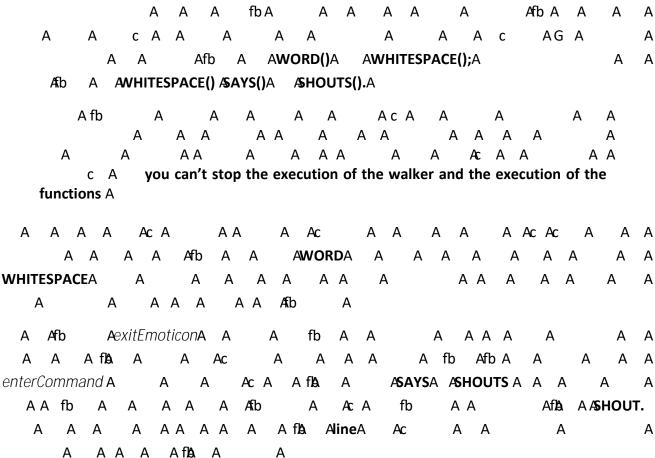
14. HtmlChatListener.js

A ACA AAA A A A BA AChatA A

Α	import antir4 from 'antir4'; A							
Α	<pre>import ChatLexer from './ChatLexer.js'; A</pre>							
А	<pre>import ChatParser from './ChatParser.js';A</pre>							
Α	import ChatListener from './ChatListener.js'; A							
А	A							

```
export default class Html ChatListener extends ChatListener {
       constructor(res) {A
Α
            super(); A
Α
            this. Res = res; A
Α
 Α
       }A
A A
       enterName(ctx) {
 Α
            this. Res. write("<strong>");
 Α
 Α
       }A
 Α
       Α
       exitName(ctx) {
 Α
                             Α
            thi s. Res. wri te(ctx. WORD().getText()); A
 Α
            this. Res. write("</strong> "); A
 Α
 Α
       }A
 Α
        exitEmoticon(ctx) {
 Α
 Α
            var emoticon = ctx.getText();
                                                   Α
 Α
            Α
            if(emoticon == ':-)' || emoticon == ':)')A
 Α
            {
 Α
                     Α
                ctx. text = "@"; A
 Α
            }A
 Α
 Α
            if(emoticon == ':-(' || emoticon == ':(')A
 Α
            {
 Α
                       Α
                ctx. text = "@"; A
 Α
 Α
            }A
 Α
       }
           Α
 Α
 Α
       enterCommand(ctx) {
                                      Α
```

```
if(ctx. SAYS() != null)A
           this. Res. write(ctx. SAYS().getText() + ':' + ''); A
Α
Α
     Α
     if(ctx.SHOUTS() != null)A
Α
                this. Res. write(ctx. SHOUTS().getText() + ':' + '<p
  style="text-transform: uppercase">'); A
Α
     }A
Α
     exitLine(ctx) {
                           Α
Α
       this. Res. write("");
Α
     }A
  }A
fb A A Afb A A A A AHtmlChatListenerA A AChatListener. A
     A f60A A A A A
         AAA AffaA fbA A AfbA A A
 A ctxA
                                                           Α
         A A fb A
                                        A fb A
                                                           Α
```



john SHOUTS:

0

15. Working with a Listener

А	enterColor(ctx) { A
Α	<pre>var color = ctx.WORD().getText();</pre>
Α	this. Res. write(' '); A
Α	}A
Α	A
Α	exitColor(ctx) { A
Α	this. Res. write(""); A
Α	}A
Α	A
Α	exitMessage(ctx) { A
Α	this. Res. write(ctx.getText()); A
Α	}A

Α

john SHOUTS:

ALLE .MT TANGE OFFICE OF COLVET WITH WITH STUDGEN OF TANKEN

Α		ΑA	١	Α	Α	Α	Α	c A	Α	Α	Α	Α	Α	Α	Α	A A	fames	sageA
Α		Α	Α		Æfb	Α	Α	Α	Α	Α	fb A		Α		A f lame	essage A	Α	АА
Α	Α																	
А		Α			A A	. A			Α	A c	Α	Α	Α	Α	Α	Αf	Afb	A A
Α		Α			A c	Α			Α									

