

Riconoscitore di grammatiche LR(1)

Documentazione

Luca Filice, Matteo Gusmini, Davide Presciani



Università degli Studi di Bergamo
Dipartimento di Ingegneria Gestionale, dell'Informazione e della trasmissione

Indice

1. Prefazione	3
1.1 Version History	3
1.2 Sommario modifiche	3
2. Introduzione	4
2.1 Definizioni e glossario	4
2.2 Scopo	4
2.3 Toolchain	5
3. Iterazione 0	6
3.1 Requisiti funzionali	6
3.1.1 Struttura corretta della grammatica	6
3.1.2 Errori lessicali	6
3.1.3 Errori sintattici	7
3.1.4 Errori semantici	7
3.2 Casi d'uso	8
3.2.1 UC1 - Analisi di una grammatica	8
3.3 Struttura del programma	9
3.3.1 Scelta architetturale	9
3.4 Programma di testing	10
4. Iterazione 1	11
4.1 Design Architetturale	11
4.1.1 Class diagram	11
4.2 Testing	13
4.2.1 Descrizione del prodotto software	13
4.2.2 Funzionalità oggetto di test	14
4.2.2.1 Elenco delle funzionalità oggetto di test	14
4.3 Esiti dei test	16
4.3.1 Esiti e copertura dei test di unità	16
4.3.1.1 Esiti dei test per la classe Stato	16
4.3.1.2 Esiti dei test per la classe RegolaDiProduzione	17
4.3.1.3 Esiti dei test per la classe NonTerminale	17
4.3.1.4 Esiti dei test per la classe Solver	18
4.3.1.5 Esiti dei test per la classe Terminale	18
4.3.2 Esiti e copertura dei test di sistema	19
5. Iterazione 2	20
5.1 Design Architetturale	20
5.1.1 Librerie utilizzate	21
5.1.2 Class Diagram	21
5.2 Casi d'uso	23
5.2.1 UC1 - Analisi di una grammatica	23
5.2.2 UC2 - Visualizzazione della grammatica analizzata	24
5.3 Testing	25
5.3.1 Descrizione del prodotto software	25

5.3.2	Funzionalità oggetto di test	26
5.3.2.1	Elenco delle funzionalità oggetto di test	26
5.4	Esiti dei test	28
5.4.1	Esiti e copertura dei test di unità	28
5.4.1.1	Esiti dei test per la classe Stato	28
5.4.1.2	Esiti dei test per la classe RegolaDiProduzione	29
5.4.1.3	Esiti dei test per la classe NonTerminale	29
5.4.1.4	Esiti dei test per la classe Solver	30
5.4.1.5	Esiti dei test per la classe Terminale	30
5.4.2	Esiti e copertura dei test di sistema	31
6.	Strutture dati e algoritmi	32
6.1	Terminale	32
6.1.1	Dati	32
6.1.2	Operazioni	32
6.2	NonTerminale	32
6.2.1	Dati	32
6.2.2	Operazioni	33
6.2.3	Algoritmi	34
6.2.3.1	calcolaInizi	34
6.3	RegolaDiProduzione	35
6.3.1	Dati	35
6.3.2	Operazioni	35
6.4	Stato	35
6.4.1	Dati	36
6.4.2	Operazioni	36
6.4.3	Algoritmi	36
6.4.3.1	aggiungiCore	37
6.4.3.2	espandiStato	40
6.4.3.3	aggiungiCompletamento	44
7.	Sviluppi futuri	45

1. Prefazione

1.1 Version History

- Versione 0.0 - 16/04/2019
- Versione 1.0 - 04/05/2019
- Versione 2.0 - 09/10/2019
- Versione 2.1 - 27/10/2019

1.2 Sommario modifiche

- Versione 0.0 - Definizione dei requisiti funzionali, dei casi d'uso, della struttura generale del software, della struttura dei test e degli strumenti utilizzati per lo sviluppo;
- Versione 1.0 - Introduzione dell'iterazione 1 con relativo piano di test ed esiti;
- Versione 2.0 - Introduzione dell'iterazione 2, modifiche design architetturale, nuovi casi d'uso, piano di test ed esiti;
- Versione 2.1 - Revisione documentazione.

2. Introduzione

2.1 Definizioni e glossario

Grammatica libera da contesto

In informatica e in linguistica, una grammatica libera dal contesto è una grammatica formale in cui ogni regola sintattica è espressa sotto forma di derivazione di un simbolo a sinistra a partire da uno o più simboli a destra.

Ciò può essere espresso con due simbolismi equivalenti:

1) $S := \alpha$

2) $S \rightarrow \alpha$

dove S è un simbolo detto *non terminale*, sostituibile con altri simboli non terminali e/o con simboli terminali, e α è una sequenza di simboli non terminali e/o terminali, ossia simboli che non possono essere sostituiti con null'altro.

L'espressione "libera dal contesto" si riferisce al fatto che il simbolo non terminale S può sempre essere sostituito da α , indipendentemente dai simboli che lo precedono o lo seguono; un linguaggio formale si dice libero dal contesto se esiste una grammatica libera dal contesto che lo genera.

Parser

Un parser LR è un parser di tipo bottom-up per grammatiche libere da contesto che legge il proprio input partendo da sinistra verso destra, producendo una derivazione a destra. Laddove indicato come parser $LR(k)$, il k si riferisce al numero di simboli letti (ma non "consumati") per prendere le decisioni di parsing.

2.2 Scopo

Scopo del programma che si andrà a realizzare è il riconoscimento di una grammatica $LR(1)$ contenuta in un file di input selezionato dall'utente.

2.3 Toolchain

Per lo sviluppo di questo progetto è stata scelta la seguente toolchain:

- GitHub, utilizzato per il controllo di versione, così da consentire il lavoro a più mani sul progetto anche qualora non ci si trovi tutti nello stesso ambiente. L'utilizzo di GitHub prevederà la creazione di branch secondari per gli sviluppi incrementali del programma, così da conservare sempre una versione funzionante del programma stesso all'interno del branch "master" della repository ospitante il progetto;
- GitKraken, client che fornisce un'interfaccia grafica intuitiva per l'utilizzo di GitHub, implementa funzioni che velocizzano l'utilizzo di git implementando un grafo rappresentante l'avanzamento della repository;
- Modelio, utilizzato per la creazione di diagrammi UML, si è scelto questo tool visto che implementa delle funzioni che partendo da codice java generano il modello UML del codice fornito. Essendo questo progetto un'espansione di un progetto già in essere realizzato per il corso Linguaggi Formali e Compilatori questo tool agevolerà la creazione del modello UML del codice già esistente;
- TexMaker, utilizzato per la redazione della documentazione;
- Eclipse, utilizzato per lo sviluppo dell'applicazione, inoltre integra dei tool utili al fine di analizzare la copertura del codice;
- ANTLR, tool per la generazione di parser, una volta definita una grammatica questo genera un parser e tutto il codice di contorno per l'individuazione di errori lessicali e semantici;
- JUnit, tool utilizzato per la definizione dei casi di test e l'esecuzione dei casi stessi;
- SpotBugs e PMD, plugin di Eclipse utilizzati per l'analisi statica del codice.

3. Iterazione 0

3.1 Requisiti funzionali

Il programma deve consentire all'utente le seguenti azioni:

- selezione del file di input da sottoporre al parsing e all'identificazione.

Il programma deve fornire le seguenti funzionalità:

- effettuare il parsing del file ricevuto in input, individuando eventuali errori sintattici, lessicali o semantici;
- qualora vengano individuati errori di qualsiasi genere nella fase di parsing, il programma deve comunicare i dettagli relativi agli errori individuati all'utente;
- qualora non vengano individuati errori nella fase di parsing, il programma deve procedere nell'identificare la grammatica come grammatica $LR(1)$ o non $LR(1)$.

3.1.1 Struttura corretta della grammatica

Il programma deve riconoscere come formalmente corrette (quindi prive di errori sintattici, lessicali e/o grammaticali) soltanto grammatiche che presentino la seguente struttura:

- una prima regola pr che abbia come elemento di sinistra il non terminale S_0 , definita come segue

$$SZ \ EQ \ NT \ TER \ SC$$

- altre $n \geq 1$ regole di produzione ar , che formano il resto della grammatica, definite come segue

$$NT \ EQ \ (NT|CT)^* \ SC$$

I blocchi componenti le regole appena definite sono così traducibili:

Simbolo	Caratteri
SZ	S_0
EQ	$- \ > \ \ :=$
NT	$A \ \dots \ Z$
CT	$a \ \dots \ z \ \ 0 \ \dots \ 9 \ \ + \ \ - \ \ * \ \ /$
TER	$/swa \ \ /cjswa$
SC	$;$

Tabella 1: Corrispondenza tra caratteri della grammatica e blocchi di definizione delle regole

Nota: Per la definizione della struttura delle regole è stata utilizzata la notazione formale di Backus-Naur estesa (EBNF).

3.1.2 Errori lessicali

L'utilizzo di qualsiasi carattere non riconducibile alla colonna "Caratteri" della Tabella 1 corrisponde a un errore lessicale.

3.1.3 Errori sintattici

Gli errori sintattici sono dati dal mancato rispetto della struttura delle regole *pr* e *ar* come definite nel paragrafo "Struttura corretta della grammatica" a pagina 6.

3.1.4 Errori semantici

Gli errori semantici si verificano nei seguenti casi:

- nella grammatica è presente un carattere non terminale che non presenta regole di produzioni associate;
- nella grammatica è presente una regola duplicata (nota bene questo non è un errore bloccante).

