Казанский (Приволжский) Федеральный Университет

На правах рукописи

УДК xxx.xxx

Тощев Александр Сергеевич

Разработка эффективного подхода обработки производственных задач прикладного характера в области обслуживания программного обеспечения и информационной инфраструктуры предприятия на основе стохастического поиска, вероятностно-логических рассуждений и машинного обучения

Специальность 05.13.01 —
«Системный анализ, управление и обработка информации (по отраслям)»

Диссертация на соискание учёной степени
Кандидат технических наук

Научный руководитель:
уч. степень, уч. звание
Елизаров А.М.

Казань — 2015

Оглавление

Введение

В настоящее время в области IT набрало большую популярность системы удаленной поддержки информационной инфраструктуры, так называемый «Аутсорсинг». Ввиду развития рынка компаниям становится невыгодно держать свой штат службы поддержки, и они отдают свою инфраструктуру сторонней компании. Ввиду возросшей интенсивности данного бизнеса возникла потребность автоматизации работы. В данном контексте рассматривается автоматизация обработки инцидентов, начиная с разбора инцидентов на естественном языке и заканчивая поиском решения и применением решения. Главными требованиями к системе являются

1. Обработка естественного языка
2. Возможность обучения
3. Общение с специалистом
4. Проведение логических рассуждений: аналогия, дедукция, индукция
5. Умения абстрагировать решение и экстраполировать его на другие решения

На данный момент многие компании ведут разработку подобных систем. Примером такой системы является набирающая популярность система IBM Watson [1]. Подобный класс система также называют вопросно-ответными системами. Другим примером является система Wolfram Alpha [2]. В данной работе был сделан акцент на попытку создания мыслящий системы на основе модели мышления Марвина Мински [3].

Целью данной работы является создание архитектуры и реализация базового прототипа программного комплекса обеспечивающего разбор и формализацию входного запроса пользователя и поиск решения данной проблемы.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. Исследовать целевую область
2. Вычислить возможность автоматизации целевой области
3. Исследовать модель мышления Марвина Мински
4. Адаптировать модель для прикладной реализации
5. Создать архитектуру приложения на основе модели

6. Реализовать прототип на основе архитектуры

Основные положения, выносимые на защиту:

1. Возможность автоматизации области предоставления удаленной поддержки информационной инфраструктуры
2. Прикладное применение модели мышления Марвина Мински для решения задачи автоматизации
3. Возможность программной реализации модели мышления Марвина Мински
4. Экстраполяция программной системы для других областей

Научная новизна:

1. Впервые была представлена реализация модели мышления Мински на практике
2. Была представлена новая модель данных для модели мышления
3. Было выполнено оригинальное исследование модели мышления …

Научная и практическая значимость …

Степень достоверности полученных результатов обеспечивается результатами выполнения тестов на контрольных примерах. Результаты находятся в соответствии с результатами, полученными другими авторами и экспертными системами

Апробация работы. Основные результаты работы докладывались на:

– RCDL-2014
– AINL-2013
– WCIT-2012

Личный вклад. Автор принимал активное участие в разработке архитектуры приложения, реализации прототипа, проработки теории, тестировании.

Публикации. Основные результаты по теме диссертации изложены в XX печатных изданиях [4–8], X из которых изданы в журналах, рекомендованных ВАК [4–6], XX — в тезисах докладов [7, 8].

Объем и структура работы. Диссертация состоит из введения, четырех глав, заключения и двух приложений. Полный объем диссертации составляет XXX страница с XX рисунками и XX таблицами. Список литературы содержит XXX наименований.

Глава 1

ПОСТАНОВКА ЗАДАЧИ ПОЛУЧЕНИЯ, АНАЛИЗА И ОБРАБОТКИ ЭКСПЕРТНОЙ ИНФОРМАЦИИ

1.1. Возникновение области

В настоящее время в области IT набрало большую популярность системы удаленной поддержки информационной инфраструктуры, так называемый «Аутсорсинг». Ввиду развития рынка компаниям становится невыгодно держать свой штат службы поддержки, и они отдают свою инфраструктуру сторонней компании. Большинство проблем, которые решает удаленная служба поддержки носят весьма тривиальный характер :

- Установить приложение
- Переустановить приложение
- Решить проблему с доступом к тому или иному ресурсу

