Диссертация магистра по направлению
«Проектирование и технология производства ЭС»

КОМПЛЕКС ЗАЩИТЫ СИСТЕМЫ ЭЛЕКТРОННОЙ ПОЧТЫ ОТ НЕЗАПРАШИВАЕМОЙ РАССЫЛКИ

Ларютин А.В.
ИУ4, 2006 г.

Научный руководитель: доцент, к. т. н. Власов А. И.

Постановка задачи

Цель работы:

Разработка многопользовательской многоагентной обучаемой пользователями автоматизированной системы адаптивной фильтрации, обеспечивающей выявление и блокировку незапрашиваемых электронных почтовых сообщений.

Решаемые задачи:

- Исследование и анализ существующих методов распространения незапрашиваемой рассылки и функциональных возможностей существующих систем фильтрации.
- Исследование и разработка математических моделей, обеспечивающих адаптивную фильтрацию потоков информации в группах пользователей.
- Исследование и разработка архитектуры адаптивной системы фильтрации, позволяющей пользователям участвовать в ее обучении
- Исследование и выбор аппаратной и программной реализации средств фильтрации и разработка реализации адаптивной системы фильтрации в виде программно-аппаратного комплекса;
- Экспериментальное исследование предложенных моделей, методов и алгоритмов

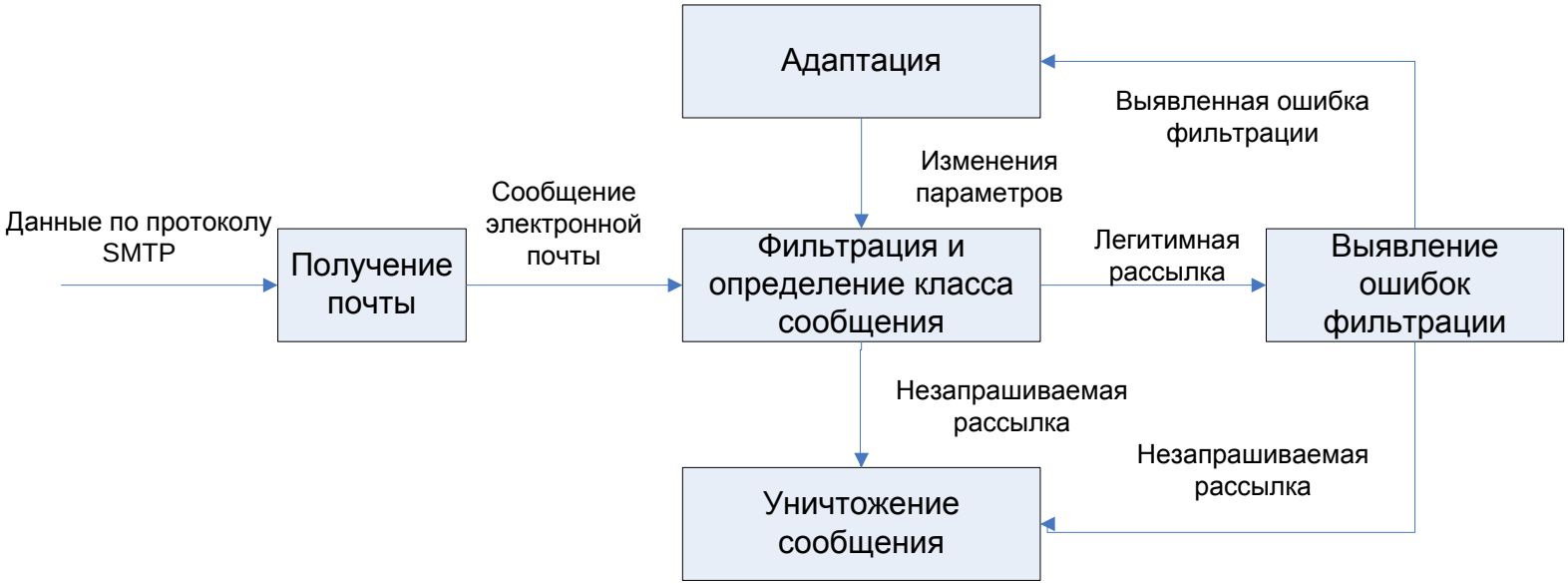
Незапрашиваемая рассылка (НР) – сообщения электронной почты, носящие рекламный или информационный характер, доставляемые адресатам, которые не выразили явного или неявного желания получать данные сообщения, а также выразили нежелание получать их.

Основные методы обхода фильтров НР:

- модификация адреса отправителя
- модификация текста сообщения
- ретрансляция сообщения через промежуточные хосты