3.2 Casi d'uso

In questa fase preliminare si è identificato un solo caso d'uso riportato di seguito

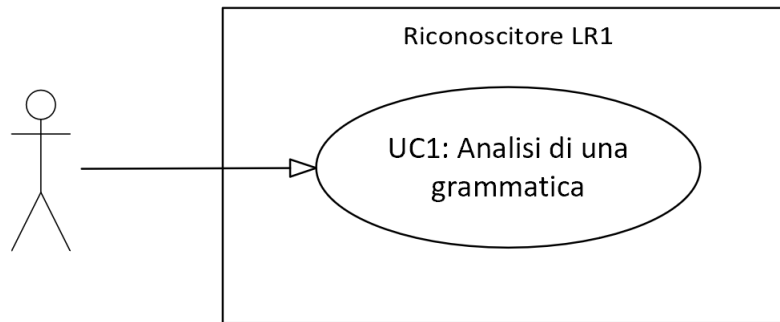


Figura 1: UML Use Case Diagram dell'iterazione 0

3.2.1 UC1 - Analisi di una grammatica

Descrizione: identificazione di una grammatica $LR(1)$

Attori coinvolti: utente

Precondizioni: esistenza del file d'input contenente la definizione della grammatica

Trigger: necessità di identificare una grammatica

Post condizioni: la classificazione della grammatica viene mostrata a schermo

Procedimento standard:

1. l'utente avvia il programma;
2. l'utente seleziona il file di input desiderato;
3. il programma esegue il parsing e l'analisi del file di input;
4. il programma mostra a schermo l'esito dell'analisi della grammatica;
5. l'utente chiude il programma.

Procedimenti alternativi o eccezioni:

- allo step 3, il programma rileva errori nella grammatica
viene mostrato un messaggio d'errore contenente informazioni relative all'errore individuato, l'utente chiude il programma e, corretta la grammatica, la procedura riparte dallo step 1

3.3 Struttura del programma

Il programma, definito in linguaggio Java, sarà composto di tre moduli:

- un primo modulo che verrà implementato all'interno del modulo generato da ANTLR, che si occuperà della lettura di un file .txt contenente la grammatica da identificare.
- un modulo generato da ANTLR, che si occuperà del parsing e dell'individuazione di errori lessicali e sintattici;
- un modulo scritto manualmente che si occuperà dell'analisi della grammatica e della sua classificazione.

L'interfacciamento con l'utente nella prima versione avverrà attraverso una CLI (Command Line Interface).

3.3.1 Scelta architetturale

Vista la struttura del programma si ritiene che la struttura ideale sia di tipo PIPES AND FILTERS, dove si possono identificare specifici moduli e ogni modulo genera dei risultati intermedi che verranno poi utilizzati dai moduli successivi.

In figura è riportata una bozza della struttura implementata nella fase 1.

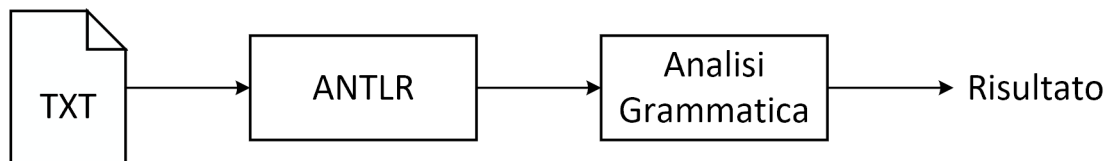


Figura 2: Bozza dell'architettura

3.4 Programma di testing

Il testing del programma sarà incentrato sul modulo *non* generato da ANTLR, assumendo che i controlli su quello generato da ANTLR siano già stati svolti dal produttore del tool, e sarà composto di:

- casi di test per ogni classe generata, scritte in JUnit, per i quali sarà richiesta una coverage del codice $\geq 95\%$;
- verifiche di coverage del codice scritto tramite l'esecuzione del programma con diversi input, al fine di garantire l'assenza di dead code;
- analisi statica del codice tramite strumenti quali SpotBugs e PMD.

A ogni revisione del codice, nonché in precedenza al rilascio di una nuova versione, saranno svolti test di non regressione per evitare l'introduzione di nuovi difetti nel codice.

4. Iterazione 1

In questa iterazione è stata creata la versione CLI (Command Line Interface), è stata inserita un'unica interfaccia grafica utilizzata per permettere all'utente la selezione del file che contiene la grammatica da analizzare, l'esito dell'analisi è riportato all'interno della CLI indicando se la grammatica è o meno $LR(1)$, l'elenco di tutti gli stati ed infine un elenco di tutte le transizioni.

4.1 Design Architettuale

In questa prima versione il design architettuale è rimasto quello previsto nella iterazione 0.

4.1.1 Class diagram

Per la generazione del class diagram della fase 1 è stato utilizzato il tool "Java Designer" di Modelio che permette di estrapolare dal codice già presente il diagramma UML delle classi. In particolare si può notare come la struttura sequenziale del programma è molto simile a quella ipotizzata nella fase zero a pagina 9.

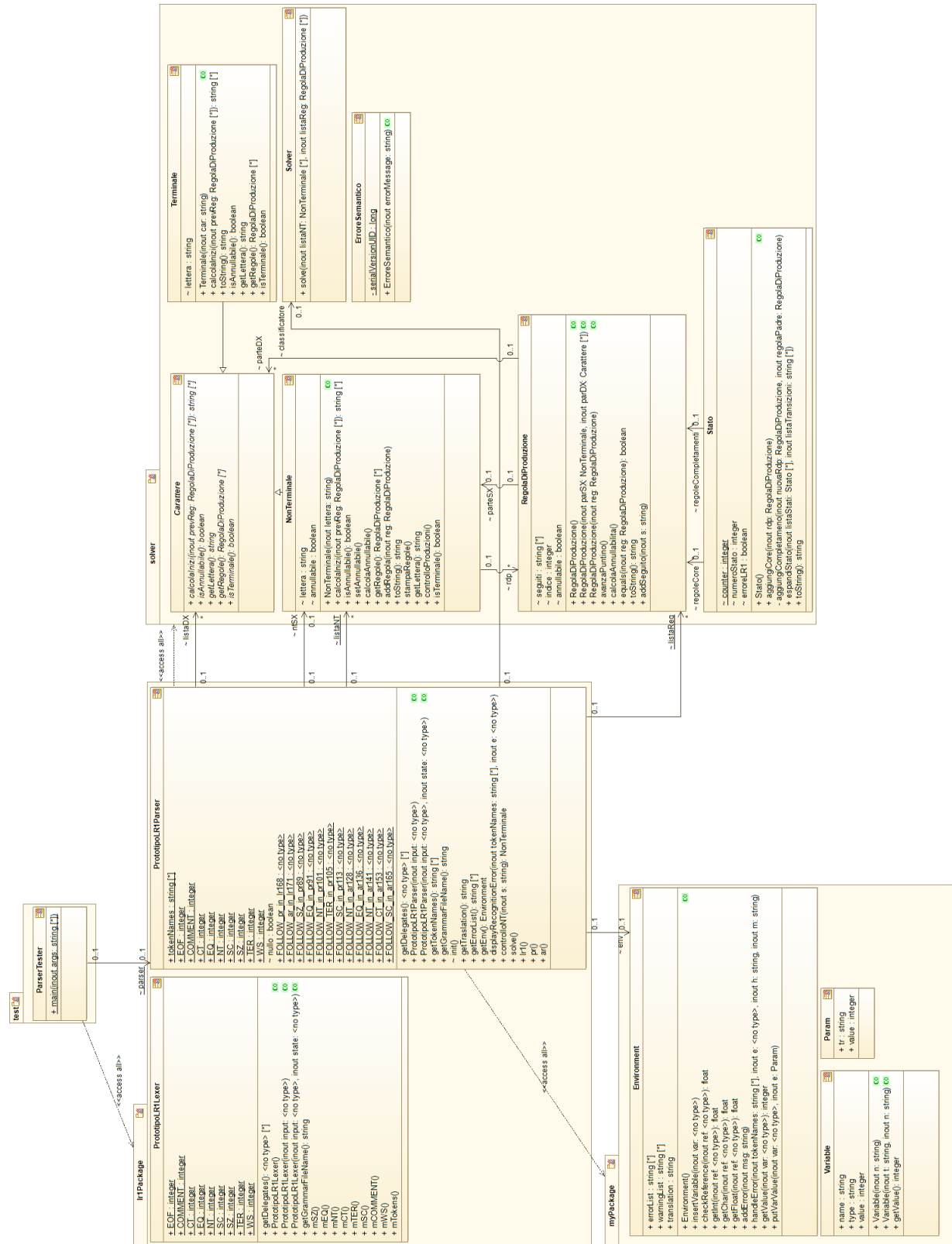


Figura 3: UML Class Diagram delle funzionalità dell'applicazione nell'iterazione 1

4.2 Testing

4.2.1 Descrizione del prodotto software

Il software, sviluppato in linguaggio Java in ambiente di sviluppo Eclipse, è suddiviso nei 4 packages descritti di seguito:

- **lr1package** - questo package contiene due degli output di ANTLR, strumento utilizzato per generare il compilatore, ossia *PrototipoLR1Lexer.java* e *PrototipoLR1Parser.java*; il terzo e ultimo file di output (*PrototipoLR1.tokens* è invece situato all'esterno di questo package;
- **main_package** - questo package contiene il file *Riconoscitore.java*, all'interno del quale è definito il metodo *main* che permette l'effettiva esecuzione del programma;
- **myPackage** - questo package contiene il file *Environment.java*, all'interno del quale è definita la classe *Environment*, necessaria per il corretto funzionamento dei file output di ANTLR ed ereditata da un progetto di esempio;
- **solver** - questo package contiene tutti i file che contengono definizioni di classi custom utilizzate per il processo di riconoscimento delle grammatiche e per rappresentare gli elementi costituenti delle grammatiche.