Данные проблемы решают специалисты технической поддержки. Обычно техническая поддержка делится на несколько линий:

1. Первая линия. Решение уже известных, задокументированных проблем, работа напрямую с пользователем
2. Вторая линия. Решение ранее неизвестных проблем
3. Третья линия. Решение сложных и нетривиальных проблем
4. Четвертая линия. Решение архитектурных проблем инфраструктуры

Каждая линия поддержки представлена специалистами. В среднем команда, обслуживающая одного заказчика насчитывает 60 человек. Процентное соотношение специалистов разных линий поддержки отображено на Диаграмме 1.1

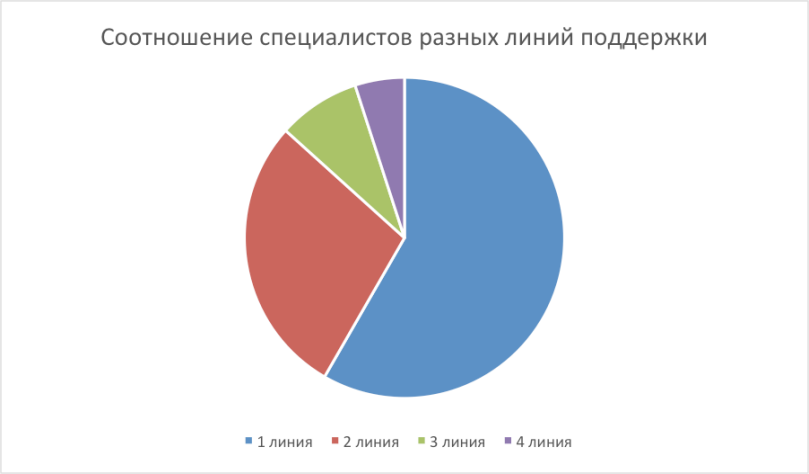

Рисунок 1.1: Диаграмма состава команд

Работа специалиста 1 линии поддержки состоит из множества рутинных и простых задач. На Диаграмме 1.2 показано соотношение разных типов проблем, встречающихся во время работы поддержки

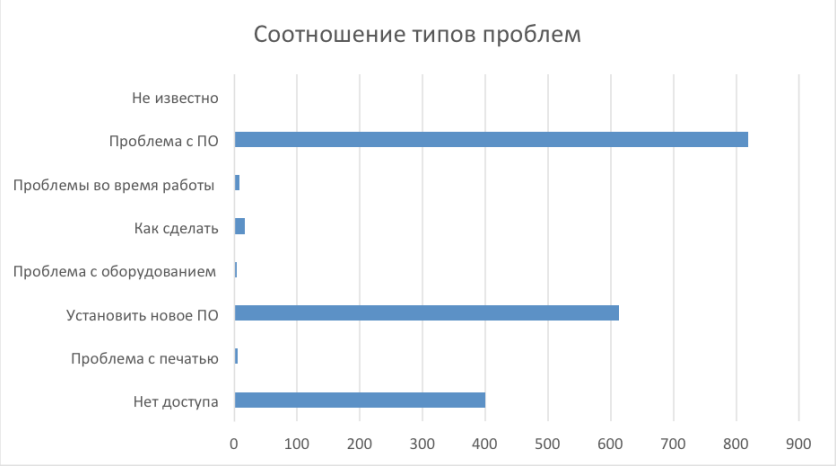

Рисунок 1.2: Диаграмма соотношений типов проблем

Решение части задач может быть автоматизировано, а специалисты получат дополнительное время на решение более интересных задач. Проблема заключается в автоматизации решения рутинных задач в области удаленной поддержки инфраструктуры.

1.2. Прогноз развития области

Основной тенденцией в развитии области удаленной поддержки инфраструктуры является попытки удешевить и улучшить стоимость предоставления услуг.

Компании, работающие на этом рынке вкладывают большие деньги в автоматизацию. Кроме того современное развитие науки и техники, а точнее вычислительных мощностей позволяет автоматизацию даже самых наукоемких процессов.

