

XII международная конференция  
CEE-SECR / РАЗРАБОТКА ПО

28 - 29 октября, Москва

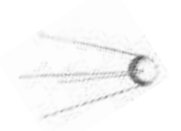


# Titanic.csv — Как заметить макушку айсберга в океане багов?

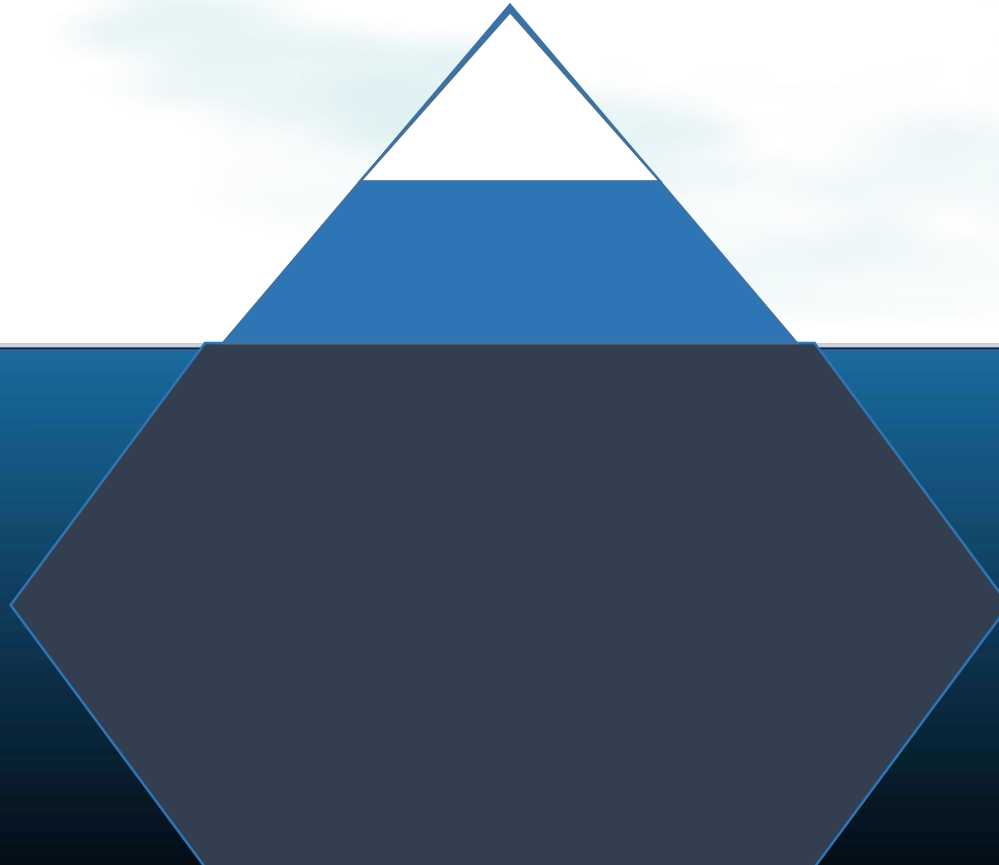
Панков Иван



**RANEPА**  
THE RUSSIAN PRESIDENTIAL ACADEMY  
OF NATIONAL ECONOMY  
AND PUBLIC ADMINISTRATION



# Как заметить макушку айсберга в океане багов?



Знать, что такое айсберг



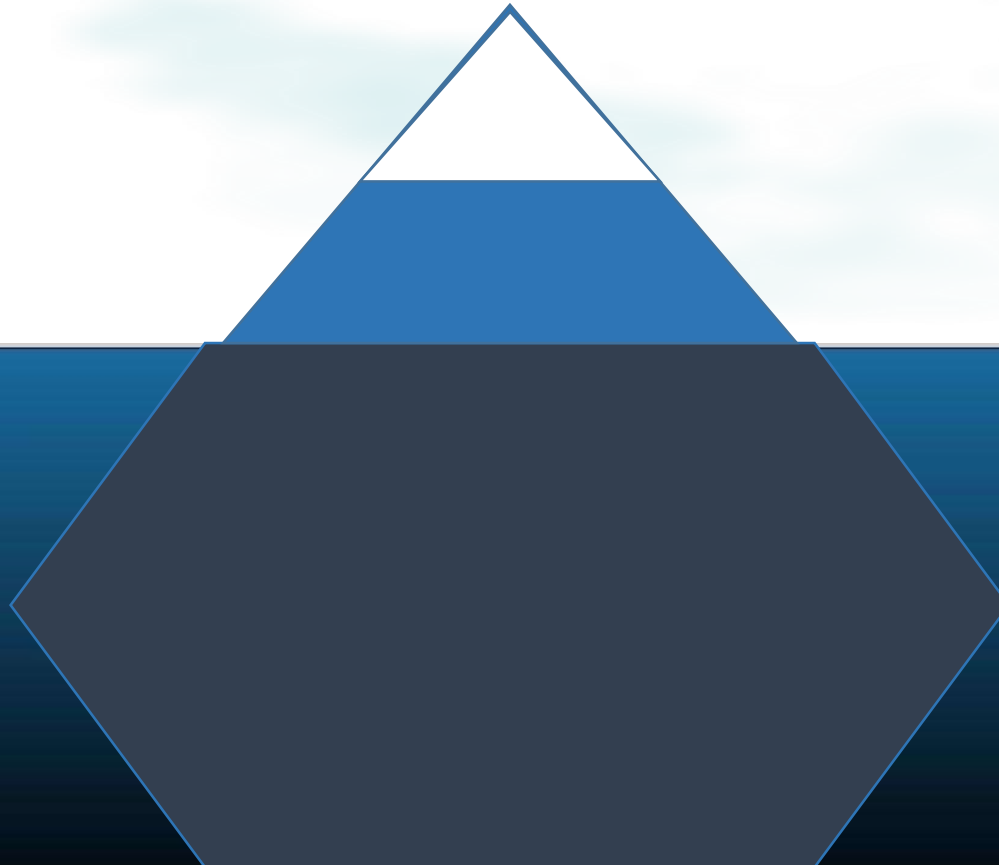