La versione oggetto di test è la 1.0, pubblicata sul branch *master* della repository di GitHub del progetto (<https://github.com/d-presciani/progettoLFC>) in data 26/03/2019.

4.2.2 Funzionalità oggetto di test

Le funzionalità sottoposte a test di unità saranno quelle definite nelle classi contenute all'interno del package *solver*, essendo queste le classi scritte manualmente e quelle che consentono l'effettivo funzionamento del programma; si trascurano di effettuare test di unità sui restanti package in quanto generati automaticamente da ANTLR e difficilmente sottoponibili a test di tale granularità.

È previsto anche un test di sistema, fornendo al sistema diversi input, per verificare la corretta integrazione tra i vari componenti e, al contempo, il corretto funzionamento del programma nel suo complesso.

Oltre ai suddetti test, è prevista anche l'esecuzione di un'analisi statica del codice tramite i due plugin di Eclipse *SpotBugs* e *PMD*.

4.2.2.1 Elenco delle funzionalità oggetto di test

Di seguito vengono riportate le classi, con i relativi metodi, per cui saranno effettuati i test di unità.

- Stato
 - private void aggiungiCompletamento(RegolaDiProduzione nuovaRdp, RegolaDiProduzione regolaPadre)
 - public void aggiungiCore(RegolaDiProduzione rdp)
 - public Stato()
 - public void espandiStato(LinkedList<Stato> listaStati, LinkedList<String> listaTransizioni)
 - public String toString()
- RegolaDiProduzione
 - public RegolaDiProduzione()
 - public RegolaDiProduzione(NonTerminale parSX, List<Carattere> parDX)
 - public RegolaDiProduzione(RegolaDiProduzione reg)
 - public void addSeguito(String s)
 - public void avanzaPuntino()
 - public void calcolaAnnullabilita()
 - public boolean equals(Object o)
 - public String toString()
- NonTerminale
 - public NonTerminale (String lettera)
 - public void addRegola (RegolaDiProduzione reg)
 - public void calcolaAnnullabile()
 - public List<String> calcolaInizi(LinkedList<RegolaDiProduzione> prevReg)
 - public void controlloProduzioni()

- public String getLettera()
- public List<RegolaDiProduzione> getRegole()
- public boolean isAnnullabile()
- public boolean isTerminale()
- public void setAnnullabile()
- public void stampaRegole()
- public String toString()
- Solver
 - public boolean solve(LinkedList<NonTerminale> listaNT,
LinkedList<RegolaDiProduzione> listaReg)
- Terminale
 - public Terminale(String car)
 - public List<String> calcolaInizi(LinkedList<RegolaDiProduzione> prevReg)
 - public String getLettera()
 - public List<RegolaDiProduzione> getRegole()
 - public boolean isAnnullabile()
 - public boolean isTerminale()
 - public String toString()

4.3 Esiti dei test

4.3.1 Esiti e copertura dei test di unità

Di seguito sono riportati gli esiti dei test e la copertura degli stessi per ognuna delle classe precedentemente evidenziate come oggetto di test di unità; nel complesso tali test hanno consentito una copertura del package *solver* pari al 96,4%.














▼  solver	 96,4 %	2.101	79	2.180
>  Stato.java	 95,6 %	1.485	68	1.553
>  RegolaDiProduzione.java	 95,6 %	238	11	249
>  Carattere.java	100,0 %	3	0	3
>  ErroreSemantico.java	100,0 %	4	0	4
>  NonTerminale.java	 100,0 %	177	0	177
>  Solver.java	 100,0 %	165	0	165
>  Terminale.java	100,0 %	29	0	29

Figura 4: Copertura dei test per il package solver

4.3.1.1 Esiti dei test per la classe Stato

Per la classe Stato sono stati scritti, all'interno della classe StatoTest, 22 test, superati correttamente dal programma, che garantiscono una copertura della classe Stato pari al 95,6%.



Figura 5: Esiti dei test della classe Stato

4.3.1.2 Esiti dei test per la classe RegolaDiProduzione

Per la classe RegolaDiProduzione sono stati scritti, all'interno della classe RegolaDiProduzione-Test, 6 test, superati correttamente dal programma, che garantiscono una copertura della classe RegolaDiProduzione pari al 95,6%.

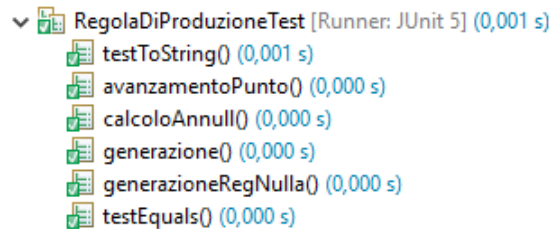


Figura 6: Esiti dei test della classe RegolaDiProduzione

4.3.1.3 Esiti dei test per la classe NonTerminale

Per la classe NonTerminale sono stati scritti, all'interno della classe NonTerminaleTest, 13 test, superati correttamente dal programma, che garantiscono una copertura della classe NonTerminale pari al 100,0%.

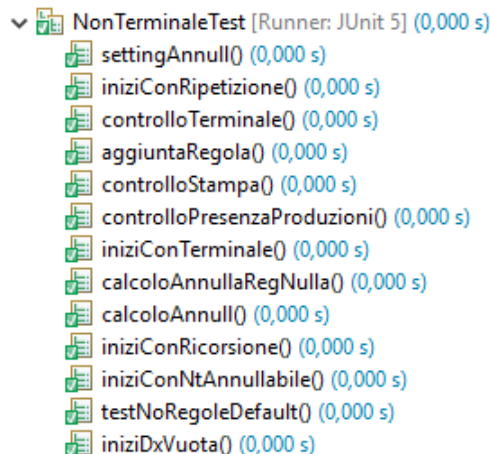


Figura 7: Esiti dei test della classe NonTerminale

4.3.1.4 Esiti dei test per la classe Solver

Per la classe Solver sono stati scritti, all'interno della classe SolverTest, 2 test, superati correttamente dal programma, che garantiscono una copertura della classe Solver pari al 100,0%.

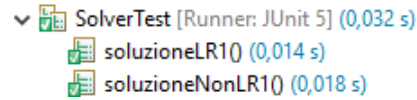


Figura 8: Esiti dei test della classe Solver

4.3.1.5 Esiti dei test per la classe Terminale

Per la classe Terminale sono stati scritti, all'interno della classe TerminaleTest, 7 test, superati correttamente dal programma, che garantiscono una copertura della classe Terminale pari al 100,0%.

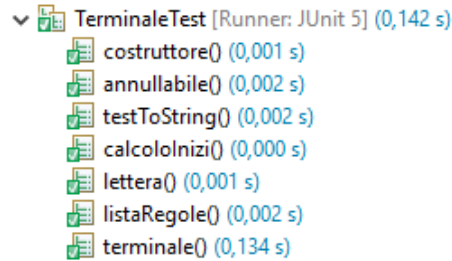


Figura 9: Esiti dei test della classe Terminale

4.3.2 Esiti e copertura dei test di sistema

I test di sistema sono stati effettuati eseguendo il metodo *main* della classe *Riconoscitore* più volte, passandogli come input i file contenuti all'interno della cartella "lfc\resources", presente nella repository del progetto (<https://github.com/d-presciani/progettoLFC>). Come oracoli per valutare la correttezza dei test sono stati utilizzati i temi svolti in classe durante il corso di Linguaggi formali e compilatori.

Durante i test di sistema è stata rilevata anche la copertura del codice nelle varie esecuzioni, nella seguente immagine è possibile vedere la copertura complessiva (ottenuta combinando la copertura delle singole esecuzioni).












Element	Coverage	Covered Instructions	Missed Instructions	Total Instructions
▼ Ir1package	 67,1 %	937	459	1.396
> ▢ PrototipoLR1Lexer.java	 48,9 %	343	358	701
> ▢ PrototipoLR1Parser.java	 85,5 %	594	101	695
▼ solver	 96,4 %	2.128	80	2.208
> ▢ Solver.java	 86,5 %	167	26	193
> ▢ Stato.java	 98,5 %	1.530	23	1.553
> ▢ NonTerminale.java	 89,3 %	158	19	177
> ▢ RegolaDiProduzione.java	 96,0 %	239	10	249
> ▢ Terminale.java	93,1 %	27	2	29
> ▢ Carattere.java	100,0 %	3	0	3
> ▢ ErroreSemantico.java	100,0 %	4	0	4
▼ myPackage	22,2 %	16	56	72
> ▢ Environment.java	 22,2 %	16	56	72
▼ main_package	 78,6 %	99	27	126
> ▢ Riconoscitore.java	 78,6 %	99	27	126

Figura 10: Copertura del codice relativa ai test di sistema

5. Iterazione 2

In questa iterazione è stata implementata un interfaccia grafica che permette di interagire con il programma in un ambiente completamente grafico.

È stata aggiunta inoltre la possibilità di visualizzare graficamente il risultato dell'analisi generando un grafo della grammatica analizzata.

5.1 Design Architetture

In questa seconda versione la struttura è rimasta di tipo PIPES AND FILTERS, ma è stata ampliata per includere la generazione e la visualizzazione del grafo, come si può vedere in figura.

