

# **UNIVERSITÀ DEGLI STUDI DI SALERNO**

## **Facoltà di Scienze Matematiche Fisiche e Naturali**



### **CORSO DI LAUREA MAGISTRALE IN INFORMATICA**

INGEGNERIA GESTIONE ED EVOLUZIONE DEL SOFTWARE  
ANNO ACCADEMICO 2016/2017

### **MINING MAILING LISTS**

MANUALE UTENTE

VERSIONE 2.0  
22 DICEMBRE 2016

Coordinatore del progetto:

Nome
Prof. Andrea De Lucia

Candidato:

Nome	Matricola
ADP – Antonio De Piano	0522500397

### Revision History

Data	Versione	Descrizione	Autore
22/12/2016	1.0	Stesura documento	Antonio De Piano
23/12/2016	2.0	Revisione documento	Antonio De Piano

Sommario

1. Prerequisiti e architettura ..... 5

2. Mining di un singolo progetto ..... 6

2. Mining di più progetti ..... 7

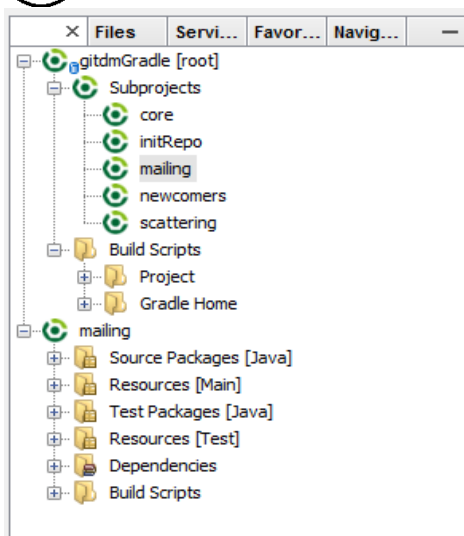
## 1. Prerequisiti e architettura

L'utente per poter eseguire l'applicazione Mining Mailing Lists "MML" dovrà disporre di una connessione a Internet e dell'IDE NetBeans con libreria Gradle.

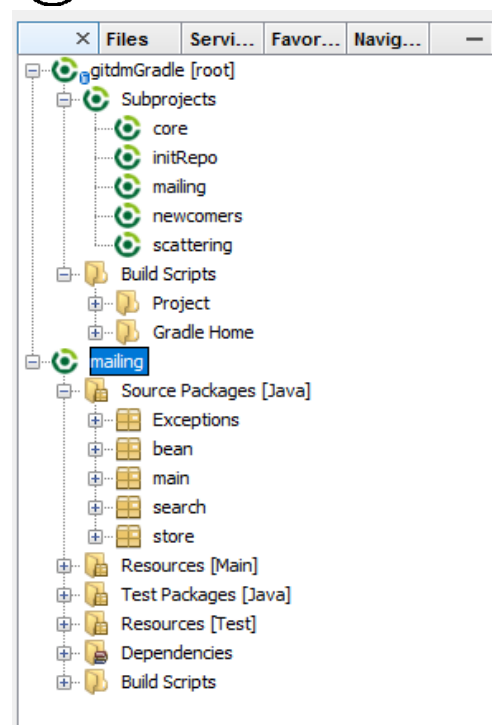
Il linguaggio di programmazione utilizzato per sviluppare l'applicazione è JAVA.

Di seguito viene mostrata la struttura del sottosistema MML di GitDM.

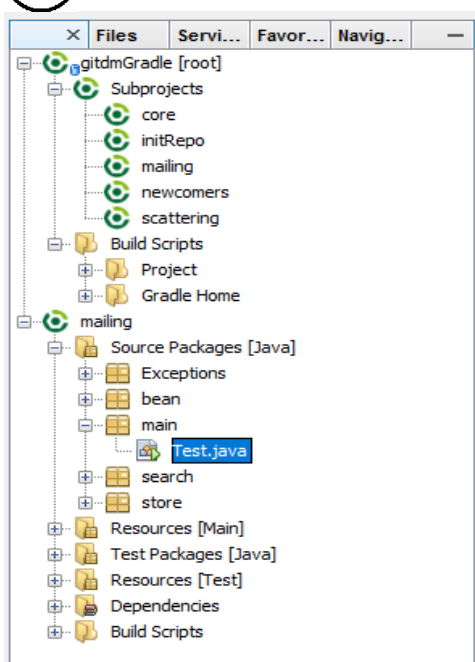
1



2



3



## 2. Mining di un singolo progetto

Per effettuare il mining mailing lists di un singolo progetto ([http://mail-archives.apache.org/mod\\_mbox/](http://mail-archives.apache.org/mod_mbox/)) basterà modificare la classe **Test.java** contenuta nel package main.

```
public class Test
{
    public static void main(String[] args)
    {
        try
        {
            Store st=new Store("lucene-general",2010,10,2010,12);
            //Scarico i messaggi dall'archivio
            if(st.download())
            {
                ArrayList<String> files=LuceneTester.test("Fork");
                ArrayList<Msg> list_msg=st.analyzer(files);
                for(int i=0;i<list_msg.size();i++)
                {
                    System.out.println(list_msg.get(i).toString());
                }
                st.saveMsgs(list_msg);
            }
            else
                System.out.println("\nError: invalid project name!\n");
        }
        catch(Exception e)
        {
            System.out.println(e.getMessage());
        }
    }
}
```

In quest'esempio verrà effettuato il download delle mailing lists dal progetto *lucene-general* da *ottobre 2010* a *dicembre 2010*. I nomi dei progetti disponibili sono presenti al seguente link ([http://mail-archives.apache.org/mod\\_mbox/](http://mail-archives.apache.org/mod_mbox/)).

Il download verrà effettuato nella directory store\_messages che l'utente potrà consultare sul suo desktop.

La directory store\_messages conterrà al suo interno, i file scaricati dall'archivio email di Apache Foundation e due sotto-directory. La directory messages, conterrà tutti i singoli messaggi, splittati e indicizzati da Apache Lucene, mentre la directory output conterrà i risultati di ricerca prodotti dall'applicazione.

Di seguito, il contenuto della directory store\_messages secondo l'input fornito nella classe **Test.java**.

messages	22/12/2016 11:35	Cartella di file	
output	22/12/2016 11:35	Cartella di file	
201010	22/12/2016 11:35	Documento di testo	217 KB
201011	22/12/2016 11:35	Documento di testo	652 KB
201012	22/12/2016 11:35	Documento di testo	359 KB

Per poter modificare le directory di lavoro di MML, basterà modificare la classe **Store.java** presente nel package Store, fornendo nuovi Path.

Di seguito i Path attuali utilizzati dall'applicazione.

```
this.url_base="http://mail-archives.apache.org/mod_mbox/"+this.name_project;

this.dir_output_msg=new File("C:\\Users\\depiano.it\\Desktop\\store_messages\\output");
this.dir_store = new File("C:\\Users\\depiano.it\\Desktop\\store_messages");
this.dir_split_msgs=new File("C:\\Users\\depiano.it\\Desktop\\store_messages\\messages");
```

## 2. Mining di più progetti

Per poter effettuare il mining di più progetti, bisognerà seguire la procedura illustrata per il mining di un singolo progetto. Dovrà essere modificata la classe **Test.java** e la classe **Store.java**, fornendo Path differenti per l'applicazione e ovviamente il nuovo nome del progetto.

Questa procedura, dovrà essere eseguita ad ogni esecuzione differente dell'applicazione, per non riscrivere i risultati prodotti dall'esecuzione precedente.