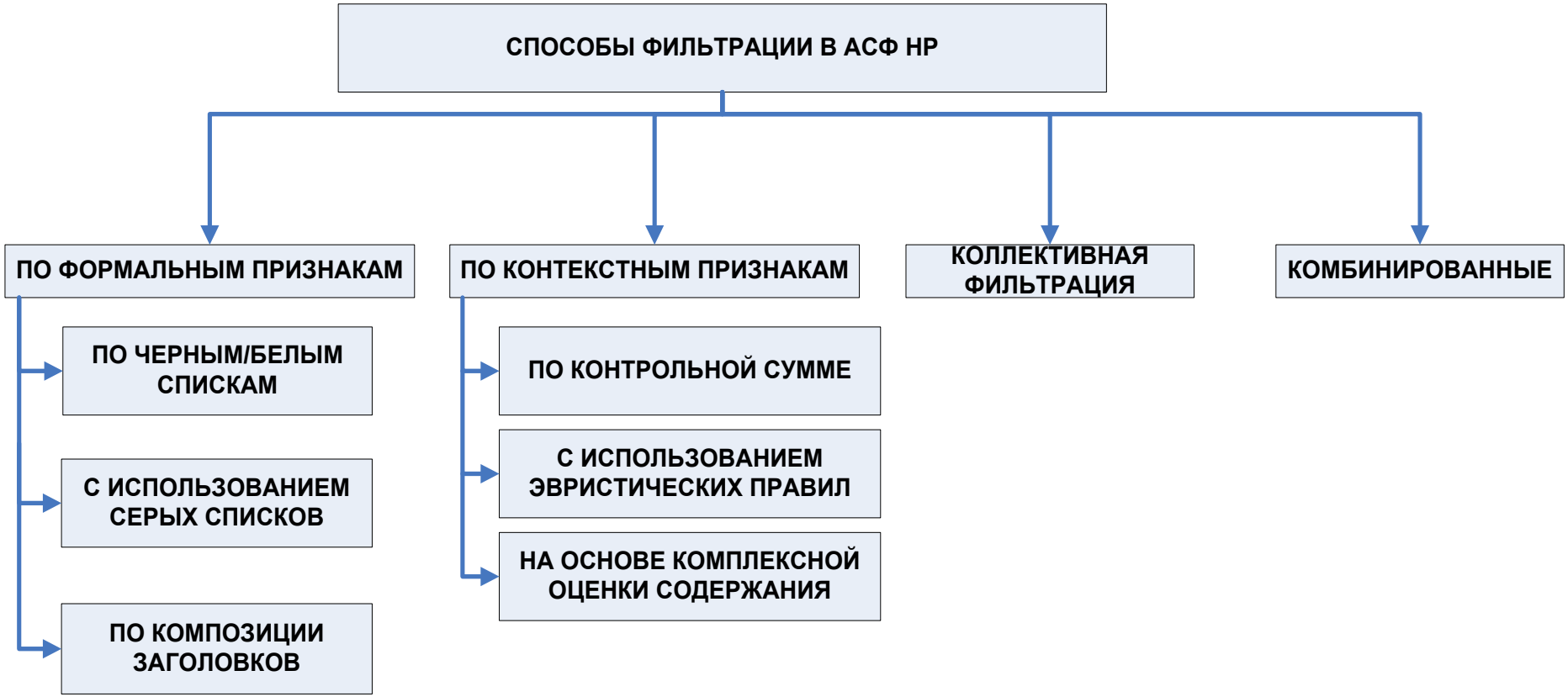
Основные функции систем фильтрации HP:

- получение сообщений электронной почты
- фильтрация и определение класса сообщений
- выявление ошибок фильтрации и адаптация

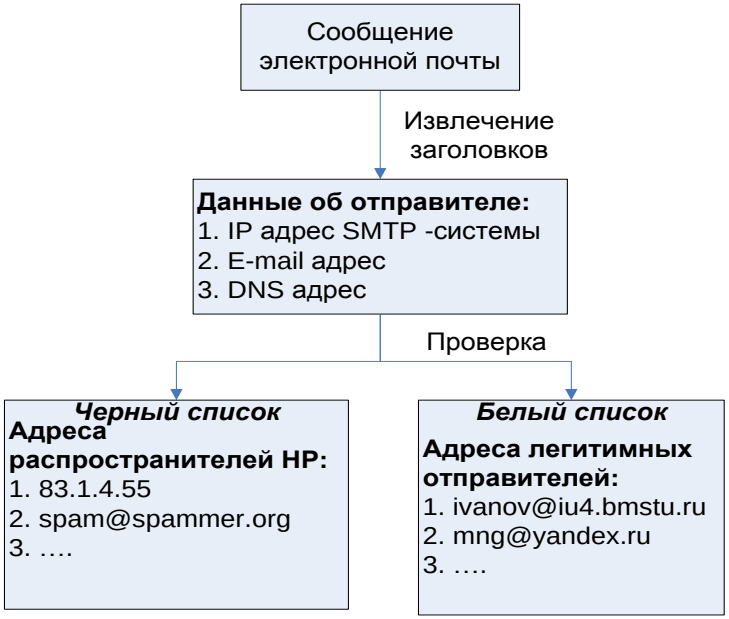


Методы фильтрации сообщений электронной почты:

- по формальным признакам
- по контекстным признакам
- коллективная фильтрация
- комбинированная фильтрация

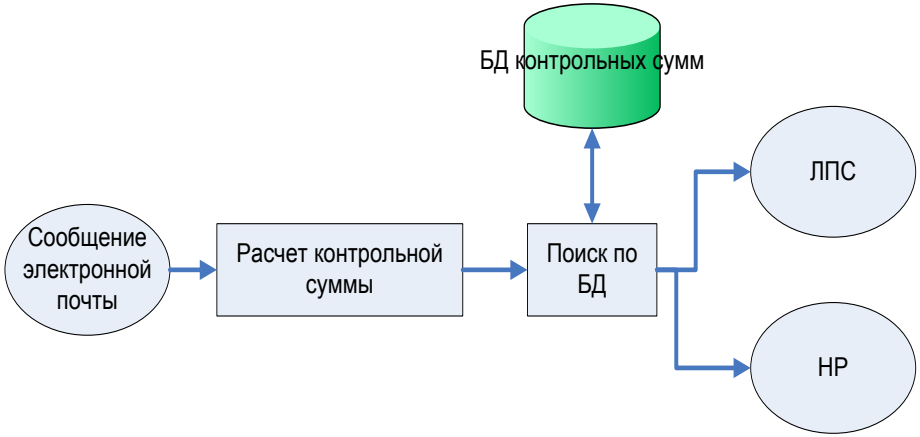


Методы формальной фильтрации



Фильтрация по спискам

- извлечение данных об отправителе
- проверка по наличию в черных/белых списках
- принятие решения

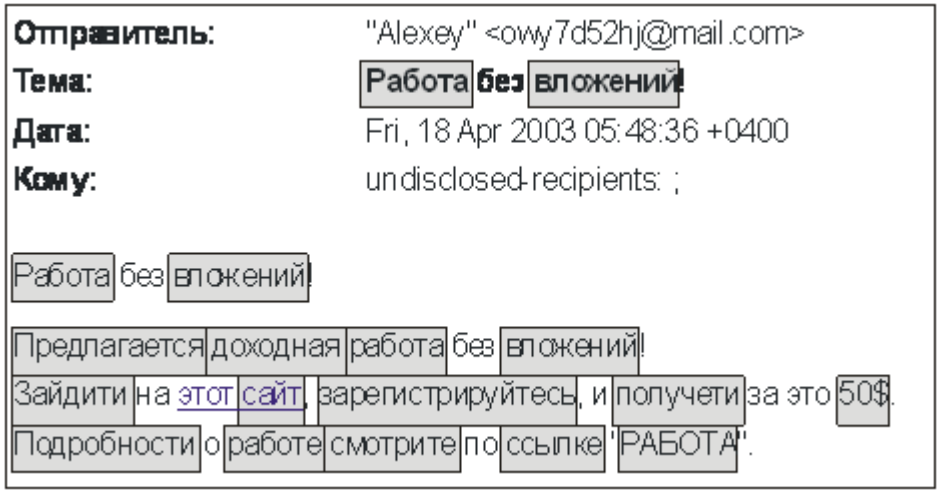


Фильтрация по контрольным суммам

- расчет контрольной суммы
- проверка по БД существующих контрольных сумм HP
- принятие решения

Метод контекстной фильтрации

- формирование списка ключевых слов экспертом
- морфологический анализ текста сообщения
- выделение характерных для HP ключевых слов
- принятие решения на основе заданного порога вероятности



Отправитель: "Alexey" <owy7d52hj@mail.com>
Тема: Работа без вложений
Дата: Fri, 18 Apr 2003 05:48:36 +0400
Кому: undisclosed-recipients: ;

Работа без вложений!