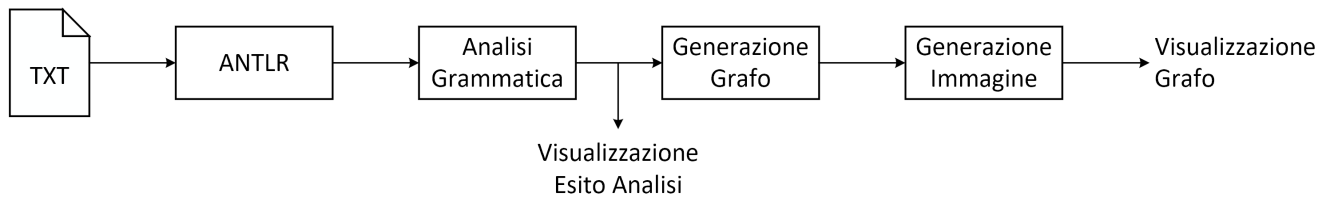


Figura 11: Bozza dell'architettura nella seconda iterazione

5.1.1 Librerie utilizzate

Per l'implementazione dell'interfaccia grafica e la creazione del grafo sono state utilizzate le seguenti librerie:

- JavaFX, framework grafico utilizzato per la realizzazione dell'interfaccia utente;
- JgraphT, libreria realizzata in swing che permette la creazione di grafi di vario tipo, implementando inoltre iteratori, può essere utilizzata per la creazione di algoritmi di vario tipo;
- mxGraph, libreria che permette la creazione di diagrammi, utilizzando SVG e HTML per il rendering delle immagini. In questo progetto questa libreria è utilizzata per la visualizzazione del grafico creato con JgraphT, in particolare questa libreria è stata scelta perché permette, se integrata in un JFrame, di modificare l'aspetto del grafico in realtime, oltre che offrire molte opzioni per la personalizzazione dei grafici creati. Nella pratica questa libreria è stata utilizzata per modificare l'aspetto delle box, cambiando colore agli stati che contengono delle regole che rendono il linguaggio non $LR(1)$, e per modificare lo stile grafico di rappresentazione.

Durante lo sviluppo di questa iterazione sono però sorte delle problematiche nell'integrazione delle librerie per la generazione del grafo e JavaFx.

Provando ad integrare all'interno dell'interfaccia utente un modulo sviluppato per testare le funzionalità di mxGraph che implementava la modifica realtime del grafo, sono stati riscontrati alcuni problemi di integrabilità essendo questa libreria realizzata in swing.

Per implementare questo modulo si è provato ad utilizzare la classe SwingNode per "incapsulare" questo modulo che integra componenti realizzate in swing e riuscire quindi a riutilizzare correttamente il modulo all'interno dell'interfaccia utente creata con JavaFX.

Dopo diversi tentativi si è riusciti ad integrare il modulo contenente il grafo con l'interfaccia utente realizzata con JavaFX, perdendo però la possibilità di editare in tempo reale il grafico, si è dunque scelto per motivi di performance di salvare l'immagine generata tramite mxGraph (la scelta del formato è lasciata all'utente) e poi caricare l'immagine salvata all'interno dell'interfaccia grafica per la visualizzazione del grafo.

5.1.2 Class Diagram

Rispetto al class diagram presente a pagina 12 in questa nuova iterazione si può vedere come è stato necessario modificare la struttura delle classi, per permettere il trasferimento di informazioni tra le stesse, in modo da poter mostrare all'interno dell'interfaccia grafica i vari messaggi da mostrare all'utente.

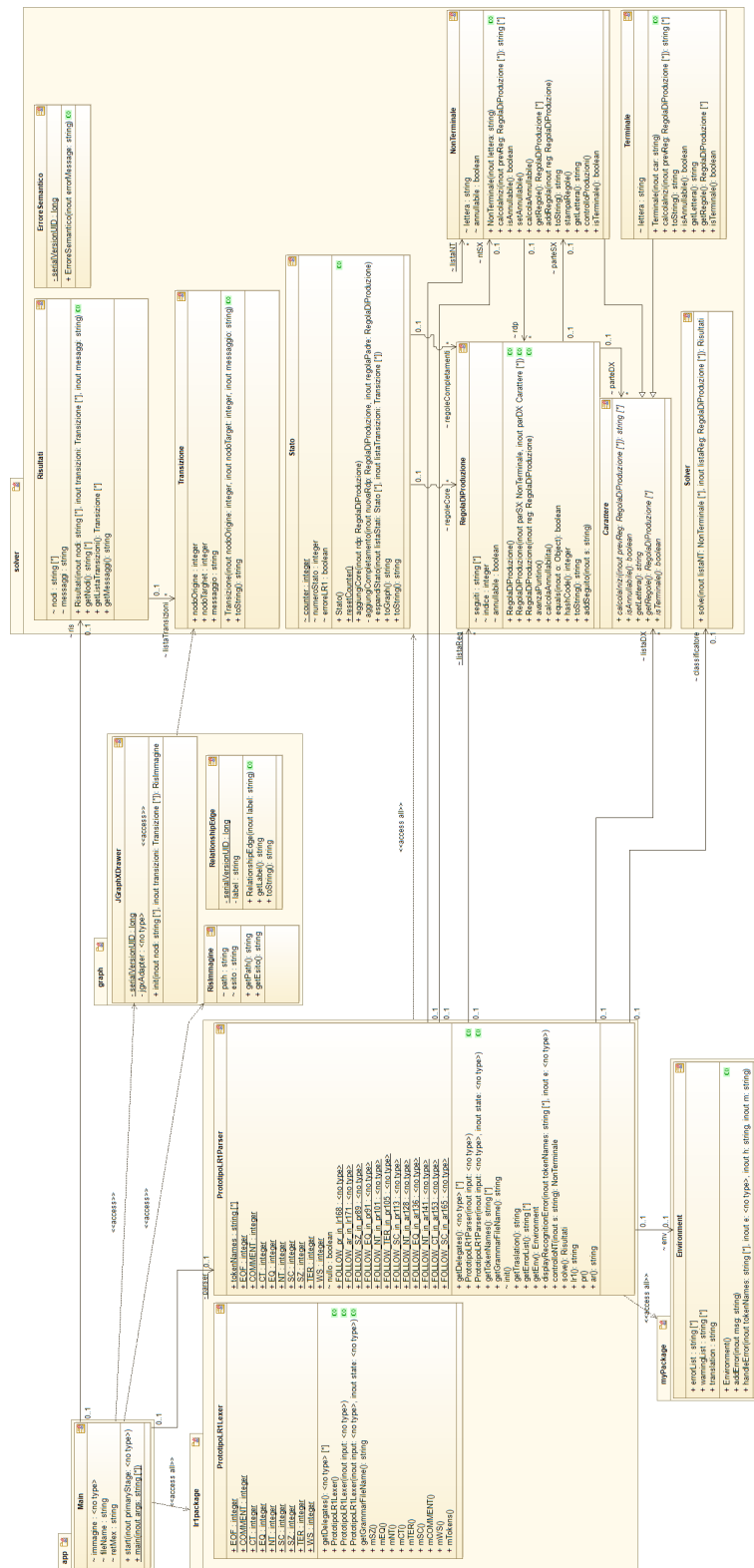


Figura 12: UML Class Diagram delle funzionalità dell'applicazione nell'iterazione 2

5.2 Casi d'uso

Vista l'introduzione di un'interfaccia grafica per l'interazione con il programma sono stati ridefiniti i casi d'uso come segue:

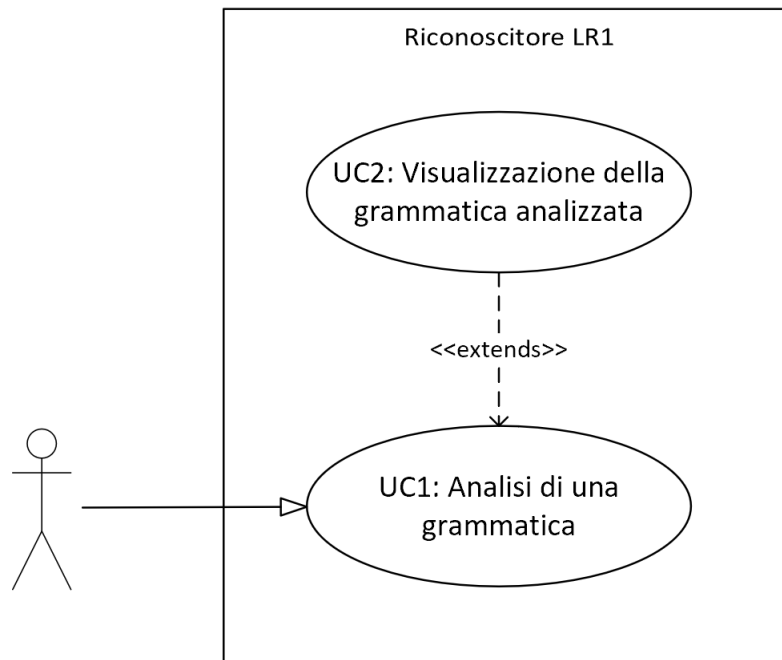


Figura 13: UML Use Case Diagram dell'iterazione 2

5.2.1 UC1 - Analisi di una grammatica

Descrizione: identificazione di una grammatica $LR(1)$

Attori coinvolti: utente

Precondizioni: esistenza del file d'input contenente la definizione della grammatica

Trigger: necessità di identificare una grammatica

Post condizioni: la classificazione della grammatica viene mostrata a schermo

Procedimento standard:

1. l'utente avvia il programma;
2. l'utente clicca il pulsante "carica file";
3. l'utente seleziona il file di input desiderato;
4. il programma esegue il parsing e l'analisi del file di input;
5. il programma mostra a schermo l'esito dell'analisi della grammatica.