Α	exitColor(ctx) { A							
Α	ctx.text += ctx.message().text; A							
Α	ctx. text += ''; A							
Α	}A							
Α	A							
Α	exitEmoticon(ctx) { A							
Α	<pre>var emoticon = ctx.getText();</pre> A							
Α	A							
Α	<pre>if(emoticon == ':-)' emoticon == ':)')A</pre>							
Α	{ A							
Α	ctx. text = "; A							
Α	}A							
Α	A							
Α	<pre>if(emoticon == ':-(' emoticon == ':(')A</pre>							
Α	{ A							
Α	ctx. text = "@"; A							
Α	}A							
Α	}A							
Α	A							
Α	exitMessage(ctx) {							

```
var text = ''; A
A A
       for (var index = 0; index < ctx.children.length; index++ ) {A</pre>
Α
           if(ctx.children[index].text != null)A
Α
               text += ctx.children[index].text; A
Α
Α
           el sea
               text += ctx.children[index].getText(); A
Α
Α
       }A
Α
  Α
      if(ctx.parentCtx instanceof ChatParser.LineContext == false)A
Α
           ctx. text = text;
Α
Α
      el sea
Α
       { A
           this. Res. write(text); A
Α
           this. Res. write(""); A
Α
      }A
  }A
```

Α

AA fbAAAAAA AcAA **Atext**AfbAAAA Α A A A A A A Af**là Aline**A Afl**à**AAA A AA AA A A Atext fb A A A A A A A A AcolorA C A A A A Α AAAfb A A f/A Α **Actx**A Α A A AC A A A A A A A A Afb A fla A A f/A Α Α Ac A A A A A A A C A Afla Afb textA AAAAA A CA AAAA Afb AAAA Afb A

16. Solving Ambiguities with Semantic Predicates

Cfb A A A A A A A C A ATEXTA A A A A A A ΑА Ac AAAAA A A A AA f/A A Asemantic predicates.A A A A Æfb A A A CAACAAfb A Α A AAc Α A c A A A A ACA AAA АА AfbA AA AAA AAAA ffb A A f&n A A АА A A A A fla A A A flactions A A A A C A C A AThe downside is that the grammar is no more language independent A AAA A ACA AfbAA A AGAA A AA ΑА A A A A A A A A A A A A A A A A BA A A A AΑ Ж Y ///////A fb A A A A fb АА Α AA A AlinkAAA Afb Acaa aa aa ATEXTA A A Ac Afb A A AAc A A A Α A A A Ac fb A A A Aflà A AAA AC AAA A AAA AAA AAA A **ATEXT**A А А A A A A A A A A A A A A A A Α Α AAAfbAA A fbAcA Ac A ffbb A A A A ffbb A AfbA AA AAA AA AA AAA fb A A Α A A A A A A A Table A A Α Α $\mathsf{A} \quad \mathsf{A} \qquad \mathsf{A} \quad \mathsf{A}$ A A A A A A Y AA AA A A Α Y AA AAAAA A A AA Α **A** A

A A A

AA

Y AA

AAAA

А	A A		
Α	А	A A A AA	
Α	A A		
Α	А	A A A AA	
Α	A A		
Α	A fb	A AA A AA	
Α	A A		
Α	А	A AA A AA	

17. Continuing the Chat in Python

Α	А	А	Α
Α			

pred A

Α	fA Y	fbA		А	Α	
А	AXXX FIZA	ААА				
Α	/XXXXXX	A fb	АА	A A fb	АА	Α

Α

18. The Python Way of Working with a Listener

Α

Α		Α	Α							
Α	fb	Α	Α	,	AΑ					
Α	fb	Α		Α	Α		Α			
Α	fb	Α		Α	А		Α			
Α	fb	Α			Α	Α			Α	
Α	Ж									
Α	flA			Α						
Α	AXXA	Α	ÆG			Α				
Α	AXXA	Α	A			Α				
Α	Α		АА					Α		
Α	/	,	AΑ			Α				
Α	/	ΑA	4		Α					
Α	Æ									
А	AWA		ΑA				Α			
А	AXXXA									
Α	AXXX		A A	١				Α		
Α	/		AΑ			Α				