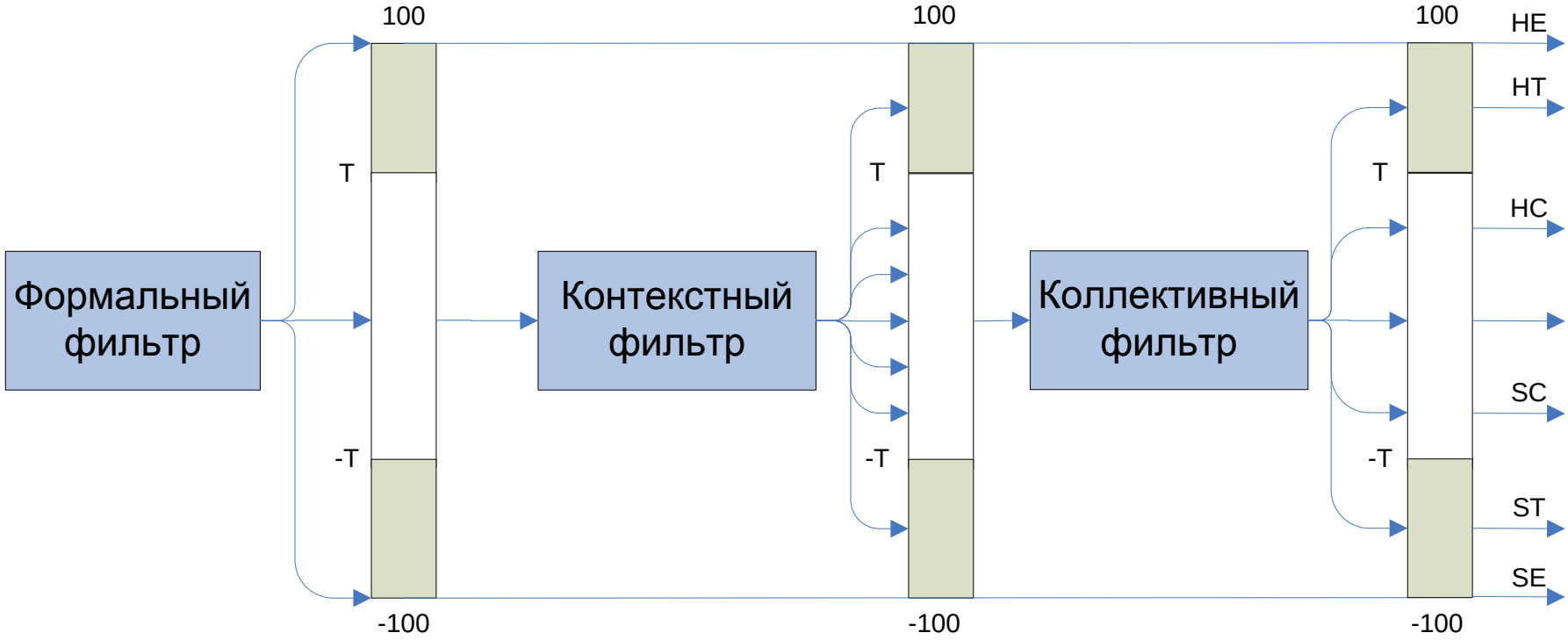
Предлагается доходная работа без вложений!
Зайдити на этот сайт, зарегистрируйтесь, и получети за это 50$.
Подробности о работе смотрите по ссылке "РАБОТА".

Лексическая Единица	*N*	*CHP*	*ЛПС*
Вложения	**3**	**0.9**	**0.2**
Доходный	**1**	**0.9**	**0.2**
Зарегистрироваться	1	0.2	0.5
Заходить	1	0.2	0.5
Подробность	1	0.1	0.7
Получить	**1**	**0.6**	**0.3**
Предлагать	**1**	**0.9**	**0.1**
Работа	**5**	**0.99**	**0.2**
Сайт	1	0.5	0.5
Смотреть	1	0.2	0.7
Ссылка	1	0.5	0.5
DDD$	***1***	***0.99***	***0.01***
Итого	***18***	***0,71***	***0,31***

