UNIVERSITÀ DEGLI STUDI DI UDINE

Dipartimento di Scienze Matematiche, Informatiche e Fisiche



RELAZIONE PROGETTO DI ALGORITMI E STRUTTURE DATI

A cura di:

* Raffaele Marseu, matricola 133879

marseu.raffaele@spes.uniud.it

* Maksym Sernyak, matricola 143087

sernyak.maksym@spes.uniud.it

* Danjel Stefani, matricola 142863

stefani.danjel@spes.uniud.it

* Marco Venir, matricola 144762

venir.marco@spes.uniud.it

Anno Accademico 2019/2020

Indice

1. Introduzione Pag. 1
2. Presentazione algoritmi Pag. 1
   1. Metodi in comune Pag. 1
   2. Alberi binari di ricerca semplici Pag. 4
   3. Alberi binari di ricerca di tipo AVL Pag. 6
   4. Alberi binari di ricerca di tipo Red-Black Pag. 8
3. La funzione “*Creatore*” Pag. 13
4. La funzione *“Generatore”*  Pag. 14
5. Il metodo per misurare il tempo Pag. 14
6. Presentazione dei risultati e considerazioni Pag. 20

1.Introduzione

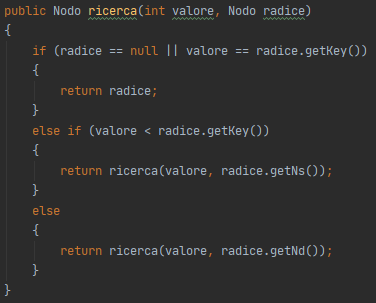
Abbiamo realizzato i programmi in Java sul programma Intellij, come ambiente di sviluppo e di run dei programmi. Abbiamo usato un computer con uno Xeon W3680 (3.33 GHz di base, 3.60 GHz Intel TurboBoost 1.0) con 6 core e 12 threads, equipaggiato con 20gb di ram DDR3 1333 MHz di velocità. Il sistema operativo usato e’ Windows 10 Pro. Come JDK abbiamo usato la versione 14.0.0.

2. Presentazione algoritmi

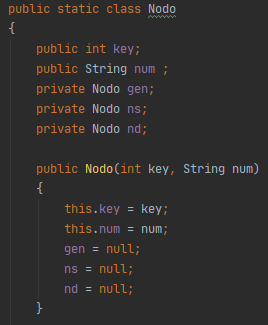
2.1 Metodi in comune:

I seguenti tre metodi sono utilizzati in tutti e tre gli algoritmi con delle piccole modifiche in base al tipo di albero utilizzato.

1. Nodo
2. Ricerca
3. Pre Order
4. Generatore interattivo



La ricerca è da manuale.



La classe “Nodo” utilizzata in tutti e 3 gli algoritmi, serve a inizializzare tutte le seguenti variabili:

key = chiave

num = nome del nodo

altezza = altezza dell’albero impostata di default come 1

ns = Nodo Sinistro

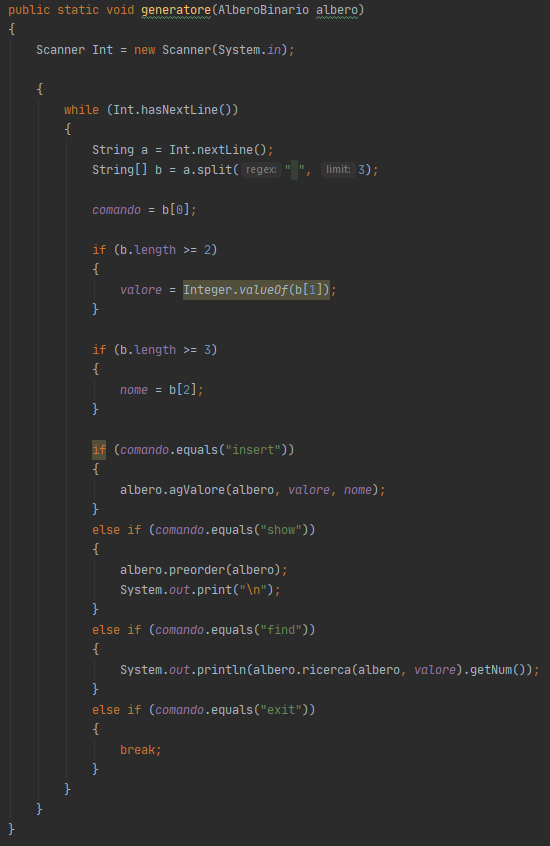
nd = Nodo Destro



PreOrder è da manuale

Generatore interattivo:

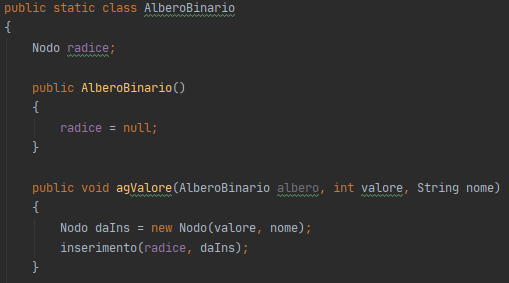
Metodo usato per rendere gli algoritmi interattivi e fornire l’output.  Utilizza la classe Scanner con il “hasNextLine” per leggere i comandi in input. Il while termina quando si inserisce la parola “exit”.   L’array “b” rappresenta una linea di input dove in prima posizione c’è il comando da eseguire, in seconda il valore, mentre la terza posizione rappresenta il nome. Gli if successivi determinano i comandi da eseguire in base all’istruzione desiderata inserita in input (che possono essere “insert”, “show”, “find” e “exit”).



2.2 Alberi binari di ricerca semplici

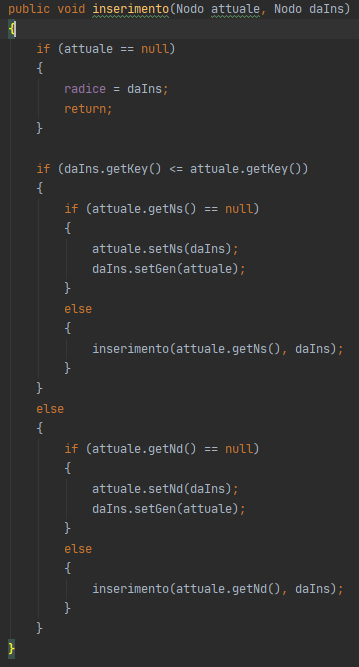
Dato che la classe “nodo” e le funzioni “ricerca” e “preorder”, sono le stesse per tutti, approfondiamo solo le funzioni “inserimento” e “agValore”, contenute nella classe “AlberoBinario”.

La funzione “agValore” serve per aggiungere all’albero un nodo, tramite la funzione “inserimento”. Il nodo da inserire è “daIns”.

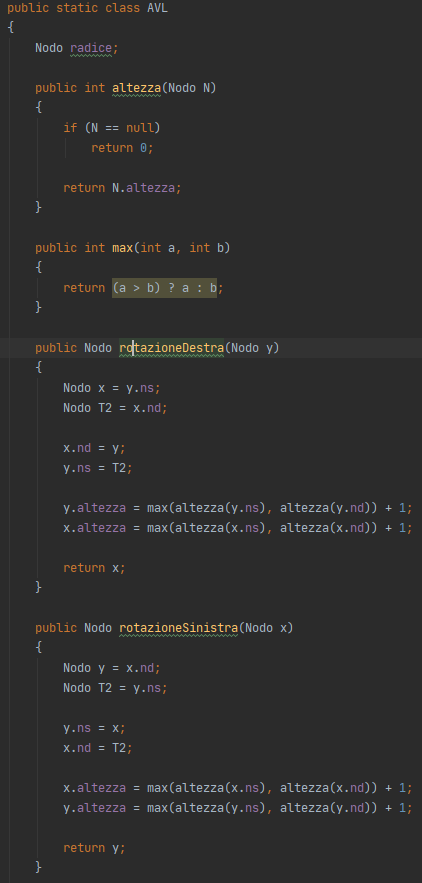


Nella funzione “inserimento”, il primo if verifica se l’albero è vuoto. In questo caso, il nodo da inserire (“daIns”), viene inserito come radice. La condizione successiva, controlla se il valore del nodo da inserire è minore o uguale dell’ultimo nodo preso in considerazione (“attuale”, nel primo caso è la radice, se esiste). In caso affermativo, si agisce sulla parte sinistra dell’albero, in caso contrario, sulla parte destra. Le istruzioni per l’aggiunta dei nodi figli sono le stesse per entrambe le parti dell’albero.

Si verifica se è presente il figlio (sinistro o destro) e in questo caso si richiama ricorsivamente la funzione “inserimento” utilizzando quest’ultimo come nodo “attuale”. Se il figlio (sinistro o destro) non è presente, si setta “daIns” come nodo “attuale” e di conseguenza quest’ultimo diventa genitore di “daIns”.



2.3 Alberi binari di ricerca di tipo AVL



Nella classe AVL, “altezza” ritorna l’altezza dell’albero, invece “max” è usato per trovare l’elemento più grande nelle due rotazioni.

