

**Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«Оренбургский государственный университет»**

**Иванова Екатерина Петровна**

# **Электронная база данных археологических находок**

Направление 09.04.02 – Информационные системы и технологии  
Магистерская программа – Информационные системы в научных исследованиях

Научный руководитель  
доктор технических наук,  
профессор Чепасов В.И.

Оренбург  
2018

### Базовая

1. Смекалов С. Л., Федорова Д. Л. Геоинформационные технологии в археологических исследованиях / Балт.гос.техн.ун-т - Санкт-Петербург, 2004. 104 с.
2. Холюшкин Ю.П., Витяев В.В., Костин В.С. Задачи археологии и методы их решения, 2013. 100 с.

### Периодическая

1. Холюшкин Ю.П., Костин В.С. Некоторые подходы к анализу данных в археологии / Вестник НГУЭУ. 2015. № 4. С. 240- 258.
2. Холюшкин Ю.П., Витяев В.В., Костин В.С. Проблемы автоматизации информационных потоков в археологии / Информационные технологии в гуманитарных исследованиях. 2011. № 16. С.58-71.
3. Федоров Д.Л., Смекалов С.Л. Особенности построения геоинформационных систем для археологических исследований / Информация и космос. 2004. № 1. С. 50-54.
4. Краснопрошин А.И., Бурнос Д.В., Андреева И.И. Анализ археологических пространственных данных в геоинформационной системе / Сборник статей Международной научно-практической конференции: В 4 частях. Актуальные проблемы современной науки отв. редактор А.А. Сукиасян. 2013. С. 173-176.

**Цель работы:** Автоматизация информационных процессов геоинформационной системы объектов археологического наследия Оренбургского района

**Анализируемые свойства:**

достоверность;  
точность.

**Целевая функция**

$$\alpha_1(P, N(T)) \leq 0,05$$

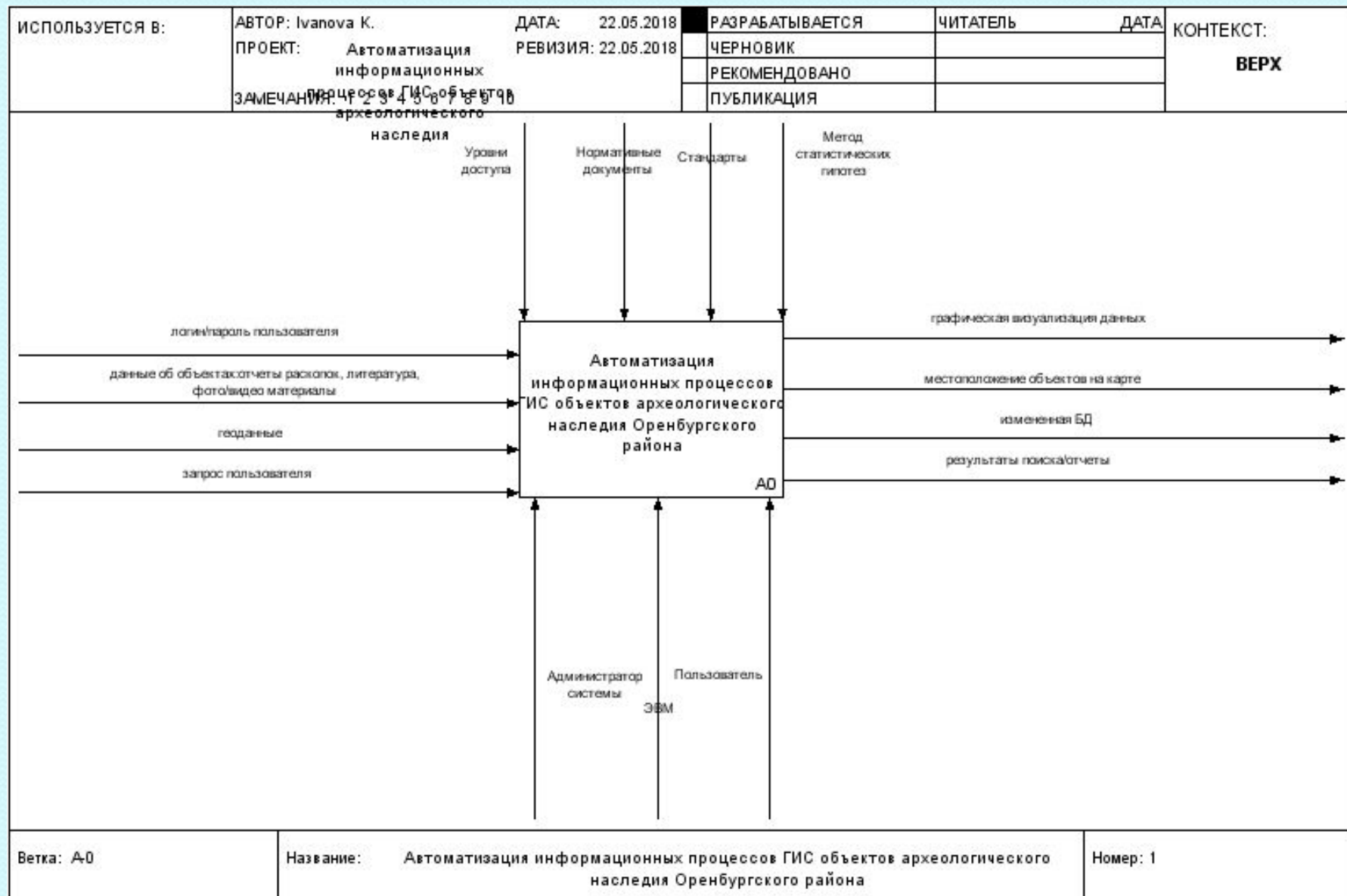
где  $\alpha_1$ - ошибка первого рода;