Α	fb A A AA
Α	fb A A A A
Α	fb A A A A
Α	AA .
Α	A AA
Α	ANNA FIDA A
Α	ANNAMA fb AA A
Α	ANNOWA fb A G c A
Α	A
Α	ANNA FIDA AA
Α	ANNOWA fb A
Α	AA .
Α	ANNA FILA FIDA AA
Α	ANNOWA fb A
Α	ANNOWA fb A A
Α	AA .
Α	ANNA FILA FIDA AA
Α	ANNONA AA A
А	######################################
Α	AA .
Α	ANNA FIBA FIDA ANNANA
Α	ANNANA A A A
Α	//////A A A A
Α	AA .
Α	ANNA FIDA AA
Α	ANNONA AA A
Α	AA .
Α	ANNOMATIA A A A A A A A
Α	AMMMMM A A © A
Α	AMA
Α	AMMMAfia A A A A A A A

А	AWWWWWA	AA 😀 A							
А	AA								
А	AXXA f/A	fbA		А					
А	AXXXXXA	AA A fb	АА	Υ	Α	Υ	Α		

19. Testing with Python

A A A A AA A

A

Α		Α	Α						
Α	fb	Α	Α	АА					
Α	fb	Α		Α	Α	А			
Α	fb	Α		Α	Α	А			

Α

Α	fb A A AA
Α	fb A A A A
Α	fb A A A A
Α	fb A A A A
Α	fb A A A A
Α	A A
Α	A A
Α	AA .
Α	A A
Α	AA .
А	ANNA FILA FILIA ANNOMINA
Α	///////A A A /////////A
А	ANNONA AA A
А	ANNONA AA A
А	AWWWA
Α	ANNONA fb AA A
А	ANNOWA fb AA A
Α	AA .
Α	ANNANA ANNANA
Α	ANNONA A A fb A
Α	AMAXXXXX
Α	AA .
Α	/////// fb A A ////////////
Α	/XXXXXXX
Α	/####A A A
Α	ANNOUNA
Α	ANNA fIA fbA
Α	ANNOWA AA fb AA

А	/////// A A ////////// A
Α	AMA
Α	ANNOWA AA fb A
А	ANNOWA AA A
А	ANNOUNA A ANNOUNNA
Α	AA .
Α	ANNONNA A A A A A C A A A A A C A A A A A C A A A A A C A A A A C A A A A C A A A A C A A A A C A A A A C A A A A C A A A A C A A A A C A A C A C A A C A C A A C A C A A C A C A A C A C A A C A C A A C
Α	ANNOWA fb fb c A A
Α	AA .
Α	ANA fla fbA
Α	ANNOWA AA fb A
Α	//////A A A //////////A A
Α	AMMA
Α	ANNONA A A fb A
Α	/#/// AA A
Α	/#### A /#########
Α	AA.
Α	ANNOWA A A A C A A A
Α	ANNOWA fb fb c A A
Α	AA .
Α	fla AAAA
Α	/##A A

20. Parsing Markup

c A AC AA A N AN AA AA fb Afb AA Afb AA AC AA AA A A A A A A Afb Α AAfb AfA A A A C A Ac A A A A A A Α Aisland languages A A A A A A A A A A AAAAΑ f**l**A A A A ACA A A AA AA AA AA Afb A A A A A A Ac A A A Alexical modes AA fb AfbAAAAAA AAABAA

21. Lexical Modes

A A A A A A A A A A A A A A A

Α	A AN A
Α	AA .
Α	AWWWWWAA A A N CC AA
Α	Y ////////AA AA
Α	AA .
Α	A A A A
Α	ACC A
Α	AA .
Α	ANNONNA AAAN AA
Α	/////////A A AA
Α	R
Α	////////A A A AA
А	ЛТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТТ
А	AMMMMMMAA A A AA
Α	AA .

AAAfbAAA AAAfb AAAfb AA Α Α Af AAA C AYou simply cannot define a lexical mode together with a parser grammar A A A A A A A A AΑ A A A AΑ AAAAAAA fb AA A f<u>bcc</u> A A A A A A Ac A Αc АА Ac A A Affb A A A A A A A A A fb A AA A Afb Af**la** A A Afb A A A Α Alexical modes Ac A A Α Afb A A Α Α АА Α A A A A A Α ΑА A A A Α Α Afb A Α Α Ас Α ΑА c A A A A fb A A A A A Α A A A A ΑА $\mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A}$ A A A A A A A A A A A A A A A C A fb A Α A A A A A A R A A fb Α Α Α Α A Ac A Α AAA

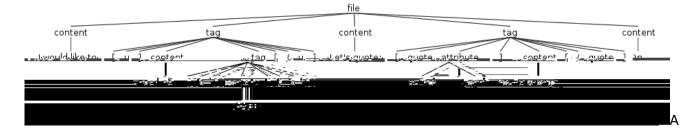
A A AA An the lexer grammar we need to define all tokens, because they cannot be defined later in the parser grammar.A