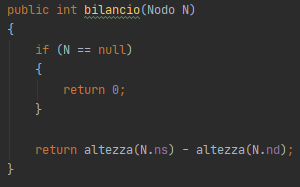
“rotazioneSinistra” e “rotazioneDestra” eseguono semplicemente le rotazioni per ribilanciare l’albero e ricalcolano l’altezza del nodo. 



Nella funzione “inserimento”, il primo if, quando trova un nodo “null”, inserisce il nuovo nodo in quella posizione. Durante la prima iterazione, viene creata la radice. I successivi due if, confrontano la key del nodo che si sta per inserire e quella del nodo di riferimento. Quindi settano ricorsivamente il figlio sinistro, se la key del nodo di riferimento è maggiore, il figlio destro altrimenti. Se le key sono uguali, ritorna il nodo.

A questo punto viene calcolata l’altezza del nodo, sommando 1 all’altezza massima tra i due figli. Successivamente si usa “bilancio”.

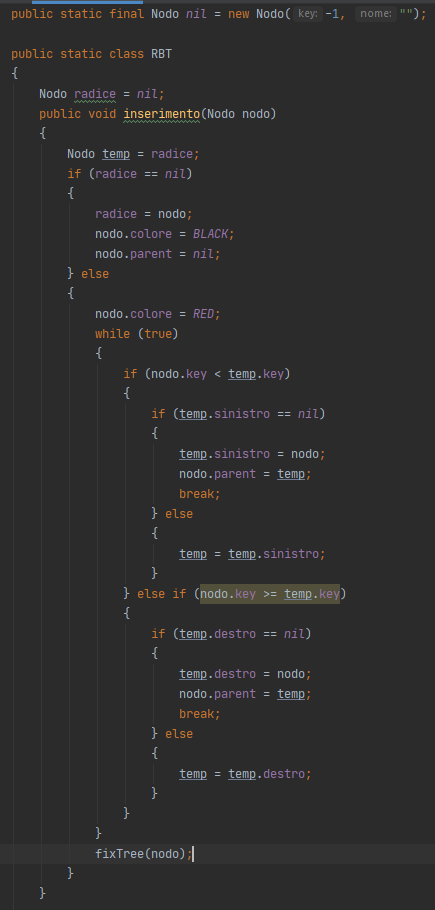
Gli if successivi stabiliscono, se ci sia bisogno e quale tipo di rotazione effettuare (se rotazione sinistra o rotazione destra).



Il metodo “bilancio”, calcola la differenza di altezza fra il figlio sinistro e il figlio destro.

2.4 Alberi binari di ricerca di tipo Red-Black

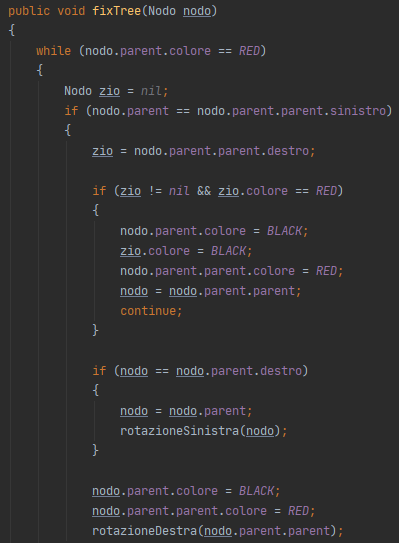
Il nodo “nil” è un nodo con key -1. Nella funzione inserimento, al primo inserimento la radice corrisponde al valore nil. Se la radice ha valore nil, il nodo da inserire (ovvero nodo), diventa radice, il colore del nodo viene settato a black e il parente del nodo diventa nil. Altrimenti, il colore del nodo viene settato a red e il suo valore viene comparato con gli altri nodi e viene inserito in una posizione vuota a sinistra o a destra in base al suo valore, come succede negli altri tipi di albero binario. Infine viene chiamata la funzione “fixTree”, che sistema l’albero.