$P_{дм}$  – достоверное местонахождение объектов археологического наследия

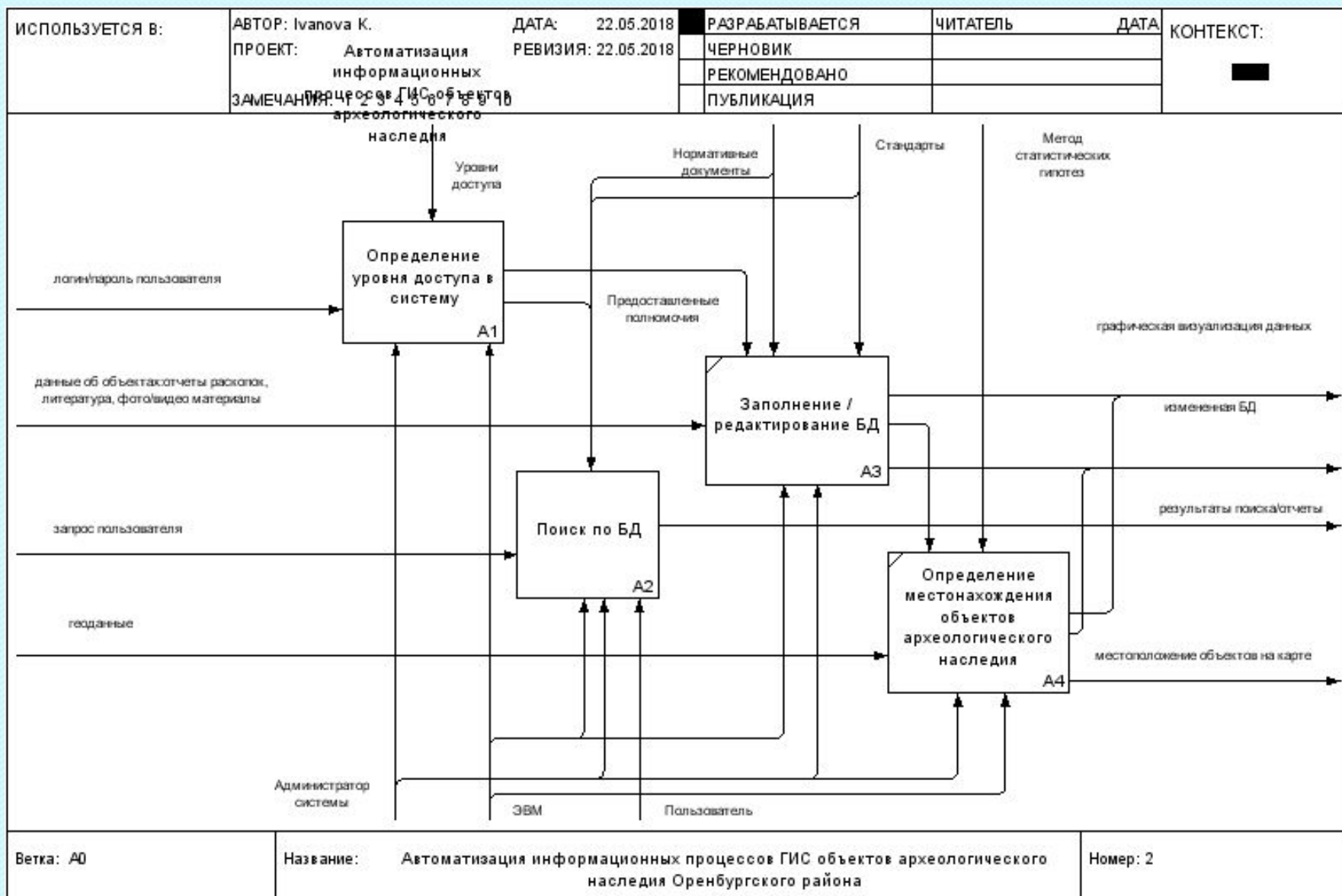
$N$  – число найденных объектов в заданном районе (ед.)