↓ ↓
CHP

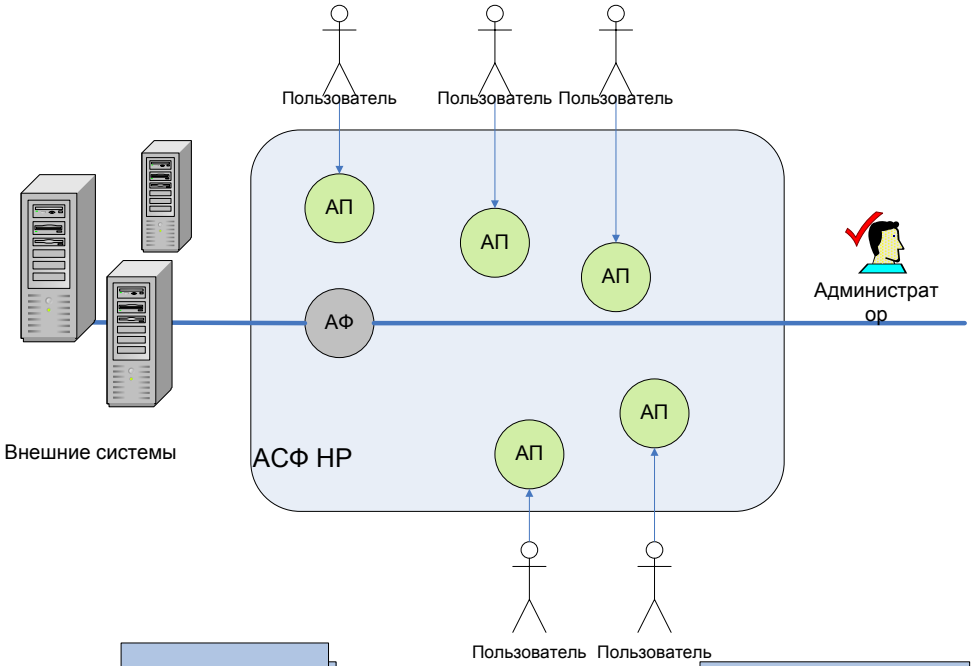
Комбинированная фильтрация

- *Формальный фильтр* – фильтрация по спискам адресов
- *Контекстный фильтр* – фильтрация по комбинации ключевых слов
- *Коллективный фильтр* – фильтрация на основе решений всех пользователей



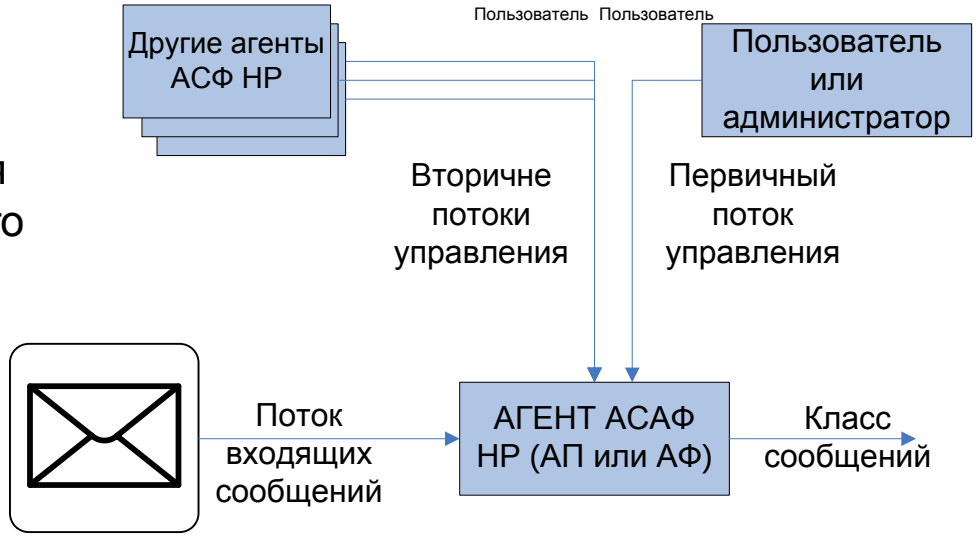
Многопользовательская структура

- *Агент фильтрации* – компонент, настроенный на решение общих задач фильтрации
- *Агент пользователя* – компонент, обеспечивающий индивидуальные для пользователя настройки

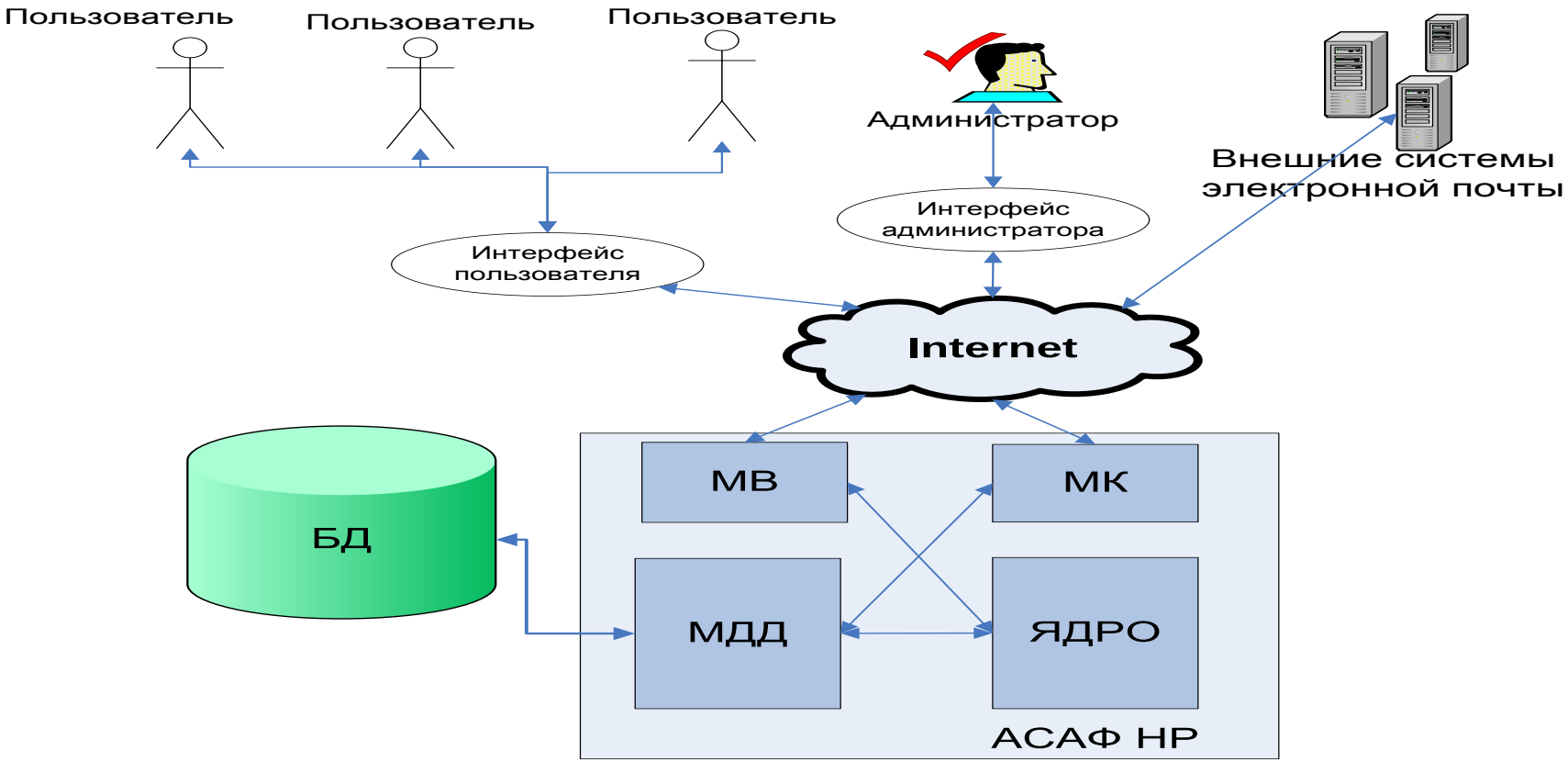


Параметры управления фильтрацией

- Первоначальное решение принимается на основе настроек пользователя данного агента
- В случае недостаточной точности оценки используются данные настройки остальных агентов АСФ НР