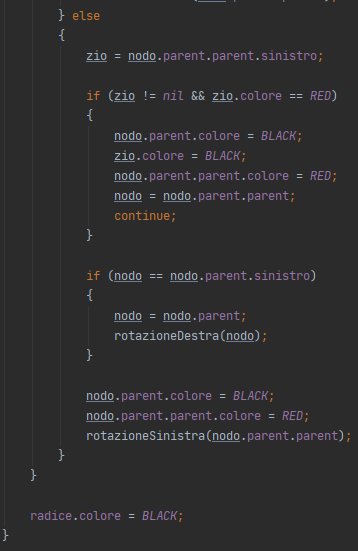


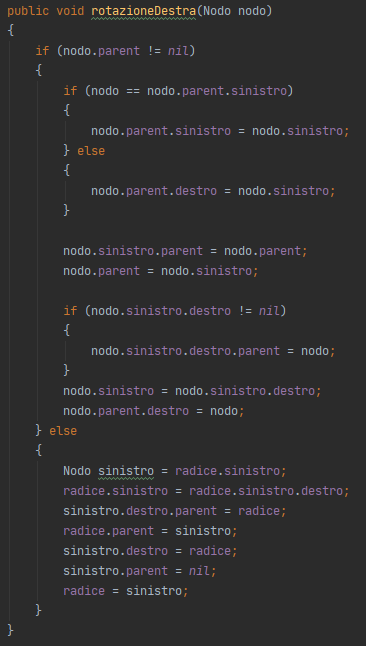
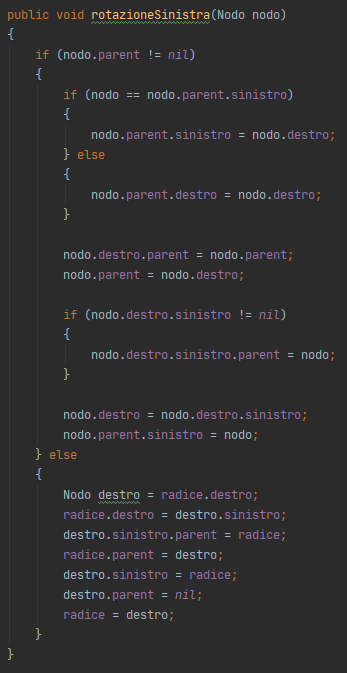
La funzione “fixTree” fa rispettare le 5 proprietà che valgono per i Red Black tree. Finchè il colore del parent del nodo (padre) è rosso, la variabile “zio” viene inizializzata a nil. Se il valore del nodo padre è uguale al nodo zio sinistro (figlio sinistro del nodo nonno), ridefiniamo la variabile “zio” come zio destro (figlio destro del nodo nonno). In seguito controlliamo se lo zio non è di colore rosso e non ha valore nil e in questo caso, vengono settati i colori: “parent” e “zio” a nero e il nodo genitore del genitore (nonno) a rosso. La variabile nodo, viene ridefinita come genitore del genitore (nonno). Se la variabile “nodo” equivale a parent.destro (al fratello destro), il nodo genitore diventa nodo e si esegue la funzione “rotazioneSinistra”. Infine, il colore del genitore viene settato a nero, mentre quello del  genitore del genitore (nonno) a rosso e si esegue “rotazioneDestra” del nonno.

Se il primo if è falso, viene eseguita la stessa procedura in modo speculare (quindi con lo zio sinistro al posto dello zio destro ecc.).

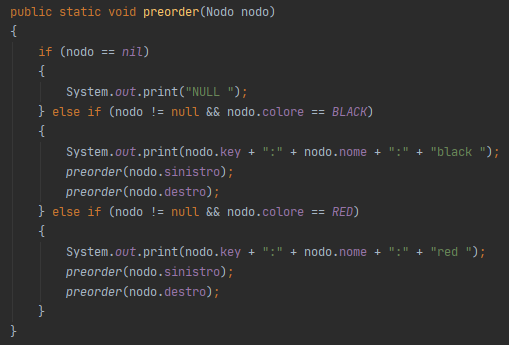
Alla fine della funzione, la radice viene colorata di nero.







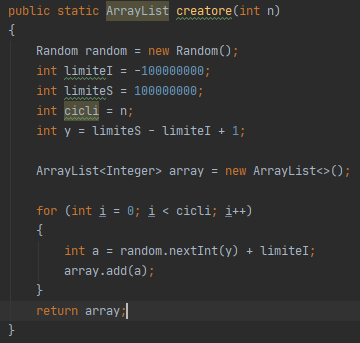
La funzione “rotazioneSinistra” è da manuale e la funzione “rotazioneDestra” è speculare alla precedente. Perciò, le spiegazioni sono state omesse.



La funzione “preorder” è da manuale, con i dovuti adattamenti affinché funzioni con i colori.

3. La funzione “*Creatore*”

Per utilizzare il “*Random*” predisposto dalla libreria di Java utilizziamo la stessa funzione che abbiamo usato nella prima Relazione. Impostiamo i limiti inferiore e superiore che rappresentano il minimo e massimo numero che il *“Random”* può selezionare. Per fare in modo che possano essere presenti anche numeri negativi, utilizziamo la variabile *“a*” in modo che l’intervallo di scelta sia *“y”*. Nella variabile *“y”*, il “+1” viene aggiunto, perché la funzione “nextInt()” non considera l’ultimo elemento del range. Sottraendo il limite inferiore, sarà possibile avere anche numeri negativi nell’array.



