

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ
Кафедра программного обеспечения информационных технологий

Факультет НиДО
Специальность ПОИТ

Контрольная работа № 1
по дисциплине «Информационное обеспечение финансовых структур»

Выполнил студент: Михаленя А.Н.
группа 801021
Зачетная книжка № 801021-23

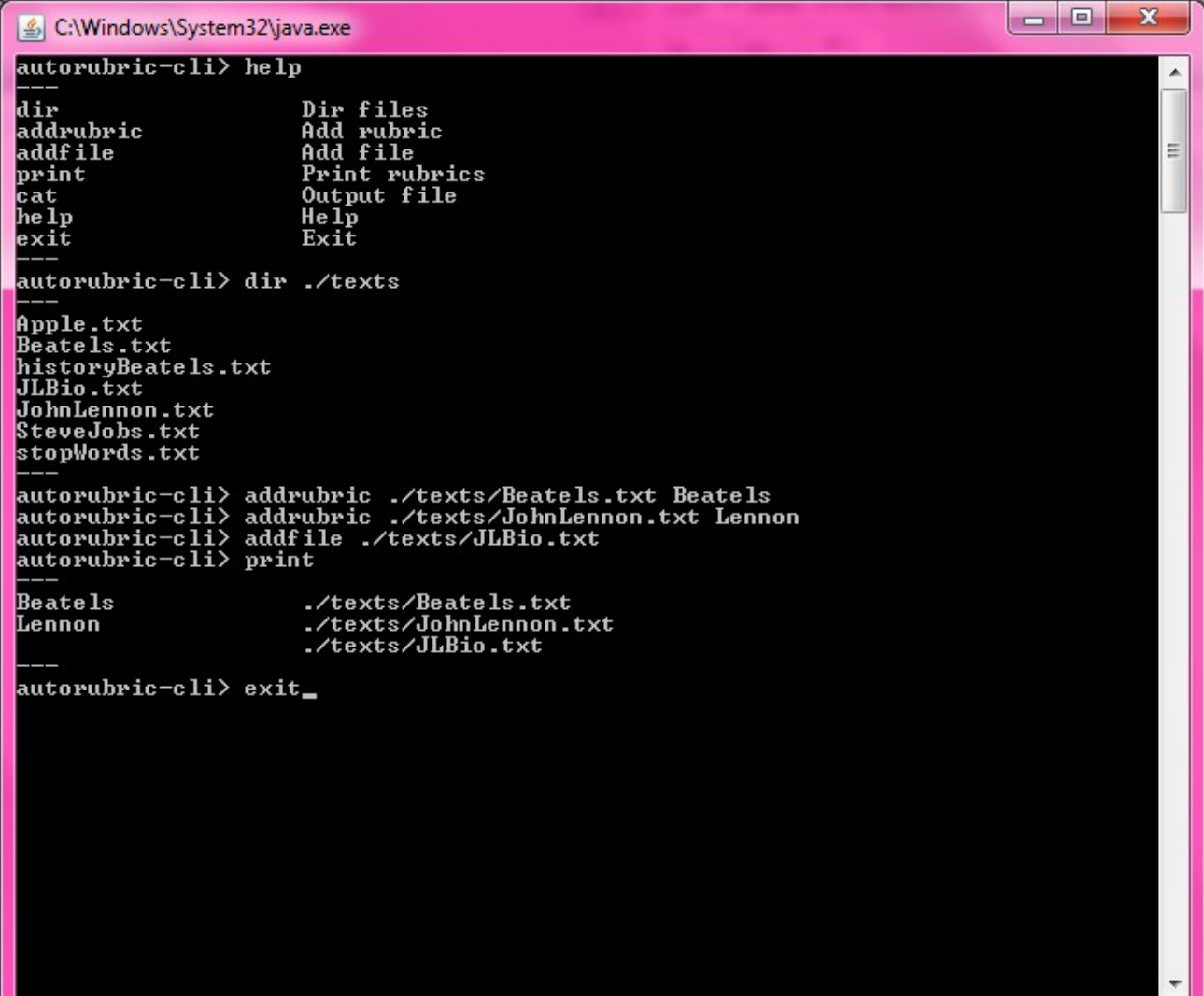
Минск 2013

АВТОМАТИЧЕСКАЯ РУБРИКАЦИЯ ТЕКСТОВОЙ ИНФОРМАЦИИ ПО ОБРАЗЦУ

Цель работы: научиться классифицировать по темам текстовую информацию, используя документы-образцы.

РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ

Для реализации алгоритма автоматической рубрикации текста было разработано CLI приложение. Вывести список команд на экран можно с помощью команды help. При запуске программы инициализируется список стоп слов, которые в дальнейшем учитываются при обработке текстов по законам Зипфа. Добавление рубрики происходит по команде addrubric (указываем последовательно файл и имя рубрики). Добавлять файлы требующие автоматической рубрикации можно с помощью команды addfiles. Рубрикация происходит сразу при добавлении. Вывод результатов – команда print. В случае если не будет указан какой либо параметр или команда будет введена не верно - система укажет на ошибки. Выход из программы – exit. Результат работы программы продемонстрирован на рисунке.



```
C:\Windows\System32\java.exe
autorubric-cli> help
---
dir                Dir files
addrubric          Add rubric
addfile            Add file
print              Print rubrics
cat                Output file
help               Help
exit               Exit
---
autorubric-cli> dir ./texts
---
Apple.txt
Beatels.txt
historyBeatels.txt
JLBio.txt
JohnLennon.txt
SteveJobs.txt
stopWords.txt
---
autorubric-cli> addrubric ./texts/Beatels.txt Beatels
autorubric-cli> addrubric ./texts/JohnLennon.txt Lennon
autorubric-cli> addfile ./texts/JLBio.txt
autorubric-cli> print
---
Beatels            ./texts/Beatels.txt
Lennon             ./texts/JohnLennon.txt
JLBio              ./texts/JLBio.txt
---
autorubric-cli> exit_
```

Рисунок 1.