Подключать профессионалов



**Настраивать свои инструменты**



# Что такое Баг с разных точек зрения



Что видно в метаданных



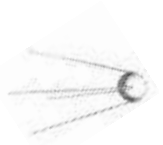
Что видно в описании дефекта



**Что известно из отзывов пользователей**

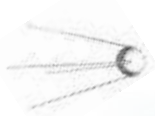


**Тайное знание разработчиков**



# Взгляд на Баг с точки зрения Text Mining





# Взгляд на Баг с точки зрения Text Mining



Описание бага



Связанные дефекты и документы



Инциденты у клиентов



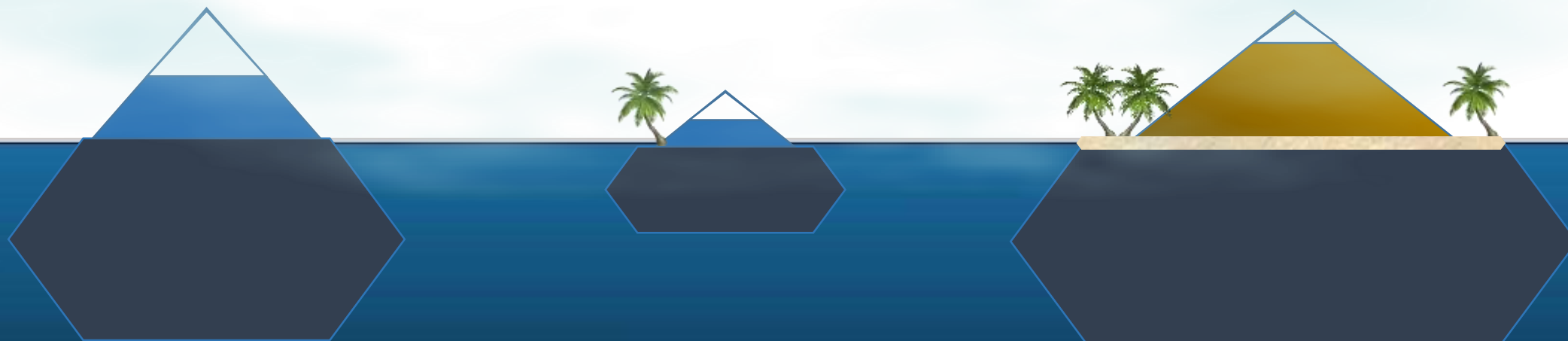
Отчеты о падениях, логи



Программный код

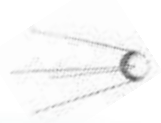


# С чего начать приёмку?

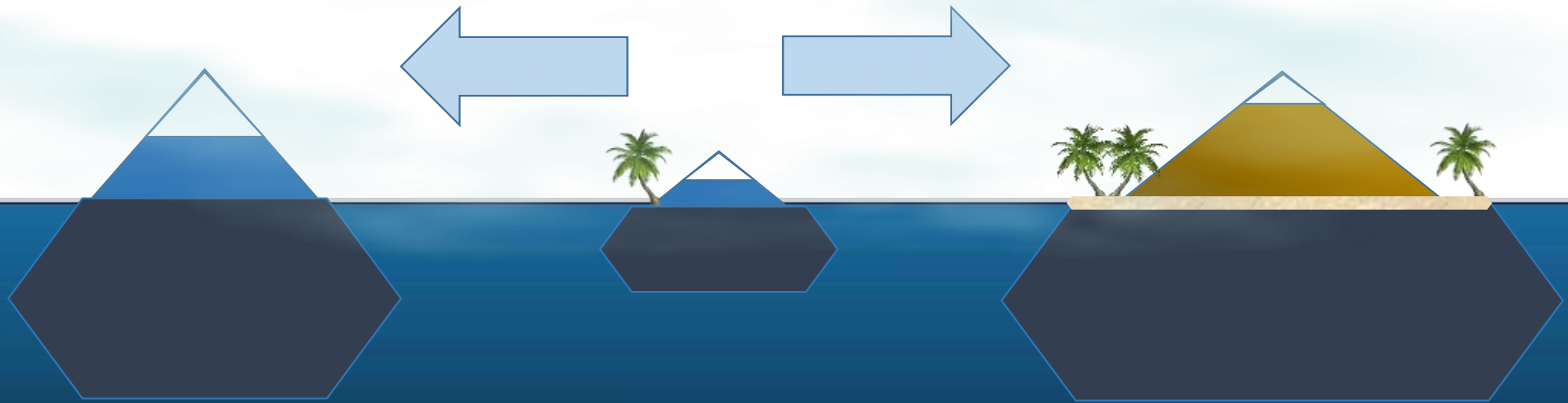


**В первую очередь обращают внимание на уже известные характеристики – размер, скорость, категория риска.**

**Но что, если они обманчивы?**



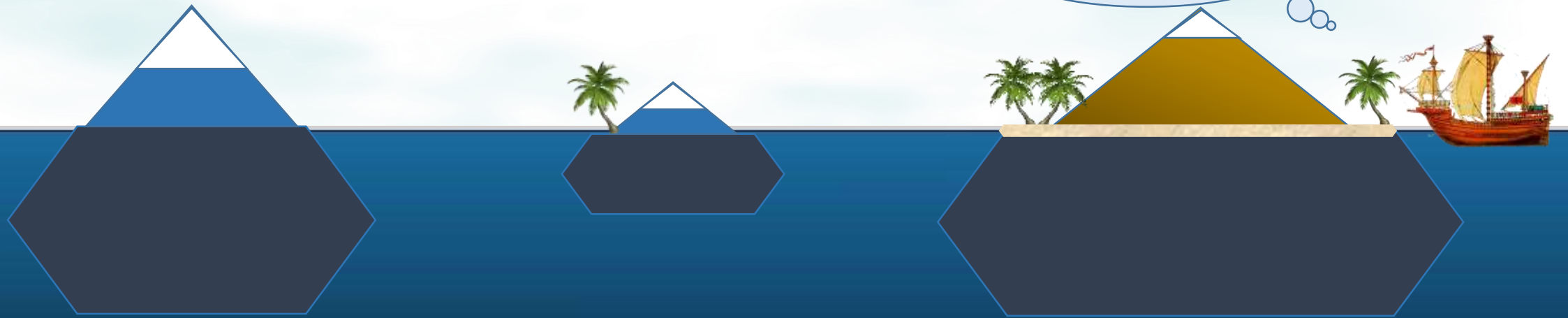
# Как сопоставлять:



**Сопоставление с другими примерами может  
выявить закономерности**

# Как сопоставлять: Скоринговые модели

...похоже на Индию

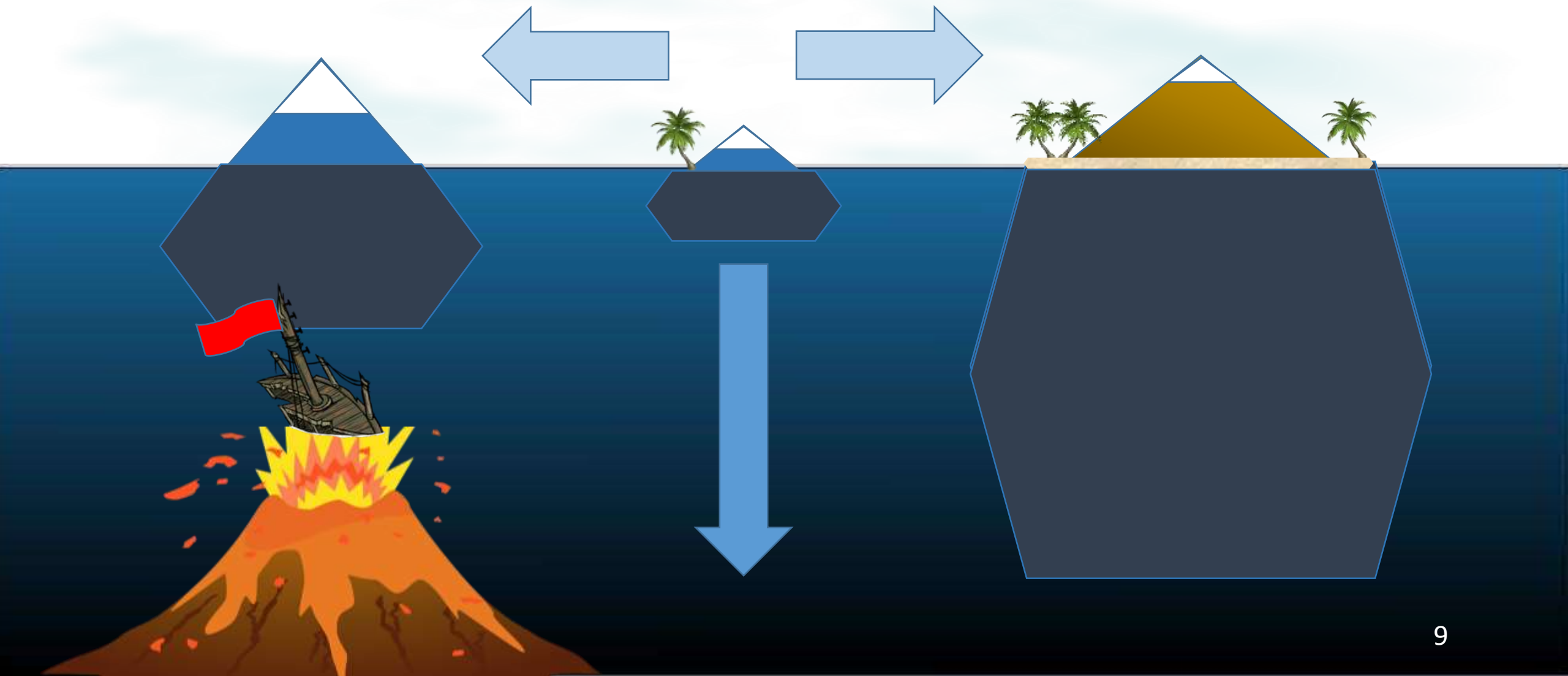


**Человек ошибается!