Дальнейшим развитием области является замена человеческих специалистов на автоматические

Таблица 1.1: Категории инцидентов

Категория	Описание
Проблема с ПО	Проблема при запуске ПО на компьютере. Решается переустановкой
Проблемы во время работы	Проблема с функционированием программного обеспечения
Как сделать	Запрос на инструкцию по работе с тем или иным компонентом рабочей станции
Проблема с оборудованием	Неполадки на уровне оборудования
Установить новое ПО	Требование установки нового программного обеспечения
Проблема с печатью	Установка принтера в систему
Нет доступа	Нет доступа к общим ресурсам

системы. Многие ведущие компании ведут разработки в этом направлении. Например, компания HP. Данная компания имеет свою системы по регистрации подобных инцидентов и сейчас ведется работа над автоматизацией системы.

1.3. Методологии, используемые в области IT аутсорсинга: ITIL и ITSM

В области IT аутсорсинга есть несколько готовых стандартов ведения работ. Одним из таких стандартов является библиотека ITIL. Данные стандарт описывает лучшие практики организации работ в области IT аутсорсинга. Используемый в библиотеки подход соответствует стандартам ISO 9000 (ГОСТ Р ИСО 9000) . Наличие стандартов в области диктует унифицированность постановки проблем, а также унифицированность алгоритмов решения. Такие предпосылки говорят о возможности частично или в некоторых случаях полной автоматизации решения проблем.

1.4. Постановка задачи

Задачами данного исследования являются:

- Изучение возможности автоматизации области удаленной поддержки инфраструктуры путем анализа области
- Выработка критериев и сравнительный анализ существующих решений в области
- Создание программного комплекса (фреймворка) для автоматизации поддержки удаленной инфраструктуры

– Подсчет статистических результатов работы комплекса

Глава 2

МЕТОДЫ И КОМПЛЕКСЫ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

2.1. Обработка Эталонных Текстов

В данном разделе проводится обзор обработчиков естественного языка. За основу были взяты инциденты из выгрузки систем поддержки ОАО ”ICL КПО-ВС”.
Ввиду специфики области основным языком был выбран английский язык. Был сформирован список из типичных эталонных фраз, на которых тестировались обработчики естественного языка. Фразы были выявлены путем анализа существующих отчетов об инцидентах. Примерами инцидентов являются следующие инциденты:
Инцидент 1 *User had received wrong application. User has ordered Wordfinder Business Economical for her service tag 7Q4TC3J, there is completed order in LOT with number ITCOORD-18125. However she received wrong version, she received Wordfinder Tehcnical instead of Business Economical. Please assist.*
Инцидент 2 *Laptop – user has almost full C: but when he looks in the properties of the files and folders on C: they are only 40GB and he has a 55GB drive.*
Инцидент 3 *User cannot find Produkt Manageron start menu. Please reinstall.*
Инцидент 4 *User needs to have pdf 995 re-installed please.*

Во время анализа были использованы следующие обработчики естественного языка:

1. Open NLP [9]
2. RelEx [10]
3. StanfordParser [11]

Результат работы вычислялся при помощи метрик, представленных в Таблице2.1.
Результаты приведены на сводной диаграмме Рисунок 2.1

Таблица 2.1: Таблица метрик

Метрика	Описание	Формула
Аккуратность	Понимание текста обработчиком	$Ac = \frac{1-x}{y}$ где x- количество нераспознанных слов, y количество распознанных
Успешно обработанные	Успешно обработанные инциденты	$P = \frac{x}{100}$ где x успешно обработанные
Не успешно обработанные	Неуспешно обработанные инциденты	$N = \frac{y}{100}$ где y неуспешные инциденты
Результативность	Общая результативность обработчика	$R = \frac{P}{N}$
Общий бал	Общая оценка обработчика	$T = Ac + R$