22. Parser Grammars

A A A A A A A A A A A A

_	
Α	A AN A
Α	AA .
Α	AA W c N AA
Α	A
Α	fb AWWAA AA
Α	A
Α	c AAAA A AA
Α	AA .
Α	AAA Y AA
Α	AA .
Α	AA AA AA
Α	A
А	ANNA AAA C AA AAAAA

A AA AA AA AA



Advanced

23. The Markup Project in Java

Α	# use gradle to build the projectA
Α	./gradlew compileJava A
Α	# if you are not using an IDEA
Α	# and you have defined the fatJar task as in the repositoryA
Α	./gradlew fatJar
Α	java -jar .\build\libs\markup-example-gradle-all.jar

24. The Main App.java

A A A A A A A A A A A

Α	А	N	А			
Α	А	A				
Α	А	А				
Α	Æ					
Α	c A A A					
Α	А					
Α	AMA c A A A A	A AA				
Α	AMAA					
Α	A000000A A	AA ft) А			
А	AMMA A A A A C	c A	A A	Ac c	AAA	
А	AMMAMA A Afb A A	A A		А	А	
Α	A0000000N A	AA AN		Α		

Α	/ //////			Α			AA A				Α	
А	AXXXXXX		Α		АА	AN				Α		
А	AA											
Α	/ /////// N		G		Ælb	АА	f	ъ	//////////			
Α	/ ////////	W	Α	АА	AN	W	/XXXXXXXXXXXX					
Α	/ //////	f	ъ	AX	XXXXX							
А	AXX A											·
Α	А											

25. Transforming Code with ANTLR

Α	А	N	А		
Α	Æ				
Α	А	Α			
Α	А	А			
Α	А	А			
Α	Æ				
Α	c A AN	W A AN	C W	Α	

Α	A
Α	AXXA A
Α	AWACAAGN G A A
Α	AMAA
Α	AWWA A
Α	AWWA A
Α	AMMA A
Α	AMMA A
Α	AWWA A A
Α	AWAA
Α	AMMA
Α	AWA A
Α	AMA C A A N A A
Α	AMA A
Α	<i>A</i> XXXXXX Y A
Α	AMMMMA
Α	AXXXXXX A A
Α	AWA A
Α	A
Α	Afb A A A A A A A A A A
	AAAAA AAfb A Aca A AAA AAfb
	A A A A A A A A A A A A A A A A A A A
	A A A A A A A A A A A A A A A A A A A
	A AffaA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
26	. Joy and Pain of Transforming Code
	fb A A A A A A A A A A A A A A A A A A A
Α	A
Α	c A A N A ANNA

Α	AMA A AMMMMA
Α	AAMA
Α	AA
Α	A
Α	c A A N A A
Α	A
Α	AMAfb A fileN G A
Α	AMA A
Α	//////Afb A A ///////////A
Α	AMMANAAA AMMANAAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAA AMMANAAAA AMMANAAAAA AMMANAAAAA AMMANAAAAAAAA
Α	ANNOMATO A A A
Α	.ж
Α	AMA AMA
Α	AA.
Α	AMA A A
Α	А
A	
C TR	DA AAAA AAAA AAAAAAAAAAAAAAAAAAAAAAAAA
,	ACA AA
	A A A A A A A A BA A A A A A A A A A A
	A Atag A Ac A A A A A A A A
	A A Ac A A Aflà A A A AtagA AcontentA A
Α	A

A			А							
Α	С	Α	Α	N		Α	AXXX			
Α	А									
Α	/XXX \	Α	AA A	1						
Α	/XXX	Α		AA A	AA A	١				
Α	Æ									

А	ANNOUNA Ac A
Α	AAA AA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Α	<i>A</i> ₩₩₩€ A
А	/#### A A
А	ANNONNA AA AA AA AANNONNA
А	ANNONAC A
А	/#### A A
А	₩₩₩₩ A c AA c A
Α	ANNONNA C AA C C C A
Α	ANNOUNCE AA AAAA
Α	ANNOUND AA AAA AAA AAAA
Α	ЛИЛИПИИ А С АА А
Α	////////C A
Α	AMA .
Α	A
Α	AMA A A A
А	AA
Α	AMAGD AN A A A
Α	AMA_AMMAMAMAMA
Α	AWWAfb A A A
А	ANNONNA A A A A
А	AMMMAfb A A A
А	AMMMMMA A A A AMMMMMMMA
Α	AMA_AMMAMA
Α	AMMA
А	AAA A A
А	AMMA
А	AMA A AMMANA
А	A
Α	