Ad esempio:

Limite Inferiore = -100;

Limite superiore = 100;

Range Y = Limite Superiore - Limite Inferiore + 1 = 100 + (-100) + 1 = 201;

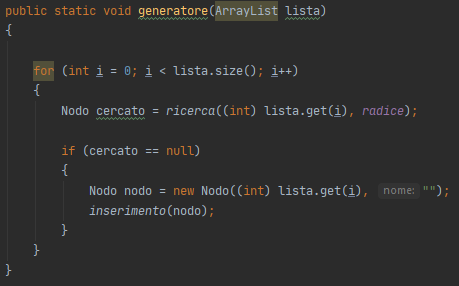
Random Y = 10;

Numero finale = Random Y + Limite inferiore = 10 + (-100) = -90;

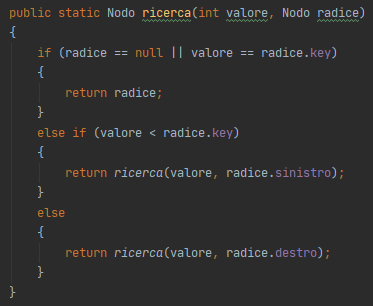
Nell’implementazione di tutti gli algoritmi, abbiamo utilizzato un range che va da

 -100.000.000 a 100.000.000.

4 La funzione *“Generatore”*



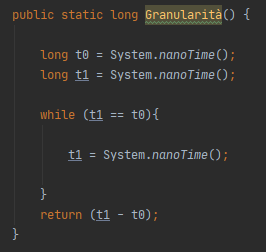
La funzione “generatore” cerca i valori della lista che è data in input. Se non vengono trovati, li inserisce come nuovi nodi nell’albero con la funzione “inserimento”.



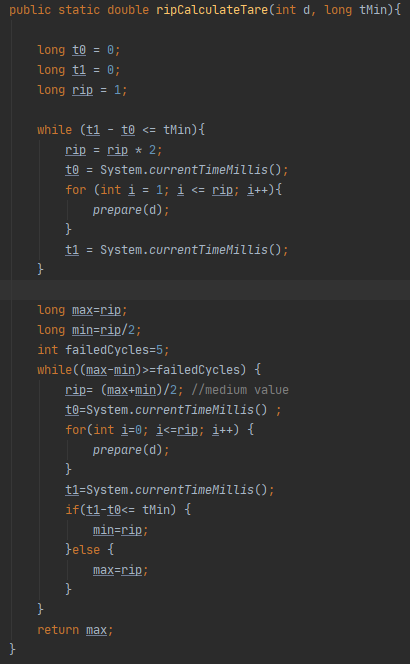
La funzione “ricerca” è da manuale, perciò non ci soffermiamo su questa spiegazione.

5 Il metodo per misurare il tempo

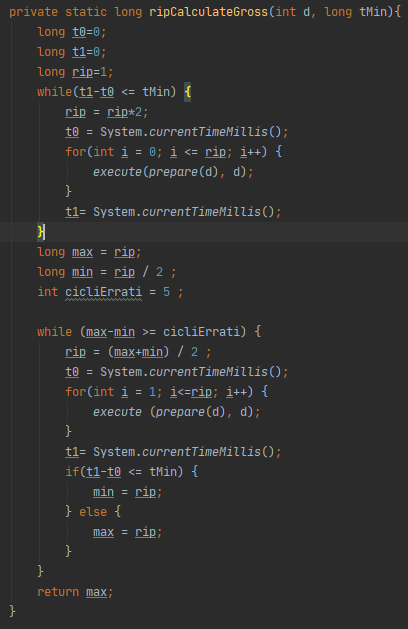
Per granularità (o risoluzione) si intende l’intervallo di misurazione necessario al sistema per caricare una singola istruzione, in questo caso noi carichiamo e tentiamo di ottenere il tempo minimo registrabile e ritorniamo la differenza tra i due risultati.



In “*ripCalculateTare*” calcoliamo il numero di ripetizioni minimo per far si che l’esecuzione di *“prepare”* sia registrato come tempo di preparazione e venga registrato come dato per conoscere il tempo per il caricamento dati tenendo conto anche dei cicli errati. Il tempo di *“prepare”* viene registrato per poterlo sottrarre al tempo totale ed avere il tempo di esecuzione vero e proprio dell’algoritmo.