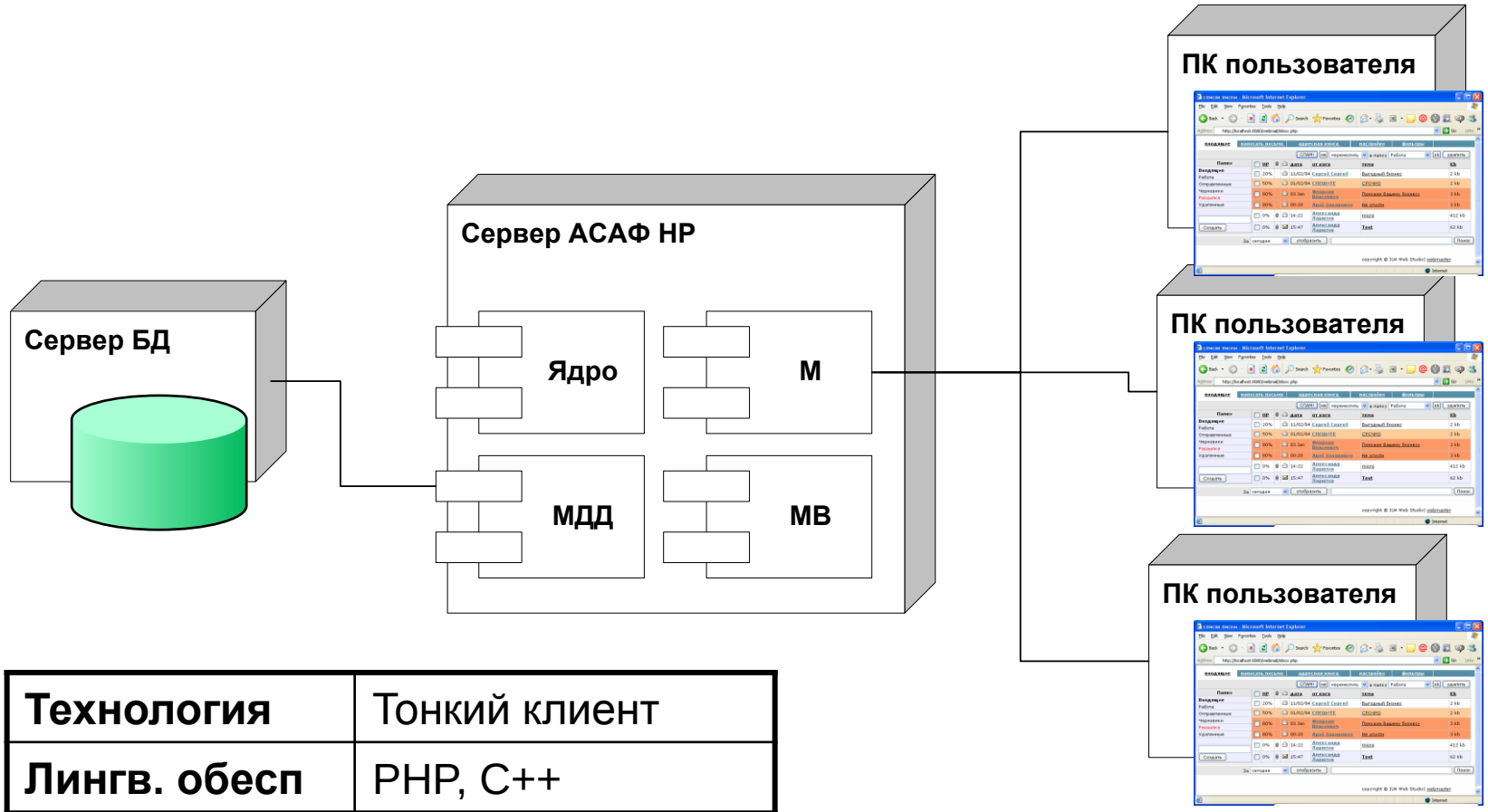


Архитектура АСАФ HP



- ***ЯДРО*** – модуль, реализующий основные функции фильтрации
- ***МК*** – модуль, реализующий отображение графического интерфейса
- ***МДД*** – модуль, обеспечивающий доступ к базе данных
- ***МВ*** – модуль, обеспечивающий взаимодействие остальных модулей

Диаграмма развертывания комплекса



Технология	Тонкий клиент
Лингв. обесп	PHP, C++
ОС	ASP Linux
СУБД	MySql
WWW-сервер	Apache