Procedimenti alternativi o eccezioni:

- allo step 3, il programma rileva errori nella grammatica
viene mostrato un messaggio d'errore contenente le informazioni relative all'errore individuato, l'utente, corretta la grammatica riprende la procedura dallo step 2

5.2.2 UC2 - Visualizzazione della grammatica analizzata

Descrizione: visualizzazione della grammatica $LR(1)$ analizzata precedentemente

Attori coinvolti: utente

Precondizioni: analisi di una grammatica avvenuta correttamente

Trigger: necessità di visualizzare la grammatica identificata

Post condizioni: creazione di un file contenente il grafo relativo alla grammatica e visualizzazione dello stesso

Procedimento standard:

1. l'utente clicca il pulsante "genera e mostra grafo";
2. l'utente seleziona la posizione dove salvare il grafo e il formato del file creato;
3. il programma genera e mostra a schermo il grafo relativo alla grammatica.

Procedimenti alternativi o eccezioni:

- allo step 3, il programma rileva un errore in fase di generazione dell'immagine
viene mostrato un messaggio d'errore contenente informazioni relative all'errore individuato, l'utente, riprende la procedura dal punto 1 seguendo le indicazioni fornite dal programma

5.3 Testing

5.3.1 Descrizione del prodotto software

Il software, sviluppato in linguaggio Java in ambiente di sviluppo Eclipse, è suddiviso nei 5 packages descritti di seguito:

- **app** - questo package contiene il file *Main.java*, oltre ad altri file di supporto, utilizzati per la creazione dell'interfaccia grafica;
- **graph** - questo package contiene tutti i file necessari alla generazione del grafo;
- **lr1package** - questo package contiene due degli output di ANTLR, strumento utilizzato per generare il compilatore, ossia *PrototipoLR1Lexer.java* e *PrototipoLR1Parser.java*; il terzo e ultimo file di output (*PrototipoLR1.tokens* è invece situato all'esterno di questo package);
- **myPackage** - questo package contiene il file *Environment.java*, all'interno del quale è definita la classe *Environment*, necessaria per il corretto funzionamento dei file output di ANTLR ed ereditata da un progetto di esempio;
- **solver** - questo package contiene tutti i file che contengono definizioni di classi custom utilizzate per il processo di riconoscimento delle grammatiche e per rappresentare gli elementi costituenti delle grammatiche.

La versione oggetto di test è la 2.0, pubblicata sul branch *master* della repository di GitHub del progetto (<https://github.com/d-presciani/progettoLFC>) in data 09/10/2019.

5.3.2 Funzionalità oggetto di test

Le funzionalità sottoposte a test di unità saranno quelle definite nelle classi contenute all'interno del package *solver*, essendo queste le classi scritte manualmente e quelle che consentono l'effettivo funzionamento del programma; si trascurano di effettuare test di unità sui restanti package in quanto generati automaticamente da ANTLR e difficilmente sottoponibili a test di tale granularità.

È previsto anche un test di sistema, fornendo al sistema diversi input, per verificare la corretta integrazione tra i vari componenti e, al contempo, il corretto funzionamento del programma nel suo complesso.

Oltre ai suddetti test, è prevista anche l'esecuzione di un'analisi statica del codice tramite i due plugin di Eclipse *SpotBugs* e *PMD*.

5.3.2.1 Elenco delle funzionalità oggetto di test

Di seguito vengono riportate le classi, con i relativi metodi, per cui saranno effettuati i test di unità.

- Stato
 - private void aggiungiCompletamento(RegolaDiProduzione nuovaRdp, RegolaDiProduzione regolaPadre)
 - public void aggiungiCore(RegolaDiProduzione rdp)
 - public Stato()
 - public void espandiStato(LinkedList<Stato> listaStati, LinkedList<Transizione> listaTransizioni)
 - public String toString()
- RegolaDiProduzione
 - public RegolaDiProduzione()
 - public RegolaDiProduzione(NonTerminale parSX, List<Carattere> parDX)
 - public RegolaDiProduzione(RegolaDiProduzione reg)
 - public void addSeguito(String s)
 - public void avanzaPuntino()
 - public void calcolaAnnullabilita()
 - public boolean equals(Object o)
 - public String toString()
- NonTerminale
 - public NonTerminale (String lettera)
 - public void addRegola (RegolaDiProduzione reg)
 - public void calcolaAnnullabile()
 - public List<String> calcolaInizi(LinkedList<RegolaDiProduzione> prevReg)
 - public void controlloProduzioni() throws ErroreSemantico

- public String getLettera()
 - public List<RegolaDiProduzione> getRegole()
 - public boolean isAnnullabile()
 - public boolean isTerminale()
 - public void setAnnullabile()
 - public void stampaRegole()
 - public String toString()
- Solver
 - public Risultati solve(LinkedList<NonTerminale> listaNT,
LinkedList<RegolaDiProduzione> listaReg)
- Terminale
 - public Terminale(String car)
 - public List<String> calcolaInizi(LinkedList<RegolaDiProduzione> prevReg)
 - public String getLettera()
 - public List<RegolaDiProduzione> getRegole()
 - public boolean isAnnullabile()
 - public boolean isTerminale()
 - public String toString()

5.4 Esiti dei test

5.4.1 Esiti e copertura dei test di unità

Di seguito sono riportati gli esiti dei test e la copertura degli stessi per ognuna delle classi precedentemente evidenziate come oggetto di test di unità; nel complesso tali test hanno consentito una copertura del package *solver* pari al 95,6%.

solver		95,6 %	2.314	107	2.421
> Carattere.java		100,0 %	3	0	3
> ErroreSemantico.java		100,0 %	4	0	4
> NonTerminale.java		100,0 %	177	0	177
> RegolaDiProduzione.java		95,6 %	238	11	249
> Risultati.java		62,5 %	15	9	24
> Solver.java		100,0 %	162	0	162
> Stato.java		96,1 %	1.674	68	1.742
> Terminale.java		100,0 %	29	0	29
> Transizione.java		38,7 %	12	19	31

Figura 14: Copertura dei test per il package solver

5.4.1.1 Esiti dei test per la classe Stato

Per la classe Stato sono stati scritti, all'interno della classe StatoTest, 22 test, superati correttamente dal programma, che garantiscono una copertura della classe Stato pari al 96,1%.

StatoTest [Runner: JUnit 5] (0,001 s)
aggiuntaConCompPres() (0,000 s)
aggiuntaCompletamenti() (0,000 s)
testToString() (0,000 s)
espansioneStatoCore() (0,000 s)
aggiuntaCoreNoInizi() (0,000 s)
aggiuntaRegolaCoreIndice() (0,000 s)
controlloLR1Spostamento() (0,000 s)
espansioneStatoCompletamentiDuplicati() (0,000 s)
espansioneStatoCoreDuplicati() (0,000 s)
espansioneStatoCompletamento() (0,000 s)
espansioneDaComp() (0,000 s)
generazione() (0,000 s)
aggiuntaRegolaCoreNT() (0,000 s)
controlloLR1RiduzioneCore() (0,000 s)
aggiuntaCoreAnnullabileSeguitiPadre() (0,000 s)
ricalcoloInizi() (0,000 s)
testToStringNoComplementi() (0,000 s)
aggiuntaCoreNullaSeguitiPadre() (0,000 s)
controlloLR1RiduzioneCompletamenti() (0,000 s)
aggiuntaRegolaCore() (0,000 s)
aggiuntaRegolaCoreNulla() (0,000 s)
aggiuntaCoreCarattereSingolo() (0,000 s)

Figura 15: Esiti dei test della classe Stato

5.4.1.2 Esiti dei test per la classe RegolaDiProduzione

Per la classe RegolaDiProduzione sono stati scritti, all'interno della classe RegolaDiProduzioneTest, 6 test, superati correttamente dal programma, che garantiscono una copertura della classe RegolaDiProduzione pari al 95,6%.

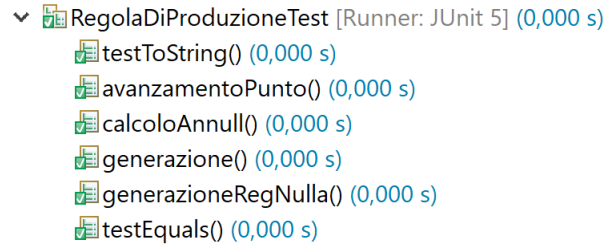


Figura 16: Esiti dei test della classe RegolaDiProduzione

5.4.1.3 Esiti dei test per la classe NonTerminale

Per la classe NonTerminale sono stati scritti, all'interno della classe NonTerminaleTest, 13 test, superati correttamente dal programma, che garantiscono una copertura della classe NonTerminale pari al 100,0%.