Α		// private variables inside the class AppTestA
Α		private MarkupErrorListener errorListener; A
Α		private MarkupLexer markupLexer; A
Α	A	
Α		@Test A
Α		<pre>public void testText()A</pre>
Α		{A
Α		MarkupParser parser = setup("anything in here"); A
Α	A	
Α		MarkupParser.ContentContext context = parser.content(); A
Α		A
Α		assertEquals("", this. errorListener.getSymbol()); A
Α		}A
Α	A	
Α		@Test A
Α		<pre>public void testInvalidText()A</pre>
Α		{A
Α		MarkupParser parser = setup("[anything in here"); A
Α	A	
Α		MarkupParser.ContentContext context = parser.content(); A
Α		A
Α		// note that this.errorListener.symbol could be emptyA
Α		// when ANTLR doesn't recognize the token or there is no error. \mathbf{A}
Α		// In such cases check the output of errorListener A

Α	assertEquals("[", this. errorListener.getSymbol()); A
Α	}A
А	A
Α	@Test A
А	<pre>public void testWrongMode()A</pre>

Α

Α	A
А	assertEquals("", this. errorListener.getSymbol()); A
Α	}A
А	A
А	@Test A
Α	<pre>public void testInvalidAttribute()A</pre>
Α	{A
Α	<pre>MarkupParser parser = setup("author=/\"john\""); A</pre>
Α	// we have to manually push the correct mode A
Α	this.markupLexer.pushMode(MarkupLexer.BBCODE); A
Α	A
А	MarkupParser.AttributeContext context = parser.attribute(); A
Α	A
Α	assertEquals("/", this.errorListener.getSymbol()); A
А	}A

AAA A AC AA A fAA A A A A A A AA ACA A A A ACA A A A А А AAA A AAA A AAAG A A A A A A ATEXTA Ac Α А А Α Af&A attributeA c A A A A A A A A A A A A A A A A A fBAAAAFbACAA A AAA fbAA AAA Α A A

28. Dealing with Expressions

AAAAffb A AAGA AA AA ΑА A A A A Ac A A A A A A f&a A AAAA C AA Affb A A A A A AGA AA AC A A A A A A C A Afb A AAA AAAAAAfbA A A A A A A A A $\mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A} \mathsf{A}$ AACA ACA Α Α A A A A A A A A A C A A $\mathsf{A} \quad \mathsf{A} \quad \mathsf{A} \quad \mathsf{A}$ Afb A

Α **/₩** Α **/W** Α /XXXXXX\ /XXXXXXX A AWA

A A AC AA A AA

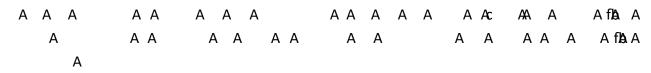
AfBAAA Aleft-recursive rules. A A A A Α AA A Afb Ac AA AA A c A A A A Α A A A A A A A C A A A A A Α AfMa fbA A A A