Диаграмма вариантов действия пользователя

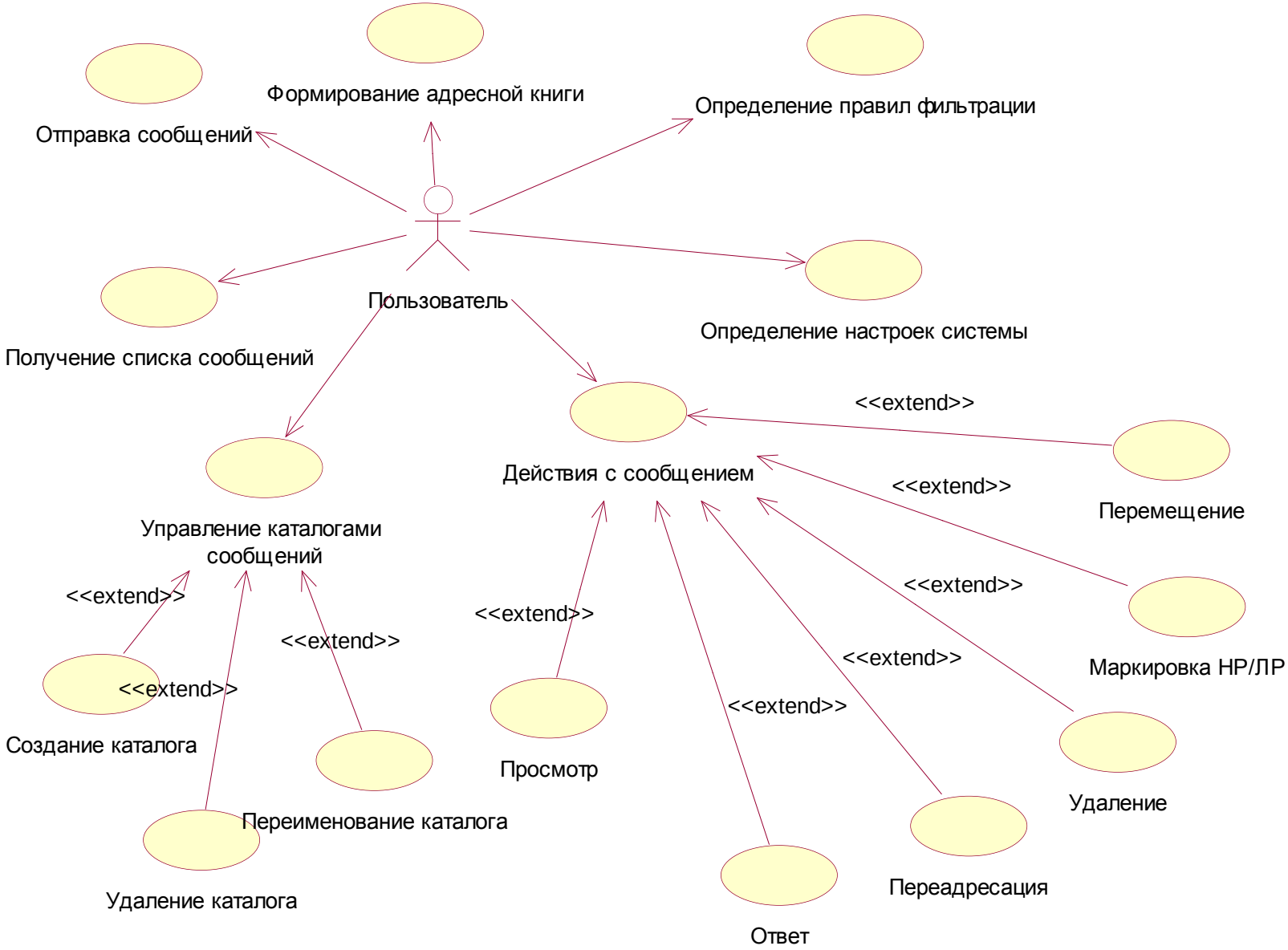


Диаграмма вариантов действия администратора