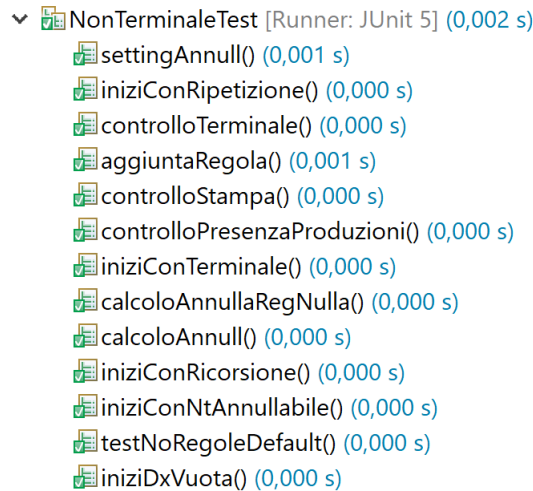


Figura 17: Esiti dei test della classe NonTerminale

5.4.1.4 Esiti dei test per la classe Solver

Per la classe Solver sono stati scritti, all'interno della classe SolverTest, 2 test, superati correttamente dal programma, che garantiscono una copertura della classe Solver pari al 100,0%.

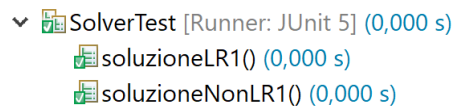


Figura 18: Esiti dei test della classe Solver

5.4.1.5 Esiti dei test per la classe Terminale

Per la classe Terminale sono stati scritti, all'interno della classe TerminaleTest, 7 test, superati correttamente dal programma, che garantiscono una copertura della classe Terminale pari al 100,0%.

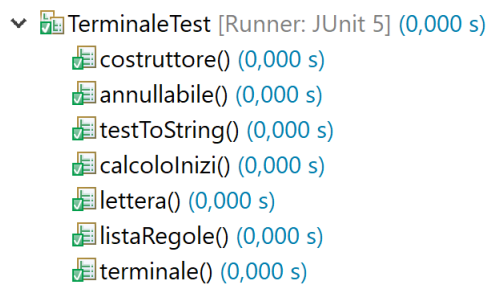


Figura 19: Esiti dei test della classe Terminale

5.4.2 Esiti e copertura dei test di sistema

I test di sistema sono stati effettuati eseguendo il metodo *main* della classe *app*, passandogli come input i file contenuti all'interno della cartella "lfc/resources", presente nella repository del progetto (<https://github.com/d-presciani/progettoLFC>). Come oracoli per valutare la correttezza dei test sono stati utilizzati i temi svolti in classe durante il corso di Linguaggi formali e compilatori.

Durante il test di sistema è stata rilevata anche la copertura del codice durante l'esecuzione, nella seguente immagine è possibile vedere la copertura complessiva.

Element	Coverage	Covered Instructions	Missed Instructions	Total Instructions
lfc	87,8 %	4.287	594	4.881
src	87,8 %	4.287	594	4.881
app	95,7 %	563	25	588
Main.java	95,7 %	563	25	588
graph	97,1 %	363	11	374
JGraphXDrawer.java	97,7 %	345	8	353
RelationshipEdge.java	75,0 %	9	3	12
RisImmagine.java	100,0 %	9	0	9
lr1package	72,6 %	1.035	391	1.426
PrototipoLR1Lexer.java	48,9 %	343	358	701
PrototipoLR1Parser.java	95,4 %	692	33	725
myPackage	100,0 %	72	0	72
Environment.java	100,0 %	72	0	72
solver	93,1 %	2.254	167	2.421
Carattere.java	100,0 %	3	0	3
ErroreSemantico.java	100,0 %	4	0	4
NonTerminale.java	86,4 %	153	24	177
RegolaDiProduzione.java	96,0 %	239	10	249
Risultati.java	100,0 %	24	0	24
Solver.java	100,0 %	162	0	162
Stato.java	93,6 %	1.630	112	1.742
Terminale.java	93,1 %	27	2	29
Transizione.java	38,7 %	12	19	31

Figura 20: Copertura del codice relativa al test di sistema

6. Strutture dati e algoritmi

In questa sezione si illustrano le strutture dati utilizzate e successivamente gli algoritmi più rilevanti implementati. Le strutture fondamentali che costituiscono le fondamenta delle strutture più complesse sono le classi `Terminale` e `NonTerminale`. Queste classi sono rappresentate di seguito.

6.1 Terminale

Questa classe viene utilizzata per rappresentare i caratteri terminali, questi caratteri sono i più semplici da gestire dato che la maggior parte delle proprietà che condividono con i caratteri `NonTerminale` sono ben definite.

6.1.1 Dati

- *lettera* una variabile utilizzata per memorizzare il carattere terminale.

6.1.2 Operazioni

- Costruttore(*car*): costruisce il `Terminale` memorizzando la variabile *car* ricevuta;
- CalcolaInizi: restituisce il carattere `Terminale`;
- toString: restituisce la rappresentazione testuale del carattere terminale;
- isAnnullabile: dato che un carattere terminale non è mai annullabile questa funzione restituisce sempre *false*;
- getLettera: restituisce la rappresentazione testuale del carattere terminale;
- getRegole: dato che un carattere terminale non ha regole associate questa funzione ritorna sempre *null*;
- isTerminale: questa funzione restituisce *true* dato che il carattere è un carattere terminale.

6.2 NonTerminale

Questa classe viene utilizzata per rappresentare i caratteri non terminali, questi caratteri hanno proprietà in comune con i caratteri terminali, ma come si può vedere dai dati memorizzati e dall'implementazione delle funzioni questi ultimi richiedono l'implementazione di particolari algoritmi per restituire alcune informazioni rilevanti.

6.2.1 Dati

- *lettera*: come per la classe `Terminale` questa variabile viene utilizzata per memorizzare la rappresentazione testuale del carattere;
- *rdp*: una `LinkedList` contenente tutte le regole di produzione che hanno questo non terminale come parte sinistra;
- *annullabile*: una flag utilizzata per memorizzare se il carattere è o meno annullabile senza dovere ricalcolare questa proprietà ogni volta.

6.2.2 Operazioni

- Costruttore(*lettera*): costruisce il Terminale memorizzando la variabile *lettera* ricevuta, inoltre inizializza la lista *rdp* come LinkedList vuota, viene infine inizializzata la variabile *annullabile* a false;
- CalcolaInizi: questa funzione calcola gli inizi del non terminale e li restituisce;
- calcolaAnnullabile: questa funzione deve essere chiamata al termine della creazione del non terminale e calcola se il terminale è o meno annullabile andando ad analizzare tutte le regole di produzione che ha associate;
- toString: restituisce la rappresentazione testuale del carattere terminale;
- isAnnullabile: getter del campo annullabile;
- getLettera: restituisce la rappresentazione testuale del carattere terminale;
- getRegole: questa funzione restituisce l'elenco delle regole di produzione associate al non terminale;
- isTerminale: questa funzione restituisce *false* dato che il carattere è un carattere non terminale.

6.2.3 Algoritmi

6.2.3.1 calcolaInizi

L'unico algoritmo di interesse in questa struttura dati è quello che consente di calcolare gli inizi del non terminale visualizzabile di seguito:

Algoritmo 1 *calcolaInizi*(*prevReg*: lista di regole di produzione già analizzate)

```

1: lista regole  $\leftarrow$  prevReg
2: creo una lista di tipo String vuota inizi
3: for RegolaDiProduzione reg : rdp do
4:   if la parte destra di reg non è vuota then
5:     if reg non è contenuta in regole then
6:       aggiungo reg a regole
7:       i = 0
8:       finito  $\leftarrow$  false
9:       do
10:        lista di String temp  $\leftarrow$  calcolaInizi(regole) per l'elemento i della parte destra
        di reg
11:        if temp non è vuota then
12:          for String ch : temp do
13:            if inizi non contiene ch then
14:              aggiungi ch ad inizi
15:            end if
16:          end for
17:        end if
18:        if l'elemento i della parte destra di reg è non è annullabile then
19:          finito  $\leftarrow$  true
20:        end if
21:        i + +
22:      while not finito && i minore della dimensione della parte destra di reg
23:    end if
24:  end if
25: end for
26: return inizi

```

Questo algoritmo contiene al suo interno una ricorsione, analizzando questa ricorsione si può notare che al massimo l'algoritmo richiama se stesso un numero di volte pari al numero di elementi contenuti nell'elenco delle regole generate dal parser. Chiamando r questo valore, chiamando s il numero di caratteri che rappresentano i seguiti, che al massimo possono essere il numero di caratteri terminali più il numero di caratteri non terminali (presi una sola volta), infine chiamando k la lunghezza della parte destra più lunga presente nella grammatica, si ottiene che questo algoritmo ha costo $O(r^2 sk)$.

6.3 RegolaDiProduzione

Questa classe viene utilizzata per rappresentare le varie regole di produzione, questa classe è stata creata come contenitore per semplificare il codice delle altre strutture dati

6.3.1 Dati

- *parteSX*: il non terminale presente a sinistra del simbolo di equivalenza;
- *parteDX*: una LinkedList di caratteri terminali e non;
- *seguiti*: una LinkedList che rappresentante i seguiti della regola di produzione (NOTA: i seguiti sono da intendersi in relazione alla posizione del puntino (*indice* in questa implementazione));
- *indice*: una variabile utilizzata per memorizzare la posizione dell'indice all'interno della parte destra della regola di produzione;
- *annullabile*: flag utilizzata per memorizzare se la regola è o meno annullabile senza dover ricalcolare questa proprietà tutte le volte.