А	AA A A
Α	ANNA A A AA
Α	AA A AA
Α	ANNONA A N.C. AA

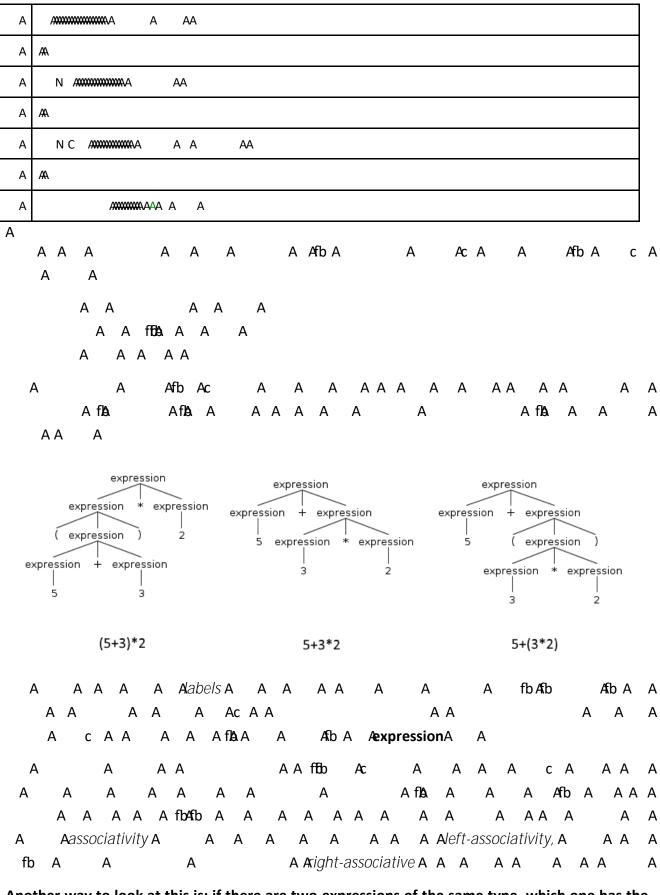
A Afb A fb A A A A C A A A Ac A A A AA Afb A Α ANTLR4 can create a similar structure automatically, so we can use a much more natural syntax A

Α		AA	АА	А	
Α	/XXXXXX	Α	АА	AXXXXXXXXXXXXXA	
Α	/XXXXXX	A NC	Α		
Α	/XXXXXX	А			

29. Parsing Spreadsheets



1	
Α	A A
Α	A
Α	######A A A ##########################
Α	/************A A A A A A
Α	/******A A N A
Α	/******************A
Α	///////A N A A A A A A A A A A A A A A A
Α	/***********A N C /********************** A
Α	///////A ///////////A A //////////////
А	/***********A
Α	AA .
А	fb A AAA AA
А	fb A AMMAA AA
А	A
Α	///////A AA
А	/////////A A AA
Α	///////A AA
Α	N AMMANA A AA
А	AA .



Another way to look at this is: if there are two expressions of the same type, which one has the precedence: the left one or the right one? A A A A A A A A A

Α	AXXXXXXXX		,	4			АА	Α	А
А	<i>/</i>		Α			АА	Α		А
А	AA								
Α	/XXXXXXXX					Α		АА	А
Α	<i>A</i>	W	Α	АА	Α		W	Α	
Α	AXXXXXXXXX								
Α	AXXXXXXXX			W				Α	
Α	AXXXXX A								
Α	AWA A								
Α	А								

31. Excel is Doomed

A A A A A A A A A A A SpreadsheetA A

Α	c A A W AA C W c A
Α	A
Α	ANNA A A A A A A
Α	AA .
Α	ANNA c A A c AW A A
Α	ANA_AWWWWA
Α	ANNONA A c NC A c fb A
Α	ANAA
Α	AA .
Α	ANA C A A C AV A A
Α	ANA.A
Α	ANNONA AAA A
Α	AA .

Α	ANNANA A A
Α	AMA A
Α	AA .
А	AMA c A A c AV
Α	AMA A
Α	ANNOWA AV A
Α	AMA A
Α	AA .
Α	AMACA ACAVIN NA AA
Α	AMA .
Α	AMMMA C A fbA AV A
Α	ANNOWA C A AAV A
Α	AMMMA C A AAA
Α	AA .
Α	AMMMAFIA A A A
Α	AMMAMA A A fbA A A
Α	AMMMAFIA A A A
Α	AMMAMA A A fbAA A
Α	AA .
Α	AXXXXXX A A
Α	AMA A
Α	A
Α	ANA A
Α	AA .
Α	ANA C A A C AV G G A A
Α	AMA A
А	AXXXXX A AA N A
Α	ANNOWA C A AAA
Α	AA .
Α	AMMANA A
Α	//////A

А	<i>A</i> ₩₩₩₩ A A
Α	AMMANAMAN A AN W A
Α	. ////////////////////////////////////
А	A
Α	/////// A A
А	//////// A /N W A
Α	. жжжжжжж
Α	ANNOVA A
Α	AA .