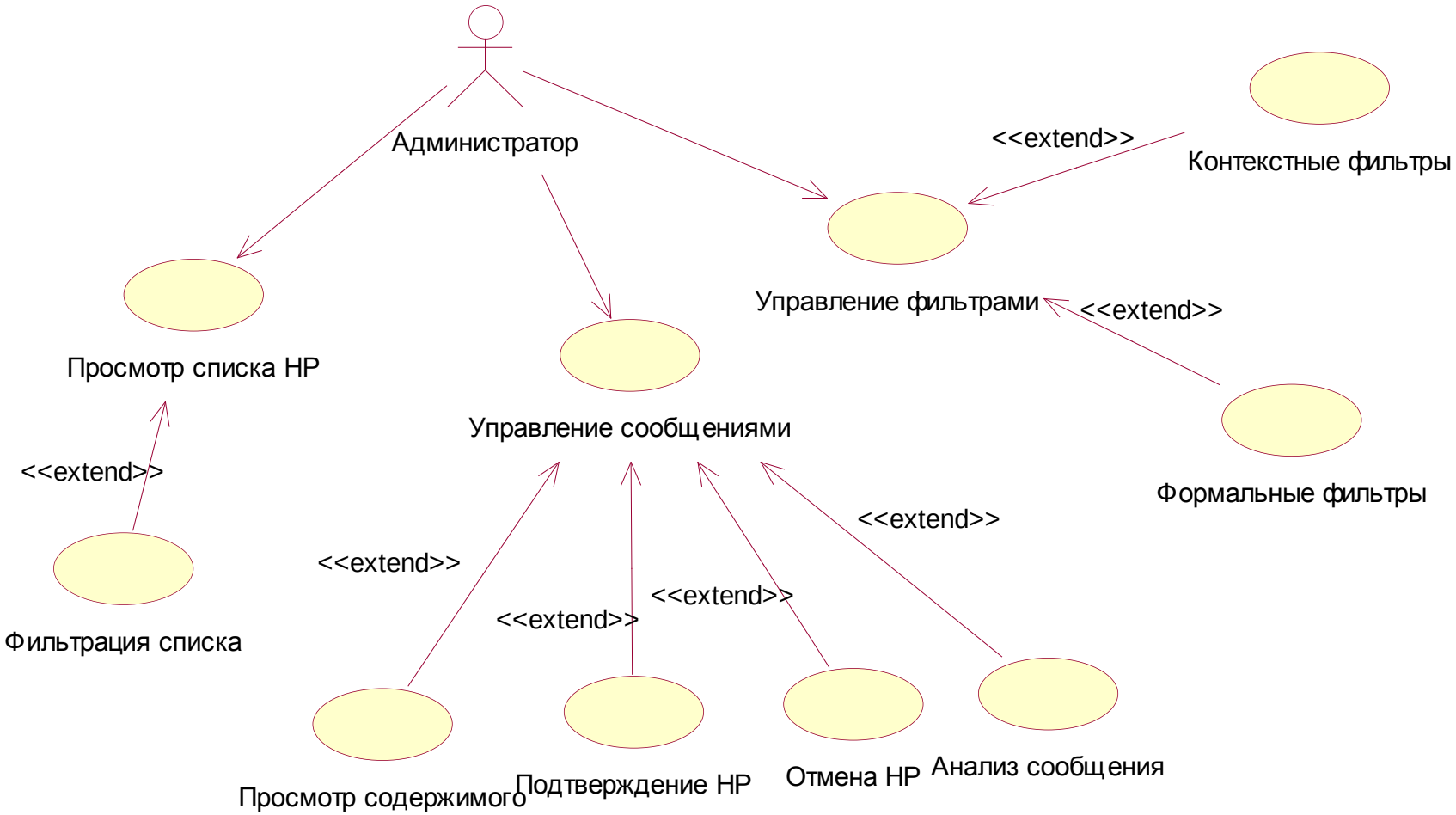
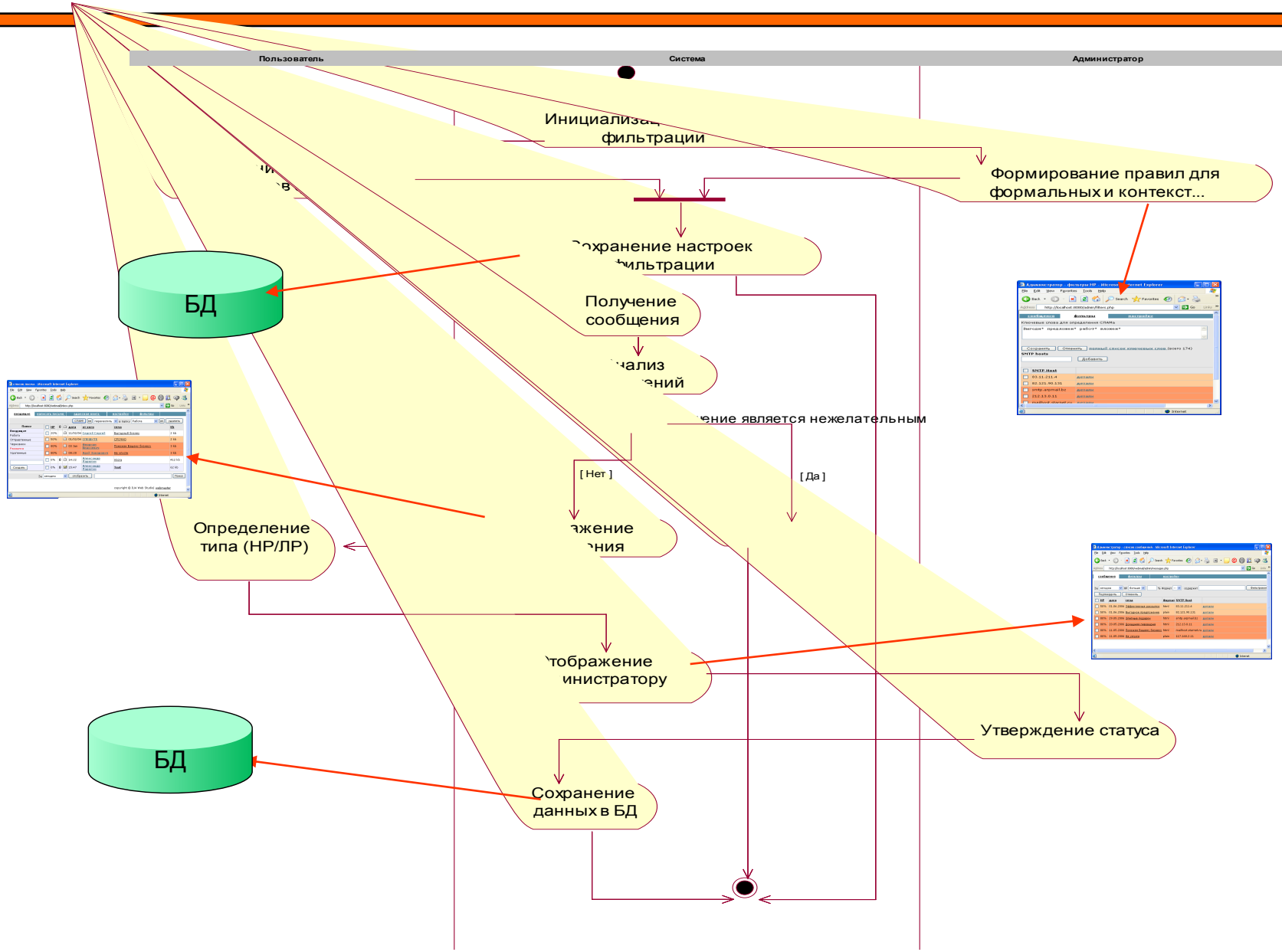


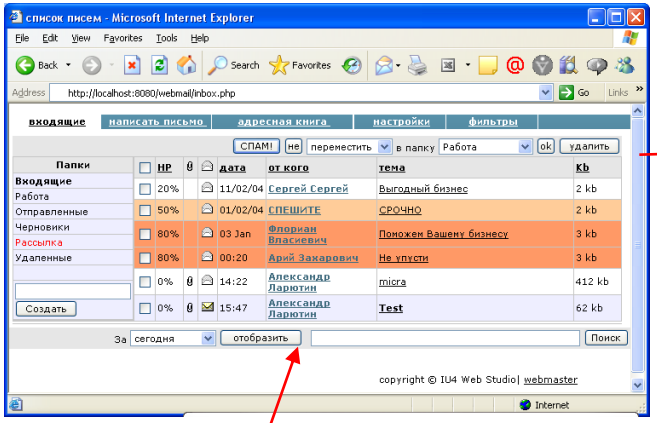
Диаграмма действий функционирования системы