**

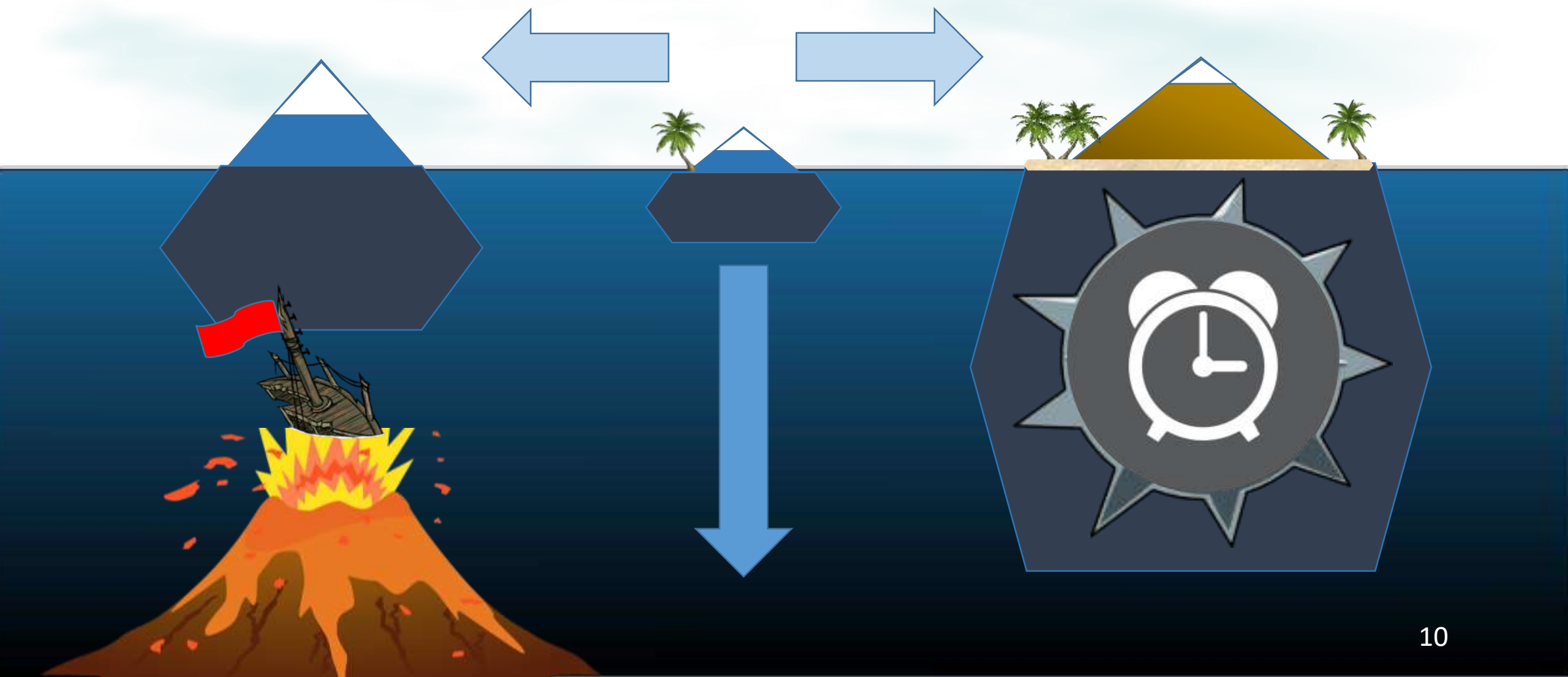
**Впрочем, как и машины**



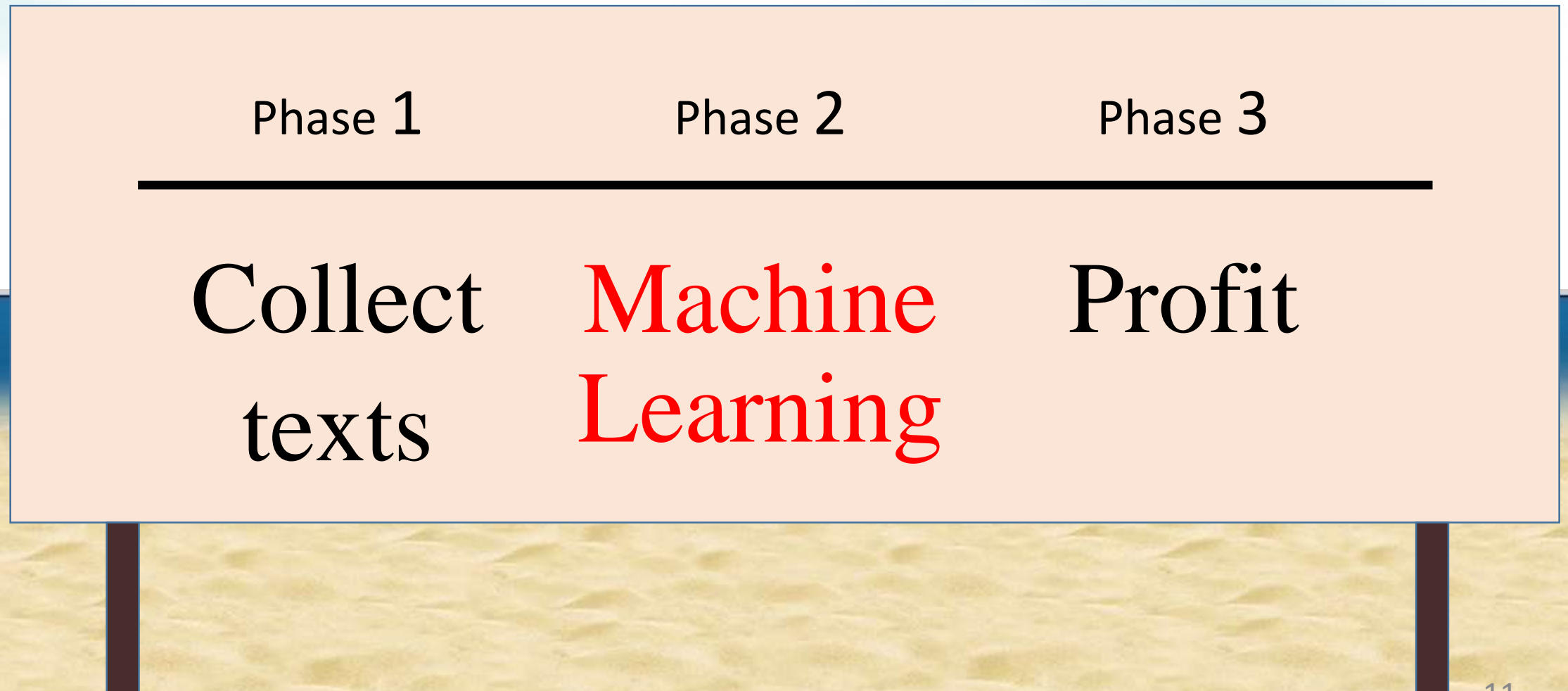
# Как сопоставлять: Подход исторический



# Как сопоставлять: Подход исторический



# Как использовать