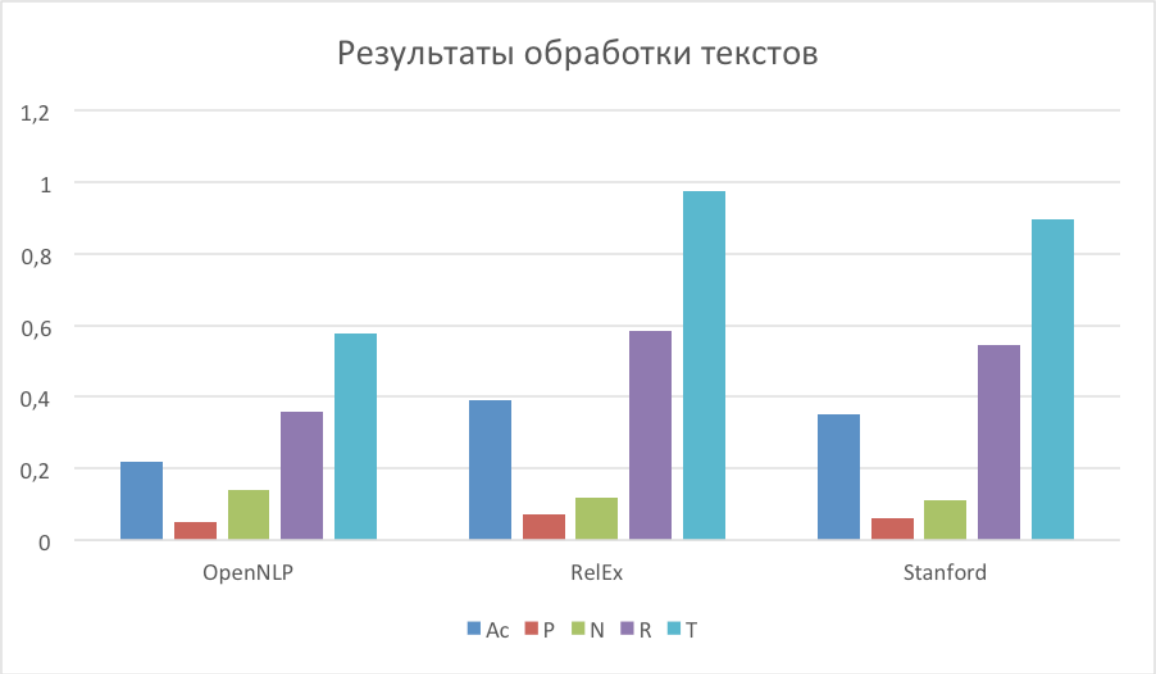

Рисунок 2.1: Результаты обработки текстов

Из диаграммы видно, что наилучшее результаты показывает обработчик RelEx [10]. После анализа необработанных инцидентов было выявлено нескольких проблем у всех обработчиков:

1. Невозможности корректировки простых грамматических ошибок, связанных с пропущенными пробелами или неверным форматированием. Ошибки первого типа.

2. Ошибки неверной интерпретации слов в предложении. Например, слово please интерпретировалось как глагол, хотя является по смыслу «формой вежливости». Ошибки второго типа.

2.2. Обработка текстов с ошибками

По результатам прошлого раздела было решено выбрать в качестве обработчика естественного языка RelEx, но были выявлены некоторые проблемы. Было принято решение исправить данные проблемы при помощи предварительной обработки текста. Предварительная обработка текста была разбита на несколько фаз:

1. Комплексная корректировка ошибок

2. Обработка при помощи внутренней базы знаний

Для того, чтобы избавится от орфографический, синтаксических ошибок используется составной корректировщик. Данный компонент имеет модульную структуру и применяет корректировку последовательно.

Для данного компонента были написаны модули корректировки:

– Google API

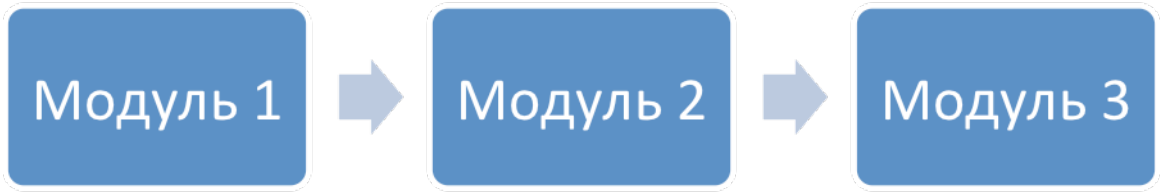

Рисунок 2.2: Архитектура предварительной обработки текста

– After The Deadline

Таким образом удалось исправить большинство ошибок, связанный с синтаксисом, грамматикой, орфографии. Также удалось исправить ошибки неверного написания: лишних пробелов, пропущенных запятых, пропущенных точек. По-прежнему остается проблема обработки неверной интерпретации слов в тексте.