$T$  – временная принадлежность (век)

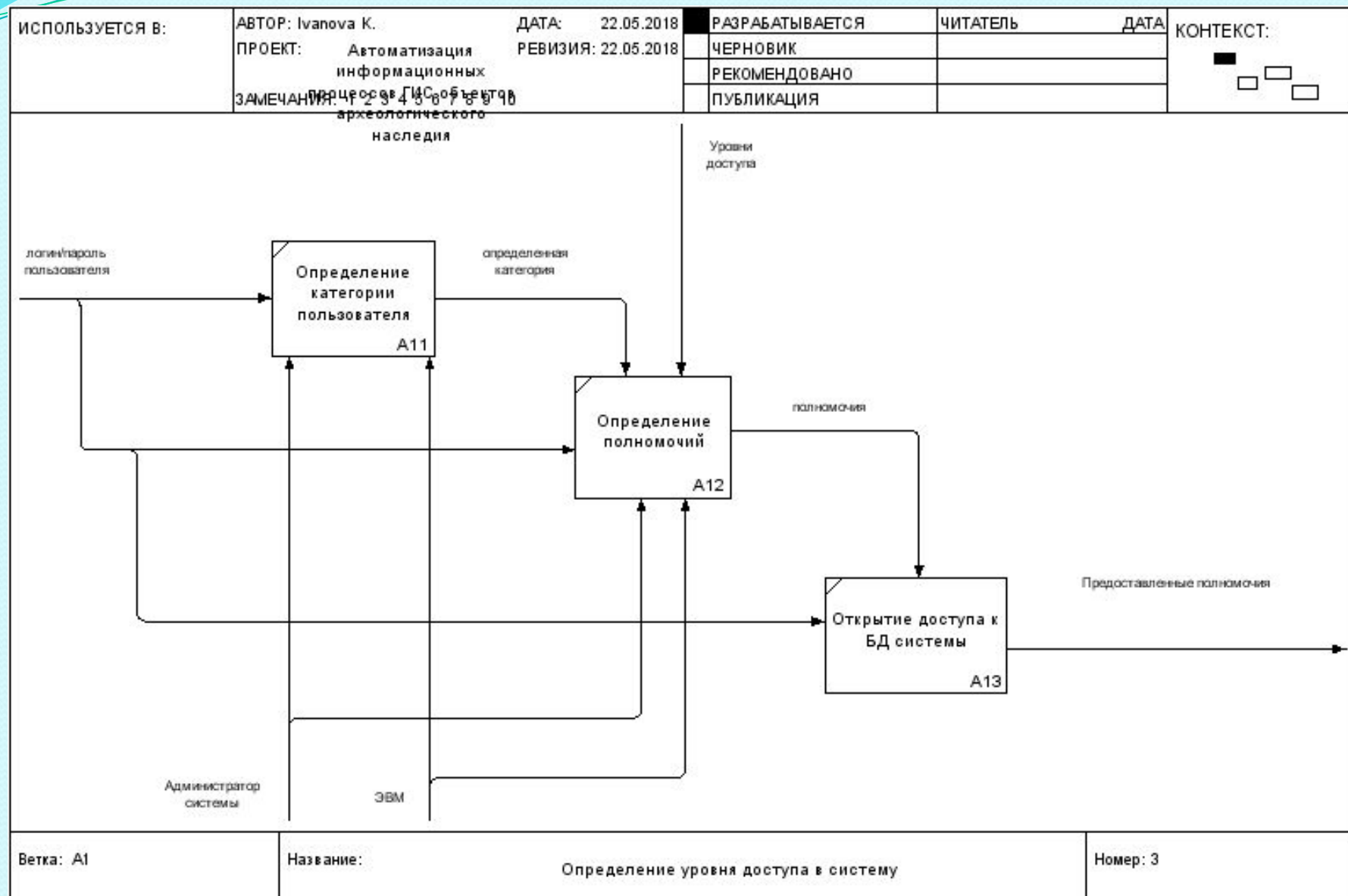
# Контекстная диаграмма