сакральное знание на практике: Структура



# Как использовать сакральное знание на практике. Phase 1: Сбор текстов

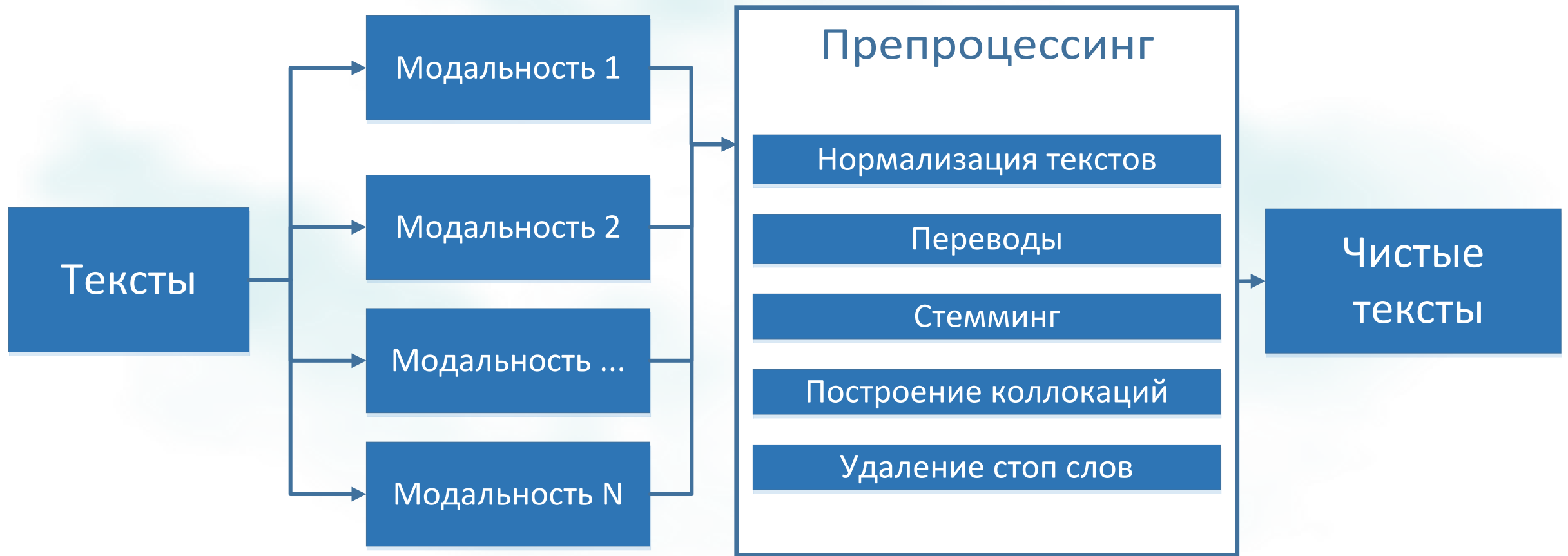
- Сам сбор текстов – заголовки, описания, национальный корпус и корпоративный
- Препроцессинг – переводы, нормальная форма, цензура, разбиение на смысловые части
- Векторизация – статистика, нормализация, добавление оптимизируемых параметров