Для корректировка ошибок второго типа был использовано инъекция в работу обработчика RelEx. В ввиду OpenSource природы проекта, модульности был подменен модуль извлечения и обработки слов в предложения. Стандартный процесс обработки был разбит на «предобработку» и «обработку». Стадия «обработки» включала в себя алгоритм работы такой же как был до этого в модули, на стадии «предобработки» управление передается модулю основному приложению, который проверяет данное слово на предмет его вхождения во внутреннюю Базу Знаний и если таковое имеется, то приложение передает соответствующие корректировки в модуль

2.3. Сравнение средств обработки русского и английского языка

Средства обработки естественного языка принято относить в большому классу средств NLP – Natural Language Processing. Для английского языка существует множество открытых средств обработки естественного языка, для русского языка найти их гораздо сложнее. Рассмотрим архитектуру средств обработки естественного языка на примере OpenCog RelEx

Глава 3

Вёрстка таблиц

3.1. Таблица обыкновенная

Так размещается таблица:

Таблица 3.1: Название таблицы

Месяц	T_{min}, К	T_{max}, К	$(T_{max} - T_{min})$, К
Декабрь	253.575	257.778	4.203
Январь	262.431	263.214	0.783
Февраль	261.184	260.381	-0.803

3.2. Параграф - два

Некоторый текст.

3.3. Параграф с подпараграфами

3.3.1. Подпараграф - один

Некоторый текст.

3.3.2. Подпараграф - два

Некоторый текст.

Заключение

Основные результаты работы заключаются в следующем.

1. На основе анализа …
2. Численные исследования показали, что …
3. Математическое моделирование показало …
4. Для выполнения поставленных задач был создан …

И какая-нибудь заключающая фраза.

Список литературы

1. *Wikipedia*. IBM Watson. — web. — 2014. https://ru.wikipedia.org/wiki/IBM_Watson.
2. *Wolfram*. Wolfram Alpha. — web. — 2014. — 00. http://www.wolframalpha.com/.
3. *Minsky Marvin*. The Emotion Machine. — Simon & Schuster, 2006.
4. *Сычёв М. С.* История Астраханвского казачьего войска: учебное пособие. — Астрахань: Волга, 2009. — 231 с.
5. *Соколов А. Н., Сердобинцев К. С.* Гражданское общество: проблемы формирования и развития (философский и юридический аспекты): монография / Под ред. Х. Шутзе. — Астрахань: Издательство, 2009. — 218 с.
6. *Гайдаенко Т. А.* Маркетинговое управление: принципы управленческих решений и российская практика. — 3-е изд, перераб. и доп. изд. — М.: Эксмо: МИРБИС, 2008. — 508 с.
7. *Лермонтов Михаил Юрьевич*. Собрание сочинений: в 4 т. — М.: Терра-Кн. клуб, 2009. — 4 т.
8. Управление бизнесом: сборник статей. — Нижний новгород: Изд-во Нижегородского университета, 2009. — 243 с.
9. *Foundation Apache Software*. Apache OpenNLP. — web. — 2012. — 04. https://opennlp.apache.org/.
10. *Goetzel Ben*. OpenCog RelEx. — web. — 2012. — 04. http://wiki.opencog.org/w/RelEx.
11. *Маннинг С. Д., Прабхакар Р.* Введение в обработку информации / Ed. by Х. Шутзе. — Кембридж: Издательство Университета Кембридж, 2009. — 581 pp.

Список рисунков

Список таблиц

Приложение А

Название первого приложения

Некоторый текст.

Приложение В

Очень длинное название второго приложения, в котором продемонстрирована работа с длинными таблицами

В.1. Подраздел приложения

Вот размещается длинная таблица:

Параметр	Умолч.	Тип	Описание
&INP			
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора

продолжение следует

(продолжение)

Параметр	Умолч.	Тип	Описание
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$) 1: генерация белого шума 2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
&SURFPAR			
kick	1	int	0: инициализация без шума ($p_s = const$)

продолжение следует

(продолжение)			
Параметр	Умолч.	Тип	Описание
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс
kick	1	int	0: инициализация без шума ($p_s = const$)
			1: генерация белого шума
			2: генерация белого шума симметрично относительно экватора
mars	0	int	1: инициализация модели для планеты Марс

В.2. Ещё один подраздел приложения

Нужно больше подразделов приложения!

B.3. Очередной подраздел приложения

Нужно больше подразделов приложения!

B.4. И ещё один подраздел приложения

Нужно больше подразделов приложения!