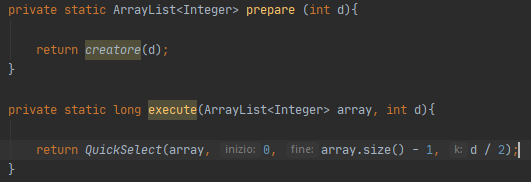
Nel primo while verifichiamo ogni quante ripetizioni la differenza dei due tempi (t0 e t1) continua ad essere minore del tMin (ovvero la granularità fratto la percentuale di errore dei dati, nel nostro caso 1%). La crescita di *“rip”* è esponenziale. Nel secondo while invece, considerando le ripetizioni, troviamo il tempo minore senza cicli falliti tramite il metodo numerico di bisezione.



*“ripCalculateGross”* calcola il tempo di esecuzione del programma in totale, compreso del tempo di esecuzione dell’algoritmo, il tempo per il caricamento dei dati e come per *“ripCalculateTare”*, vengono tenuti in considerazione anche i cicli errati che possono essere presenti. In questo caso abbiamo inserito un numero di al più 5 cicli errati.

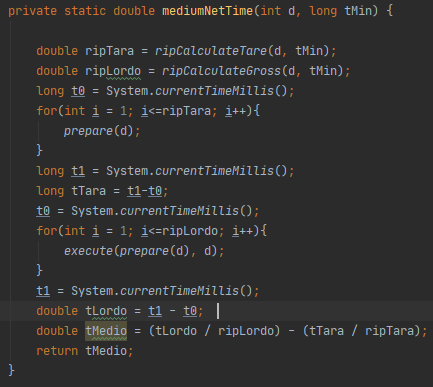
La stessa metodologia viene usata sia in *“ripCalculateTare”* sia in *“ripCalculateGross”*. Nella prima viene considerato il tempo di esecuzione della preparazione dei dati che andranno inseriti nell’algoritmo, mentre in *“ripCalculateGross”* viene considerato tutto il processo, ovvero dalla creazione dei dati allo svolgimento dell’algoritmo.

La funzione *“prepare”* inizializza l’algoritmo che successivamente andrà’ ad inserire i dati nell’array per lo svolgimento dell’algoritmo. *“execute*” invece esegue effettivamente l’algoritmo, senza contare come siano stati prodotti i dati in input.

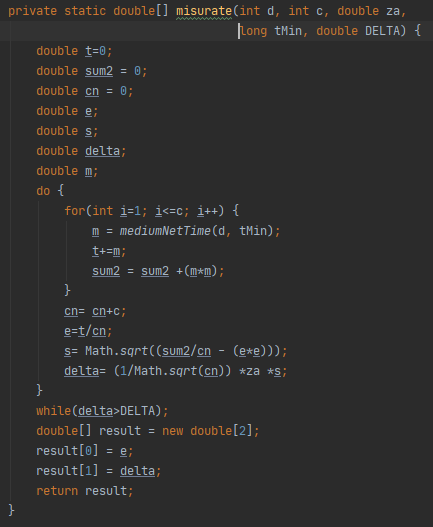


 In questa parte implementiamo il calcolo del tempo caricando il valore minimo (esecuzione della parte di caricamento dati) e il valore massimo, ovvero sfruttiamo il tempo impiegato dall’algoritmo per calcolare i risultati finali, prendendo in input i dati creati randomicamente.

In “mediumNetTime”, usiamo i vari metodi già realizzati (*“RipCalculateGross”*, *“ripCalculateTare”*, *“prepare”* ed *“execute”)*. Con il primo for controlliamo nuovamente il tempo di preparazione dei dati mentre nel secondo for, calcoliamo i tempi dell’esecuzione dell’algoritmo. Tutto ciò serve per calcolare il tempo medio sottraendo i tempi messi a rapporto con i valori precedentemente calcolati.



*“Misurate”* è la parte più importante dell’intero nucleo dell’algoritmo per la misurazione dei tempi, poiché vengono caricate tutte le variabili necessarie per inserire i valori delle ripetizioni, del *“delta”* e dell’errore della ripartizione della gaussiana (za). Tutto ciò serve proprio per collocare in un unico punto tutti gli input necessari alla ricerca dei tempi.



Con il do-while verifichiamo quando l’errore, ovvero la variabile *“DELTA”*, è maggiore del valore calcolato (variabile *“delta”*). “*delta*” è il parametro che serve a calcolare l’intervallo di confidenza. Se i valori superano l’1%, viene caricato il valore successivo. Nel for invece, viene tenuto il conto dei cicli di misurazioni da effettuare, ovvero tutti i risultati del tempo medio, tramite la funzione *“mediumNetTime”*.

