**Descrierea aplicaţiei**

**LabelMe**

1. **Motivaţie**

LabelMe este o aplicaţie desktop care permite gruparea şi organizarea e-mail-urilor, mesajelor, articolelor (în general texte scurte). Această aplicaţie reuşeşte datorită unor indicii (o listă de keywords, exemple de texte sau chiar pattern-uri). Clasificarea automată a textelor primite de la utilizator se va face în funcţie de categoriile definite anterior. Aplicaţia noastră permite clusterizarea (gruparea unor texte pe baza unor trăsături comune).

Astfel, folosindu-se de această aplicaţie, utilizatorul va avea la dispoziţie un filtru de mesaje. Acest filtru de mesaje este foarte important pentru persoanele care doresc memorarea mai rapidă şi mai eficientă a mesajelor (în mare parte a emai-urilor). Astfel, datorită aplicaţiei noastre, este posibilă memorarea textelor pe anumite categorii (servici, familie, sport, etc).

1. **Facilităţi**

În primă fază, utilizatorul va trebui să se înregistreze (să-şi creeze un cont), după care să se autentifice cu user-ul şi password-ul ales. Utilizatorul are la dispoziţie o listă de modalităţi de definire a unei categorii precum: adăugarea unei liste de keywords, adăugarea unor texte model, setarea limbii, adăugarea unor pattern-uri etc. Aplicaţia memorează toate acestea după care este afişat meniul Home. Utilizatorul poate grupa automat mesajele pe baza categoriilor predefinite (dacă acestea există). Dacă nu, acesta va fi redirecţionat pentru a seta noi categorii. Utilizatorul are la dispoziţie metoda de clusterizare (gruparea unor texte pe baza unor trăsături comune). Acesta va uploada noi texte/mesaje din arhive/fişiere, după care aplicaţia grupează aceste texte/mesaje pe posibile categorii.

1. **Tehnologii folosite**

*Apache Tika* – pentru detectarea limbii

*Algoritmul Naïve-Bayes* – clusterizarea şi definirea unor categorii

*Java* – limbaj de programare pentru a permite: Weka(data mining software), Java Text Categorizing Library(pentru selectarea limbii), Mallet(instrumente sofisticate de clasificare a documentelor), JAXP(bibliotecă pentru XML).