Α	<pre>public void testExpressionPow()A</pre>
Α	{A
Α	setup("5^3^2"); A
Α	A
Α	PowerExpContext context = parser.expression() as PowerExpContext; A
Α	A
Α	CommonTokenStream ts = (CommonTokenStream)parser.InputStream; A
Α	A
Α	Assert. Equal (SpreadsheetLexer. NUMBER, ts. Get(0). Type); A
Α	Assert. Equal (SpreadsheetLexer. T2, ts. Get(1). Type); A
Α	Assert. Equal (SpreadsheetLexer. NUMBER, ts. Get(2). Type); A
Α	Assert.Equal (SpreadsheetLexer.T2, ts.Get(3).Type); A
Α	Assert. Equal (SpreadsheetLexer. NUMBER, ts. Get(4). Type); A
Α	}A
Α	A
Α	[Fact]A
Α	<pre>public void testVisitPowerExp()A</pre>
Α	{A
Α	

```
A
  [Fact]A
  public void testWrongVisitFunctionExp()A
  { A
Α
Α
       setup("logga(100)"); A
Α
  A
             FunctionExpContext context =
                                                      parser. expressi on()
Α
                                                                               as
  FunctionExpContext; A
Α
Α
       SpreadsheetVi si tor vi si tor = new SpreadsheetVi si tor(); A
Α
       double result = visitor. VisitFunctionExp(context); A
Α
  A
      CommonTokenStream ts = (CommonTokenStream) parser. InputStream; A
Α
Α
  A
Α
      Assert. Equal (SpreadsheetLexer. NAME, ts. Get(0). Type); A
       Assert. Equal (null, errorListener. Symbol); A
Α
       Assert. Equal (0, result); A
Α
  }A
Α
  A
  [Fact]A
  public void testCompleteExp()A
Α
  { A
Α
       setup("log(5+6*7/8)"); A
Α
Α
  A
Α
       Expressi onContext context = parser.expressi on(); A
Α
  A
       SpreadsheetVi si tor vi si tor = new SpreadsheetVi si tor(); A
Α
       double result = visitor. Visit(context); A
Α
  A
```

A Assert. Equal ("1.0107238653917732", result. ToString(System. Globalization. CultureInfo. GetCultureInfo("en-US"). NumberFormat));

Catchall Rule

A AAAA

Α

Channels

Α	AA R N	I AA		
Α	A A A A A	A A		
Α	NN AA A	АА	R	N AA

Rule Element Labels

A AA fb A A AA

Problematic Tokens

Α	Ac fbA	AA AAA AfBaA fb AcA AcA
Α	AAA AA	
А	А	A AA fb AA AA A f b A
Α		A AA

A A A fBa fb A Ac fbA A A A A A A C A Ac Α CAA AAA fb AA A A fb A Ac fbA Afb A AAAΑ AA A AA AAf**l**a ААс Α A c AfA c A A A A A Α Α Α Α Α A AAA Afb AAc fbA A A Α Α A A Ac Α AAA AAAA A MAA AACA Afb A Α Α

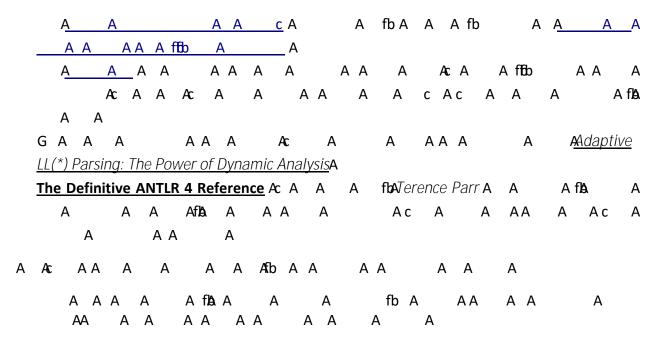
Α	Afb A A
Α	G AA A
Α	AA A GANCA
Α	A A A
Α	AA A GA GA NC A

34. Conclusions

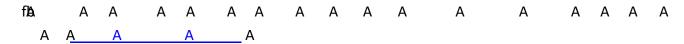
Α

- A A AA A
- A AA A AA A
- AA A A A A
- AAA AA AA AA A
- Afb A A fb C A A A A A A A A A
- A A A
- A A A A

AA Affa A A A fb Ac A A



Why program by hand in five days what you can spend twenty-five years of your A life automating?



We would like to thank Bernard Kaiflin for having revised the document and helped us improving it.

We would like to thank: Brasilio Castilho, Andy Nicholas for having spotted errors and typos in the article.

We worked quite hard to build the largest tutorial on ANTLR: the mega-tutorial! A post over 14.000 words long, or more than 70 pages, to try answering all your questions about ANTLR. Missing something? Contact us and let us now, we are here to help.