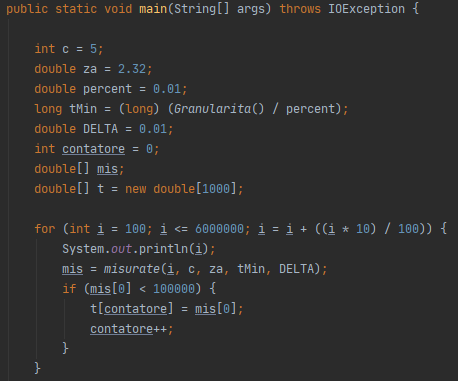
In output sono riportate le variabili *“e”* ed *”s”* che sono rispettivamente la media dei tempi e la deviazione standard.

Per calcolare la varianza campionaria abbiamo seguito questa formula. Mettendola sotto radice, troviamo la deviazione standard:

Nella formula, la prima parte (X(n,i)²) della sommatoria è rappresentata dalla variabile *“sum2”,* la seconda parte (dopo il segno meno), è rappresentata dalla variabile *“e”* al quadrato.

https://lh4.googleusercontent.com/q5VT_-RV06ZH8XBATyE4Z5DUs67RFW35AR1ncAHMKW53OwoaU1xxyevyE-Fum0f9HDP9eJLLs0HKAz2VRhXRZdy6-GGIIjN3VjDKHUz7kB1H091ThpDw9XB9bw50oy1Zd_r6SQ-b

Nel main abbiamo inizializzato le variabili principali, tra cui *“c”* che rappresenta il numero di misurazioni dell’algoritmo. Aumentando il valore, il numero di ripetizioni aumenta e di conseguenza migliora la precisione e aumenta il tempo totale di calcolo dei tempi.



Nel for memorizziamo tutti i tempi nell’array “t”. Il “contatore \* 2”, serve per inserire i dati che verranno trascritti sul foglio Excel. I dati verranno caricati con un ordine tale che il primo dei due risultati sia sempre il tempo e il secondo la deviazione standard. Abbiamo quindi inserito questo contatore solo per ordinare gli elementi dell'array “t”.



Calcoliamo il tempo ammortizzato direttamente in output, usando i tempi salvati, dividendo ogni tempo per il numero di operazioni.

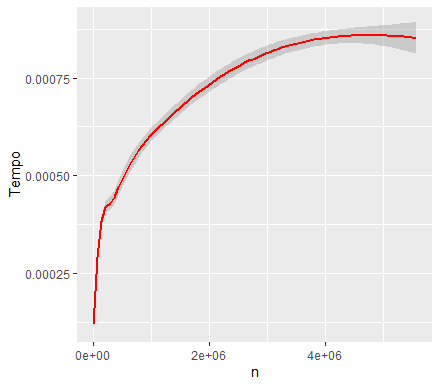
https://lh3.googleusercontent.com/Vei-7GUQ3J3Jyry9QgSi7cf5jRUgIPNPbIx5yj65bl712Tl4OSJp4tvOwDgd5-pKRvWeslE33kXKAf54lZ7LxfzR3kihdLCJY6slR_lV7NIYMNVT0TTQhzQAkxUW9GO5P7JwCEJ8

Una volta che l’algoritmo ha finito, importiamo i dati in output direttamente su Excel. La libreria che abbiamo usato per farlo è “hu.blackbelt.bundles.poi:org.apache.poi:4.0.1\_1”.

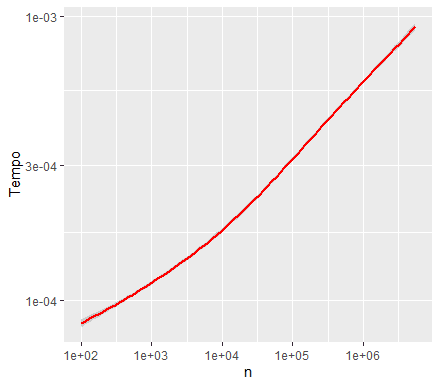
6 Presentazione dei risultati e considerazioni

Per fare i grafici abbiamo usato R. Nei grafici seguenti il Tempo è sempre rappresentato in millisecondi e collocato sull'asse y mentre *“n”* rappresenta il numero di operazioni di inserimento e di ricerca ed è collocato sull’asse x. I tempi sono stati presi con un aumento del 10% della lunghezza dell’array (il valore successivo è maggiore del 10% rispetto al precedente). Il limite superiore della grandezza del vettore è stato fissato a 6 milioni di elementi.

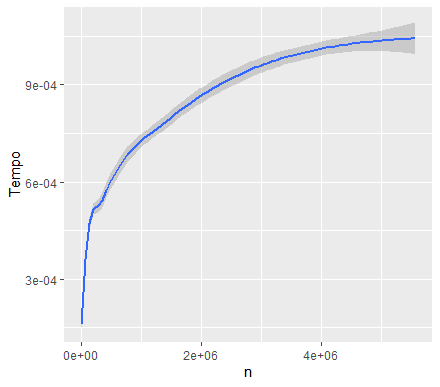
Nei grafici la linea in azzurro rappresenta regressione lineare dei dati mentre la parte in grigio esprime l’errore standard.



