

# Проект по Information retrieval

## тема

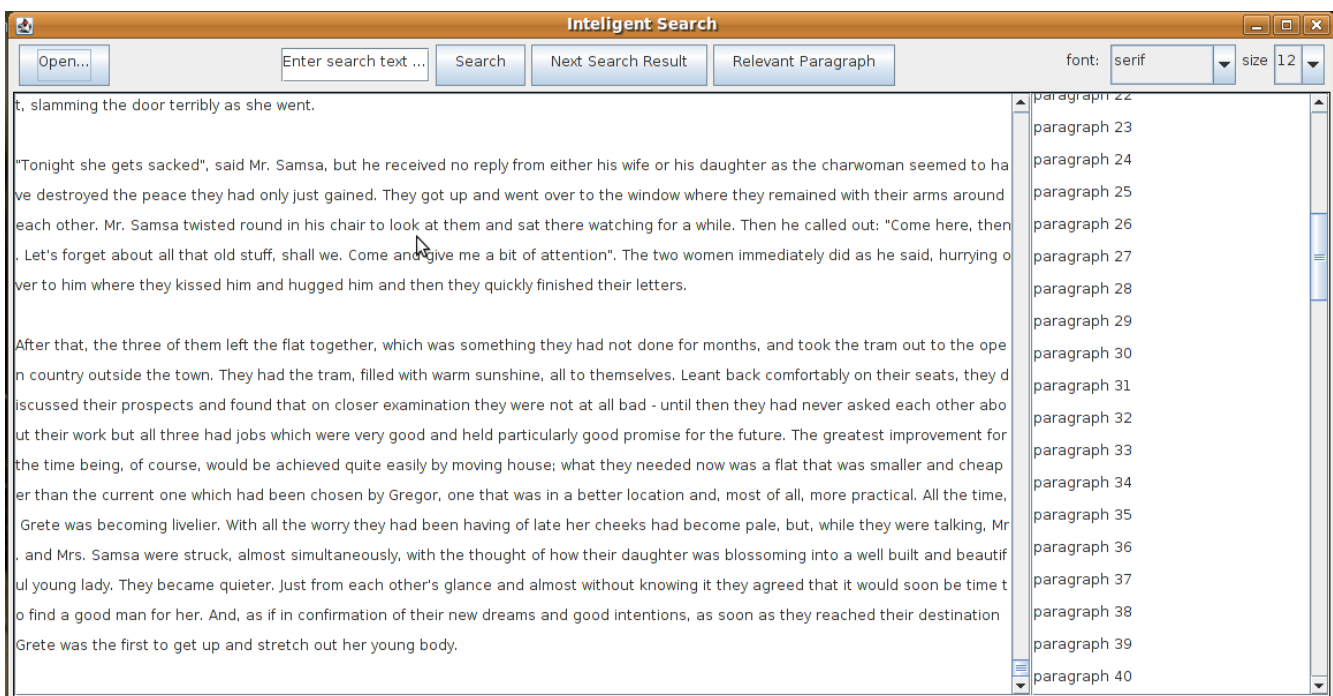
### “Програма за четене на документи, която предоставя интерфейс за търсене на релевантни параграфи”

изготвил : Мартин Обрешков

#### 1. Описание

Проекта реализира програма която е предназначена за четене на различни текстови файлове. Предлага интерфейс за търсене на релевантни параграфи в текста.

#### 2. User interface



Избирате файл, който да отворите ( за момента се поддържат само .txt, .doc формат). След като се зареди текста може да се ориентирате по бутоните.

#### 3. Developers guide

Програмата е написана на Java, използваните допълнителни библиотеки са за отварянето на различните типове файли.

file\_parser – парсване на файлове

scoring – оценяване на документите

tokenizer – разделянето на текста на параграфи и отделни думи

query – отговаря на заявки от клиента

#### **4. Todo**

1. Има нужда да се обогати графичния интерфейс.
2. Да се обработват повече типове файли.