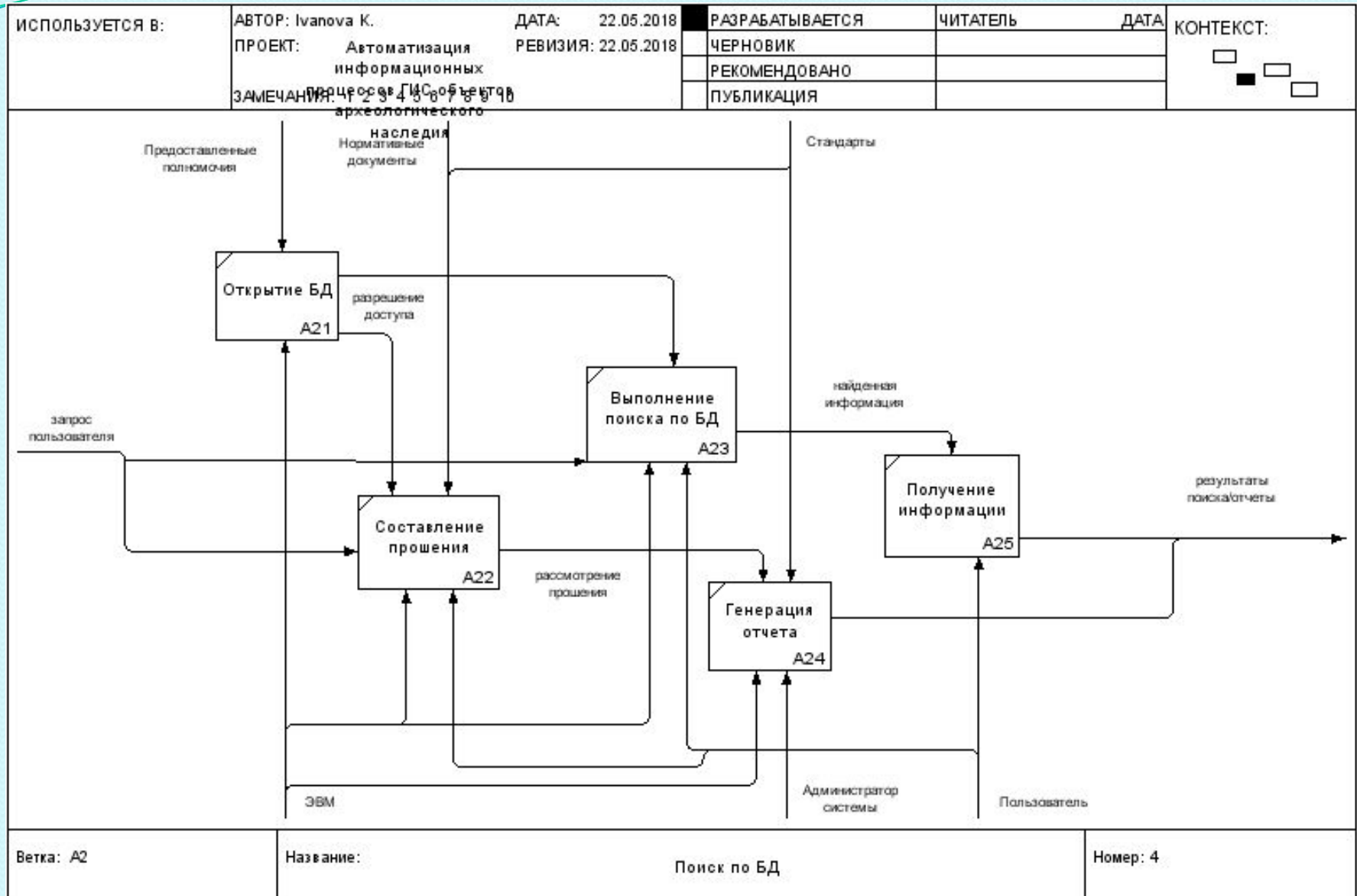
# Декомпозиция контекстной диаграммы



# Декомпозиция контекстной