6.3.2 Operazioni

- Costruttore(*parSX*, *parDX*): costruisce la regola di produzione partendo dai campi ricevuti, viene posto *indice* = 0;
- Costruttore(*reg*): costruisce la regola partendo da una regola esistente, questo costruttore viene utilizzato quando la regola è già esistente ma bisogna crearne una nuova uguale per poi incrementare il valore di *indice*;
- avanzaPuntino: incrementa di 1 il valore di *indice*;
- calcolaAnnullabilita: questa funzione deve essere chiamata al termine della creazione della regola di produzione e calcola se la regola è o meno annullabile andando ad analizzare i caratteri presenti in *parteDX*;
- equals(*o*): ridefinizione del metodo equals;
- toString: ritorna una stringa rappresentante la regola di produzione;
- addSeguito(*s*): aggiunge all'elenco dei seguiti il carattere ricevuto.

6.4 Stato

Questa è la struttura dati principale in questo progetto, qui sono presenti la maggior parte degli algoritmi che permettono di calcolare se una grammatica fornita è o meno di tipo $LR(1)$.

6.4.1 Dati

- *counter*: variabile statica utilizzata per il conteggio progressivo degli stati che vengono creati;
- *regoleCore*: LinkedList contenente le regole core dello stato;
- *regoleCompletamenti*: LinkedList contenente le regole di completamento dello stato;
- *numeroStato*: variabile utilizzata per la memorizzazione del numero dello stato;
- *erroreLR1*: flag utilizzata per memorizzare se uno stato contiene o meno delle regole che sono in conflitto e quindi rendono la grammatica non $LR(1)$.

6.4.2 Operazioni

- Costruttore: costruisce uno stato vuoto;
- resetCounter: funzione per resettare la variabile *counter* a 1;
- aggiungiCore(*rdp*): funzione che inserisce la regola ricevuta all'elenco delle regole core, successivamente analizza la regola inserita per verificare se questa aggiunge delle regole di completamento allo stato oppure modifica i seguiti delle regole già presenti;
- aggiungiCompletamento(*nuovaRdp*, *regolaPadre*): funzione che si occupa di aggiungere *nuovaRdp* alle regole di completamento e aggiunge i seguiti di *regolaPadre* in caso questa sia annullabile;
- espandiStato(*listaStati*, *listaTransizioni*): quando chiamata questa funzione genera nuovi stati esplorando ogni regola che lo compone;
- toGraph: questa funzione crea e ritorna una String che verrà utilizzata per la rappresentazione dello stato all'interno del grafo;
- toString: restituisce una string rappresentante lo stato.

6.4.3 Algoritmi

In questa classe sono implementati i principali algoritmi utilizzati per verificare che la grammatica fornita in input sia di tipo $LR(1)$, di seguito è presente lo pseudocodice degli algoritmi di maggior rilevanza.

6.4.3.1 aggiungiCore

Algoritmo 2 aggiungiCore(*rdp*: regola di produzione passata)

```

1: regolaCore  $\leftarrow$  rdp
2: if parte a destra della rdp > 0 then
3:   if parte a destra della rdp > dell'indice attuale del puntino and la parte a destra del
   puntino della rdp! = NULL then
4:     for RegolaDiProduzione regComp : regole alla destra del puntino della rdp do
5:       RegolaDiProduzione temporanea tmp  $\leftarrow$  regComp
6:       elimino i seguiti della RegolaDiProduzione tmp
7:       if indice puntino +1 < dimensione della parte a destra della rdp then
8:         i = 0
9:         do
10:          creo una lista di tipo string vuota emptyList
11:          if inizi della parte a destra del puntino +i + 1 della rdp non sono uguali alla
          lista emptyList then
12:            for string seg : inizi della parte a destra del puntino +i + 1 della rdp do
13:              if seguiti della RegolaDiProduzione tmp non contengono seg then
14:                aggiungo seg ai seguiti della RegolaDiProduzione tmp
15:              end if
16:            end for
17:          end if
18:          i ++
19:          while la regola di indice puntino +i della rdp è annullabile and la regola di
          indice puntino +i + 1 della rdp < dimensione della parte a destra della rdp
20:            if rdp è annullabile then
21:              for string seg : seguiti della rdp do
22:                if seguiti della RegolaDiProduzione tmp non contengono seg then
23:                  aggiungo seg ai seguiti della RegolaDiProduzione tmp
24:                end if
25:              end for
26:            end if
27:          else
28:            for string seg : seguiti della rdp do
29:              aggiungo seg ai seguiti della RegolaDiProduzione tmp
30:            end for
31:          end if
32:          booleanregPresente  $\leftarrow$  false
33:          for RegolaDiProduzione reg : regoleCompletamenti do
34:            if lettera a sinistra di RegolaDiProduzione reg è uguale alla lettera a sinistra
            della RegolaDiProduzione tmp and la parte a destra della RegolaDiProduzione reg è uguale
            alla parte a destra della RegolaDiProduzione tmp then
35:              for string seguito : seguiti della RegolaDiProduzione tmp do
36:                if seguiti della RegolaDiProduzione reg non contengono seguito then
37:                  aggiungo seguito ai seguiti della RegolaDiProduzione reg
38:                end if
39:              end for

```

```

40:         regPresente ← true
41:         break
42:     end if
43: end for
44:     if regPresente == false then
45:         aggiungo la RegolaDiProduzione tmp alle regole di completamento
         regoleCompletamenti
46:     end if
47: end for
48:     i ← 0
49:     while i < numero di regoleCompletamenti do
50:         RegolaDiProduzione reg ← regoleCompletamenti(i)
51:         if dimensione della parte destra della RegolaDiProduzione reg != 0 and regole a
         destra del puntino della RegolaDiProduzione reg != NULL then
52:             for RegolaDiProduzione nuoveRegole : regole dopo il puntino della
             RegolaDiProduzione reg do
53:                 booleantrovato ← false
54:                 for RegolaDiProduzione regCk : regoleCompletamenti do
55:                     if parte a sinistra di nuoveRegole == parte a sinistra di regCk and parte
                     a destra di nuoveRegole == parte a destra di regCk then
56:                         trovato ← true
57:                         for string seguito : seguiti della RegolaDiProduzione reg do
58:                             if seguiti della RegolaDiProduzione regCk non contengono
                             seguito then
59:                                 aggiungo seguito ai seguiti della RegolaDiProduzione regCk
60:                             end if
61:                         end for
62:                         break
63:                     end if
64:                 end for
65:                 if trovato == false then
66:                     inserisco reg in nuoveRegole
67:                 end if
68:             end for
69:         end if
70:         i ++
71:     end while
72:     booleanmodificato ← true
73:     while modificato == true do
74:         modificato ← false
75:         for indiceSx = 0 to indiceSx < dimensione di regoleCompletamenti do
76:             for indiceDx = 0 to indiceDx < dimensione di regoleCompletamenti do
77:                 if puntino della regola in indiceDx di regoleCompletamenti < dimensione
                 della parte destra della regola in indiceDx di regoleCompletamenti then
78:                     if carattere della parte sinistra della regola puntata a indice sinistro ==
                     carattere a destra del puntino della regola puntata a indice destro then

```

```

79:           if indice della regoleCompletamenti di indice destro +1 < dimensione
        della parte destra della regoleCompletamenti di indice destro then
80:           for stringseg : inizi della parte destra di regoleCompletamenti di
        indice destro +1 do
81:           if seguiti di indice sinistro di regoleCompletamenti non
        contengono seg then
82:           aggiungo seg ai seguiti di indice sinistro di
        regoleCompletamenti
83:           modificato  $\leftarrow$  true
84:           end if
85:           end for
86:           else
87:           for string seg : seguiti di regoleCompletamenti di indice destro
        do
88:           if seguiti di regoleCompletamenti di indice sinistro non
        contengono seg then
89:           aggiungo seg ai seguiti di indice sinistro di
        regoleCompletamenti
90:           modificato  $\leftarrow$  true
91:           end if
92:           end for
93:           end if
94:           end if
95:           end if
96:           end for
97:           end for
98:           end while
99:       end if
100: end if

```

L'algoritmo *aggiungiCore* fa uso dell'algoritmo *calcolaInizi* analizzato a pagina 34. Chiamando r il numero di regole generate dal parser, chiamando s il numero di caratteri che rappresentano i seguiti, che al massimo possono essere il numero di caratteri terminali più il numero di caratteri non terminali (presi una sola volta), e infine chiamando k la lunghezza della parte destra più lunga presente nella grammatica, si ottiene che questo algoritmo ha costo $O(rk * \text{calcolaInizi})$.

Considerando il costo analizzato precedentemente nel caso di *calcolaInizi* possiamo ottenere un costo di *aggiungiCore* pari a $O(r^3 k^2 s)$.