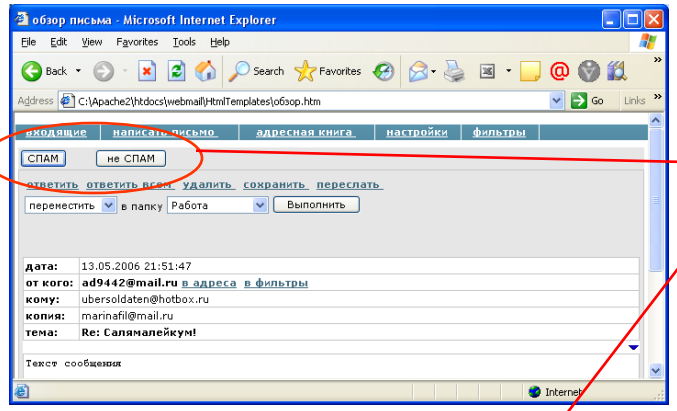
Примеры графического интерфейса

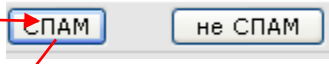
Пользователь

Список сообщений

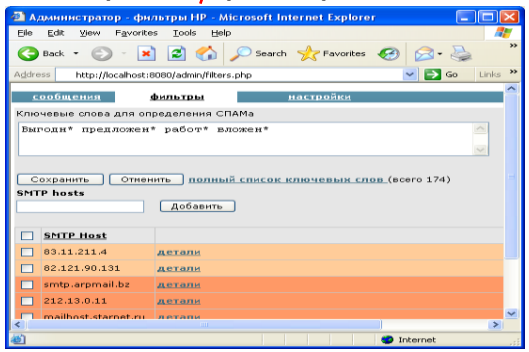


Обзор сообщения

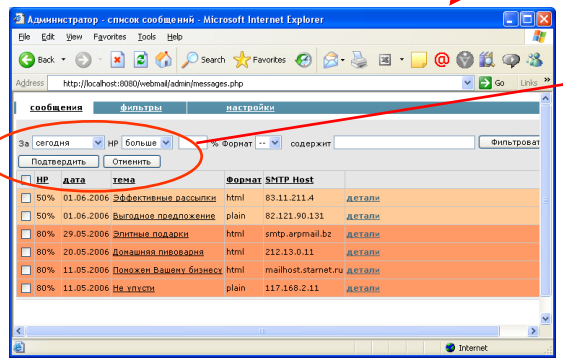




Настройки фильтрации



Список сообщений HP



Администратор

Комплекс защиты электронной почтовой записи от незапрашиваемой рассылки

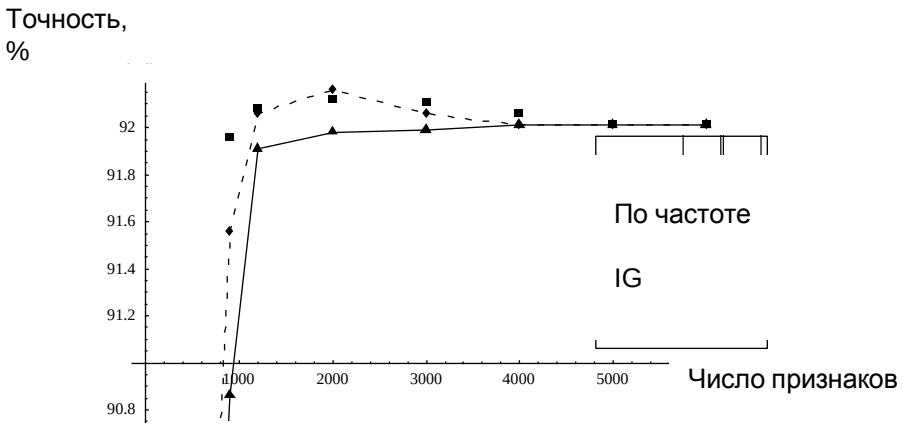
Экспериментальные исследования

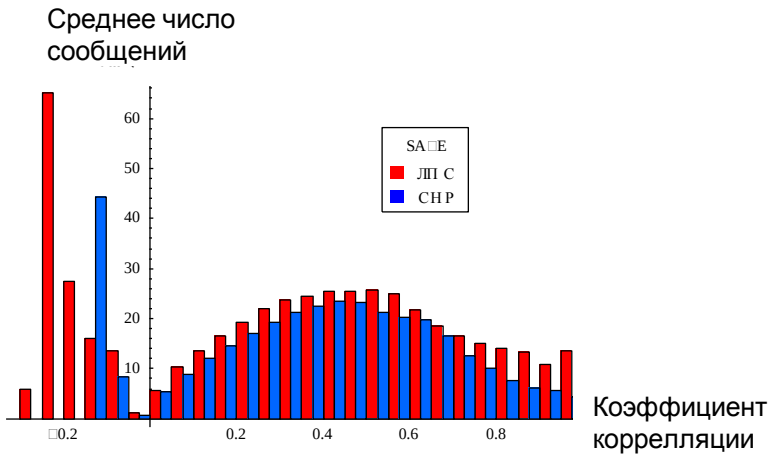
Объекты исследований:

1) Методы выбора значимых слов
2) Оценка признаков по к-там корелляции
3) Методы коллективной и комбинированной фильтрации

№	Массив	Число СНР		Число ЛПС	
		Обучение	Проверка	Обучение	Проверка
1	SA-A	1683	187	3734	415
2	SA-H			225	25
3	SA-E			3509	390

Результаты исследований





Результаты исследований

1) Наиболее эффективным при формировании словаря значимых слов является метод Хи квадрат ;
2) Достаточно ограничить число признаков до 2000;
3) Наиболее часто встречаются признаки с коэффициентом корелляции 0.5 в окрестности 0.1

Исследование пороговых значений

№ п/п	Массив	T_S	$P_{S->S}$	$P_{S->U}$	T_H	$P_{H->H}$
1	SA-E	-18	87,0%	13,0%	12	89,2%
2	SA-H	-22	83,7%	16,3%	29	78,4%

Исследование совместной фильтрации

№ п/п	Условия оценки	$P_{S->S}$	$P_{H->H}$	$P_{S->U}$
1	До применения совместной фильтрации	87,0%	89,2%	13,0%
2	После применения совместной фильтрации	91,2%	94,1%	8,8%

Таким образом, система обеспечивает фильтрацию до 91.2% незапрашиваемой рассылки

Заключение

Результаты

- Проведено исследование, классификация и систематизация существующих многопользовательских систем фильтрации НР.
- Исследована и разработана архитектура многоагентной автоматизированной системы адаптивной фильтрации НР, обеспечивающая эффективное взаимодействие пользователей системы при настройке ее фильтров.
- Исследованы и разработаны математические модели и методы анализа содержания сообщений электронной почты, что позволило повысить точность фильтрации.
- Разработан удобный графический инструментарий пользователя и администратора для работы с системой посредством общедоступных каналов связи.
- Проведены экспериментальные исследования предлагаемых в работе методов и алгоритмов, позволившие определить оптимальные параметры, обеспечивающие их максимальную эффективность.

Апробация

- Издано 3 публикации на тему диссертации
- Результаты работы использовались для получения грантов МОН РФ
- Комплекс прошел успешную апробацию в тестовом зале компании «ООО Артезио» (30 рабочих мест)