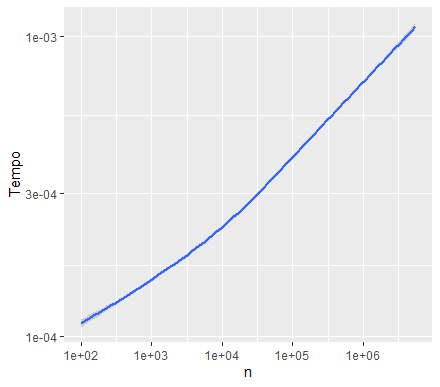
ABR con scala lineare (Albero binario di ricerca semplice)



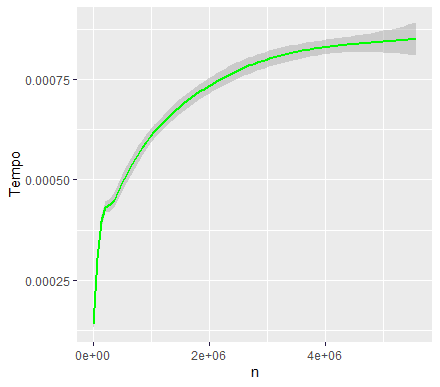
ABR con scala logaritmica (Albero binario di ricerca semplice)



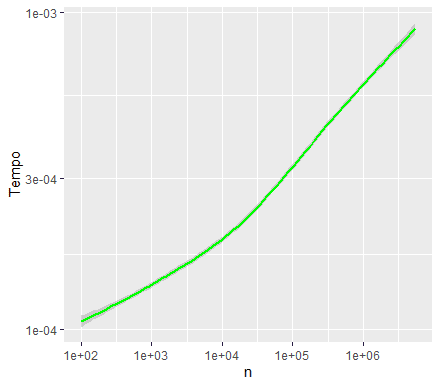
AVL con scala lineare



AVL con scala logaritmica



Red-Black con scala lineare



Red-Black con scala logaritmica

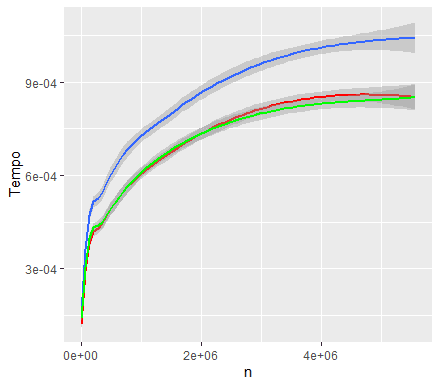
Sostanzialmente i grafici risultano coerenti con le nostre aspettative:

a)ABR ha una complessità temporale della ricerca e dell’inserimento di O(log n) nel caso medio e di O(n) nel caso peggiore (visto che i numeri sono generati in modo pseudo-randomico, il caso peggiore ha una bassa possibilità di presentarsi). Dal grafico si evince che abbiamo un andamento logaritmico, il che soddisfa le nostre aspettative sui tempi.

b)AVL ha una complessità temporale della ricerca e dell’inserimento di O(log n), sia nel caso medio sia nel caso peggiore. Dal grafico si può osservare che abbiamo un andamento logaritmico, il che rispetta le nostre aspettative.

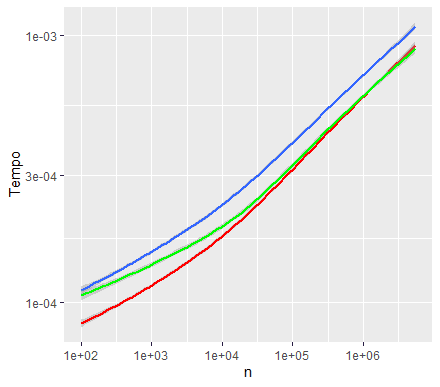
c)RedBlack ha una complessità temporale della ricerca e dell'inserimento di O(log n), sia nel caso medio sia nel caso peggiore. Dal grafico si può osservare che abbiamo un andamento logaritmico, il che rispetta le nostre aspettative.

Tutti gli algoritmi con scala lineare



(in rosso: ABR, in blu: AVL, in verde: RB)

Tutti gli algoritmi con scala logaritmica



(in rosso: ABR, in blu: AVL, in verde: RB)

Come si può vedere l’algoritmo più veloce fino a un milione è ABR. Dopo il milione, le prestazioni di Red Black e ABR sono molto simili, ma l’algoritmo RB è leggermente più veloce. AVL è l’algoritmo più lento dei tre.