6.4.3.2 espandiStato

Algoritmo 3 *espandiStato*(*listaStati*: lista degli stati già creati, *listaTransizioni*: lista delle transizioni già create)

```

1: caratteriParsati ← linkedList di String vuota
2: for ciascuna regCor di regoleCore do
3:   if regCor ha almeno un carattere da esplorare a destra del puntino then
4:     giaParstato ← false
5:     for ciascuna str in caratteriParsati do
6:       if str è uguale al carattere a destra del puntino di regCore then
7:         giaParsato ← true
8:         break
9:       end if
10:    end for
11:    if giaParsato = false then    ▷ Se non è ancora stato analizzato il carattere puntato
    dal puntino allora lo si analizza
12:      listaRegoleNuovoStato ← LinkedList di regolaDiProduzione
13:      listaRegoleNuovoStato.add(regCor)
14:      for ciascuna altraRegCor di regoleCore dopo regCor do
15:        if altraRegCor ha almeno un carattere a destra del puntino and il carattere
    puntato da regCor e altraRegCor è uguale then
16:          listaRegoleNuovoStato.add(altraRegCor)    ▷ Il nuovo stato avrà come
    regole core tutte quelle regole che vedono spostato lo stesso carattere
17:        end if
18:      end for
19:      for ciascuna regComp di regoleComplementi do
20:        if regComp ha almeno un carattere a destra del puntino and il carattere puntato
    da regCor e regComp è uguale then
21:          listaRegoleNuovoStato.add(regComp)
22:        end if
23:      end for
24:    end if
25:    CarattereNuovo ← carattere puntato dall'indice di regCor
26:    caratteriParsati.add(carattereNuovo)
27:    for ciascuna regola di listaRegoleNuovoStato do
28:      regola.avanzaPuntino ▷ Sposto per ogni regola a destra il puntino per puntare un
    nuovo carattere
29:    end for
30:    trovato ← false
31:    for ciascuno stato in listaStati do
32:      if esiste uno stato con le stesse regoleCore contenute in listaRegoleNuovoStato
    then
33:        trovato ← true
34:      end if
35:    end for

```

```

36:      if trovato = false then
37:          temp  $\leftarrow$  nuovo stato vuoto
38:          for ciascuna regola in listaRegoleNuovoStato do
39:              temp.aggiungiCore(regola)           $\triangleright$  Come illustrato in Algoritmo 1, questa
              funzione si occupa di aggiungere la regola passata al nuovo stato e genera le eventuali regole di
              completamento
40:          end for
41:          listaStati.add(temp)           $\triangleright$  Aggiunta dello stato appena creato alla lista degli stati
              presenti
42:          listaTransizioni.add(newtransizione(numeroStato, temp.numeroStato, carattereNuovo)
43:      else
44:          listaTransizioni.add(newtransizione(numeroStato, numero stato già esistente con
              le stesse regole core di quello che si stava creando , carattereNuovo)
45:      end if
46:  end if
47: end for
48:           $\triangleright$  Una volta esplorate tutte le regoleCore si procede ad esplorare tutte le
              regoleCompletamenti
49: for ciascuna regComp in regleCompletamenti do
50:     if regComp ha almeno un carattere da esplorare a destra del puntino then
51:         giaParsato  $\leftarrow$  false
52:         for ciascuna str di caratteriParsati do
53:             if str è uguale al carattere a destra del puntino di regComp then
54:                 giaPasato  $\leftarrow$  true
55:                 break
56:             end if
57:         end for
58:         if giaParsato = false then           $\triangleright$  Se non è ancora stato analizzato il carattere puntato
              dal puntino di regComp allora lo si analizza
59:             listaRegoleNuovStato  $\leftarrow$  LinkedList di regoleDiProduzione
60:             listaRegoleNuovoStato.add(regComp)
61:             for ciascuna altraRegComp di regoleCompletamenti dopo regComp do
62:                 if altraRegComp ha almeno un carattere a destra del puntino and il carattere
                    puntato da regComp e altraRegComp è uguale then
63:                     listaRegoleNovoStato.add(altraRegComp)
64:                 end if
65:             end for
66:             carattereNuovo  $\leftarrow$  carattere puntato dall'indice di regComp
67:             caratteriParsati.add(carattereNuovo)
68:             for ciascuna regola di listaRegoleNuovoStato do
69:                 regola.avanzaPuntino()
70:             end for

```

```

71:         trovato  $\leftarrow$  false
72:         for ciascuno stato in listaStati do
73:             if esiste uno stato con le stesse regoleCore contenute in listaReoleNuovoStato
then
74:                 trovato  $\leftarrow$  true
75:             end if
76:         end for
77:         if trovato = false then
78:             temp  $\leftarrow$  nuovo stato vuoto
79:             for ciascuna regola in listaRegoleNuovoStato do
80:                 temp.aggiungiCore(regola)  $\triangleright$  Come illustrato in Algoritmo 1, questa
funzione si occupa di aggiungere la regola passata al nuovo stato e genera le eventuali regole di
completamento
81:             end for
82:             listaStati.add(temp)  $\triangleright$  Aggiunta dello stato appena creato alla lista degli stati
presenti
83:             listaTransizioni.add(newtransizione(numeroStato, temp.numeroStato, carattereNuovo)
84:         else
85:             listaTransizioni.add(newtransizione(numeroStato, numero stato già esistente
con le stesse regole core di quello che si stava creando , carattereNuovo)
86:         end if
87:     end if
88: end if
89: end for
90:  $\triangleright$  Dopo aver espanso lo stato attuale, controllo che lo stato appena espanso sia o meno LR(1)
91: sovrapposizione  $\leftarrow$  LinkedList di String vuota
92: for ciascuna reg di regoleCore do
93:     if L'indice di reg non ha caratteri alla sua destra then
94:         for ciascuna str di reg.seguiti do
95:             if sovrapposizioni contiene seg then
96:                 erroreLR(1) = true
97:                 break
98:             else
99:                 sovrapposizioni.add(seg)
100:            end if
101:        end for
102:    end if
103: end for

```

```
104: for ciascuna reg di regoleCompletamenti do
105:   if L'indice di reg non ha caratteri alla sua destra then
106:     for ciascuna str di reg.seguiti do
107:       if sovrapposizioni contiene seg then
108:         erroreLR(1) = true
109:         break
110:       else
111:         sovrapposizioni.add(seg)
112:       end if
113:     end for
114:   end if
115: end for
116: for ciascun carattereMosso di caratteriParsati do
117:   if sovrapposizioni contiene carattereMosso then
118:     erroreLR(1) = true
119:     break
120:   end if
121: end for
```

Per analizzare la complessità computazionale di questo algoritmo si introduce la seguente notazione:

- *rCore*: numero delle *regoleCore* dello stato
- *cPars*: numero di caratteri in *caratteriParsati*
- *rComp*: numero delle *regoleCompletamenti* dello stato
- *rStat*: numero delle nuove regole Core dello stato da creare
- *s*: numero degli stati in *listaStati*
- *seguiti*: numero di seguiti

Questo algoritmo vede diversi cicli, analizzandoli singolarmente otteniamo che:

- il primo ha costo $O(rCore(cPars + rCore + rComp + rStat + s + rStat * aggiungiCore()))$;
- il secondo ha costo $O(rComp(cPars + rComp + rStat + s + rStat * aggiungiCore()))$;
- gli altri cicli avendo al massimo due cicli for innestati non vengono analizzati dettagliatamente dato che si è verificato che hanno un costo minore ai precedenti.

Si riscrivono ora i costi utilizzando la notazione precedente, quindi si ha che:

- $rCore = rComp = rStat = O(r)$
- $cPars = seguiti = O(s)$

Si nota quindi che entrambi gli algoritmi hanno un costo asintotico di: $O(r^2 * aggiungiCore())$, sostituendo ora il costo di *aggiungiCore()*, si ottiene un costo totale per questo algoritmo pari a: $O(r^5 k^2 s)$.

6.4.3.3 aggiungiCompletamento

Algoritmo 4 *nuovaRdp*: nuova regola di produzione, *regolaPadre*: regola di produzione padre)

```

1: duplico la regola di produzione nuovaRdp in regolaTemp
2: if indice del puntino di regolaPadre +1 < dimensione della parte destra di regolaPadre then
3:    $i \leftarrow 0$ 
4:   do
5:     for String seg : inizi dei caratteri di indice puntino + $i$  + 1 della parte destra di
       regolaPadre do
6:       if seguiti di regolaTemp non contengono seg then
7:         aggiungo seg ai seguiti di regolaTemp
8:       end if
9:     end for
10:     $i++$ 
11:   while carattere di indice puntino + $i$  della parte destra di regolaPadre è annullabile and
       indice puntino + $i$  + 1 < dimensione della parte destra di regolaPadre
12:   verificaSegAnn  $\leftarrow true$ 
13:   for  $j = \text{indice puntino} + 1$  to  $j < \text{dimensione della parte destra di } regolaPadre$  do
14:     verificaSegAnn = verificaSegAnn && carattere di indice  $j$  della parte destra di
       regolaPadre è annullabile
15:   end for
16:   if regolaPadre è annullabile || dimensione della parte destra di regolaPadre uguale a indice
       puntino +1 || verificaSegAnn == true then
17:     for String seg : seguiti di regolaPadre do
18:       if seguiti di regolaTemp non contengono seg then
19:         aggiungo seg ai seguiti di regolaTemp
20:       end if
21:     end for
22:   end if
23: else
24:   for String seg : seguiti di regolaPadre do
25:     if seguiti di regolaTemp non contengono seg then
26:       aggiungo seg ai seguiti di regolaTemp
27:     end if
28:   end for
29: end if
30: aggiungo regolaTemp a regoleCompletamenti      ▷ Una volta aggiunti i seguiti, aggiungo la
       nuova regola alle regole di completamento

```

Utilizzando la notazione precedente, si ottiene che l'algoritmo *aggiungiCompletamento* ha costo $O(sk)$.

7. Sviluppi futuri

Il programma, allo stato attuale, potrebbe essere migliorato sia sotto l'aspetto dell'interfaccia grafica che delle funzionalità, ad esempio:

- espandere le capacità di analisi implementando dei moduli in grado di riconoscere altri tipi di grammatiche come $LALR(1)$ e $LL(1)$;
- aggiungere la possibilità di scrittura della grammatica direttamente nel programma, senza l'utilizzo di un file txt, in modo da evidenziare anche nella grammatica le regole problematiche;
- migrazione della generazione del grafo a librerie compatibili con JavaFX al fine di avere un grafo editabile;
- porting dell'applicazione su piattaforma web;
- porting dell'applicazione anche per dispositivi mobili.