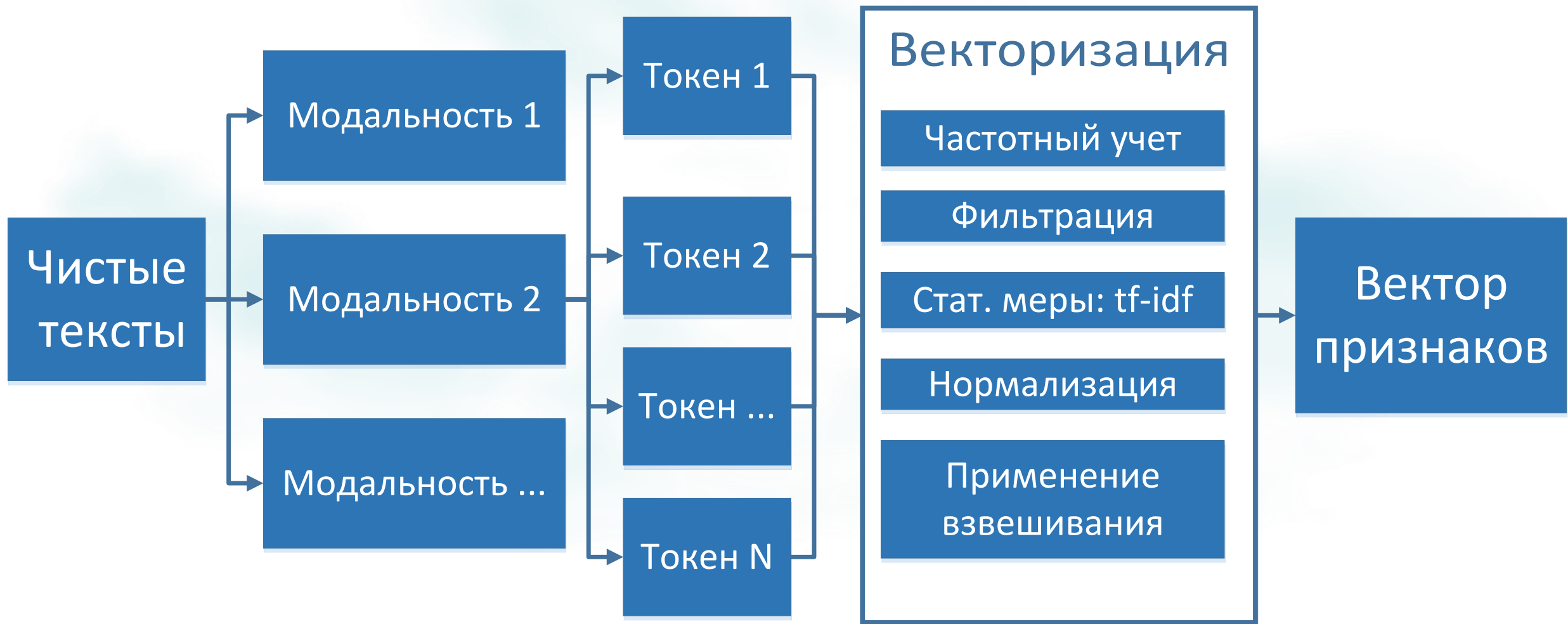
# Как использовать сакральное знание на практике. Phase 1: Сбор текстов



# Как использовать сакральное знание на практике. Phase 1: Препроцессинг



# Как использовать сакральное знание на практике. Phase 1: Векторизация



# Как использовать сакральное знание на практике. Phase 2: **Machine Learning**

Наши тексты теперь являются просто набором цифр, значит применяем стандартный арсенал Машинного обучения.

Возможные шаги:

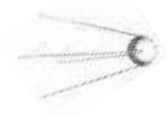
- Линейная регрессия
- Классификация методом ближайших соседей
- Наивный байесовский классификатор
- SVM
- etc.





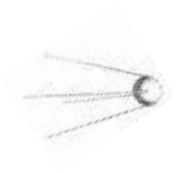
# Как использовать сакральное знание на практике. Phase 3: Profit

- Учет при очередности выполнения заданий по приёмочному тестированию
- Внедрение в системы поддержки принятия решений
- Поиск скрытых угроз по запросу или событию



# Направления развития

- Учет семантики предложения
- Автоматический поиск синонимии несловарных терминов
- Увеличение числа модальностей
- Учет времени и эволюционных процессов



# Спасибо за внимание

## Контакты

Email: [pankov.i.d@gmail.com](mailto:pankov.i.d@gmail.com)

Facebook: [www.facebook.com/ivan.pankov.7](https://www.facebook.com/ivan.pankov.7)