диаграммы



# Декомпозиция контекстной диаграммы



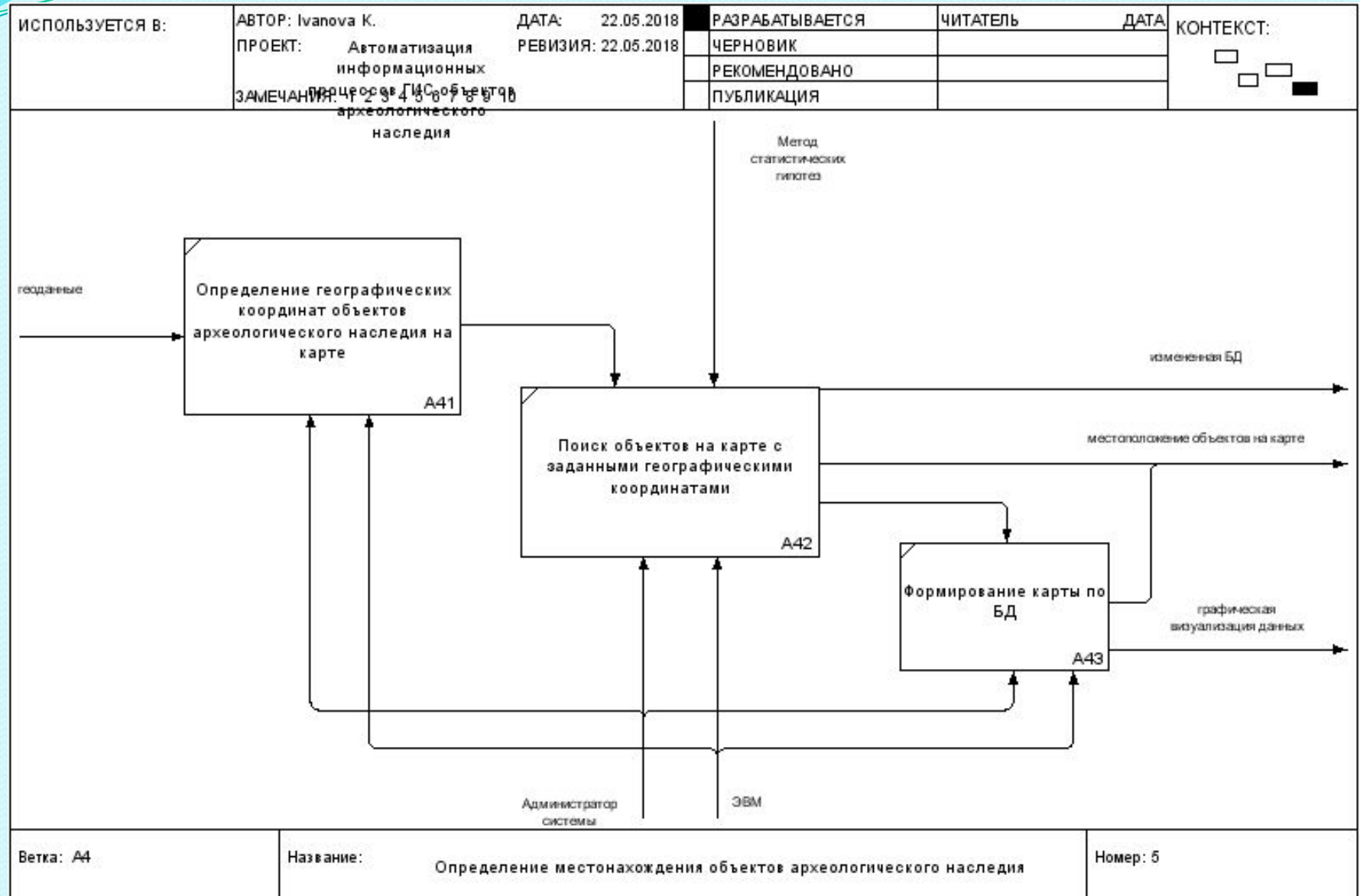
Ветка: A2

Название:

Поиск по БД

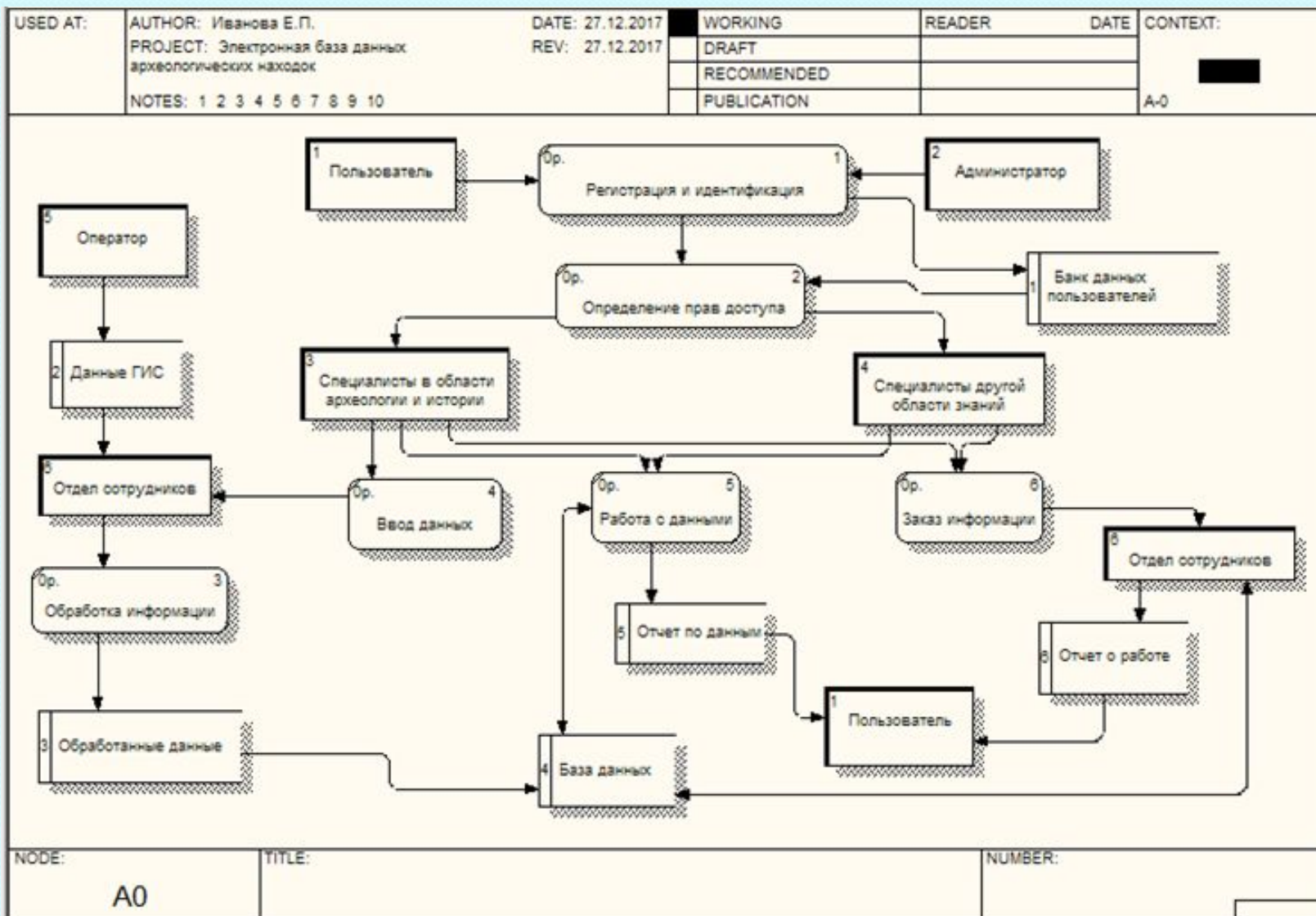
Номер: 4

# Декомпозиция контекстной диаграммы





# DFD модель



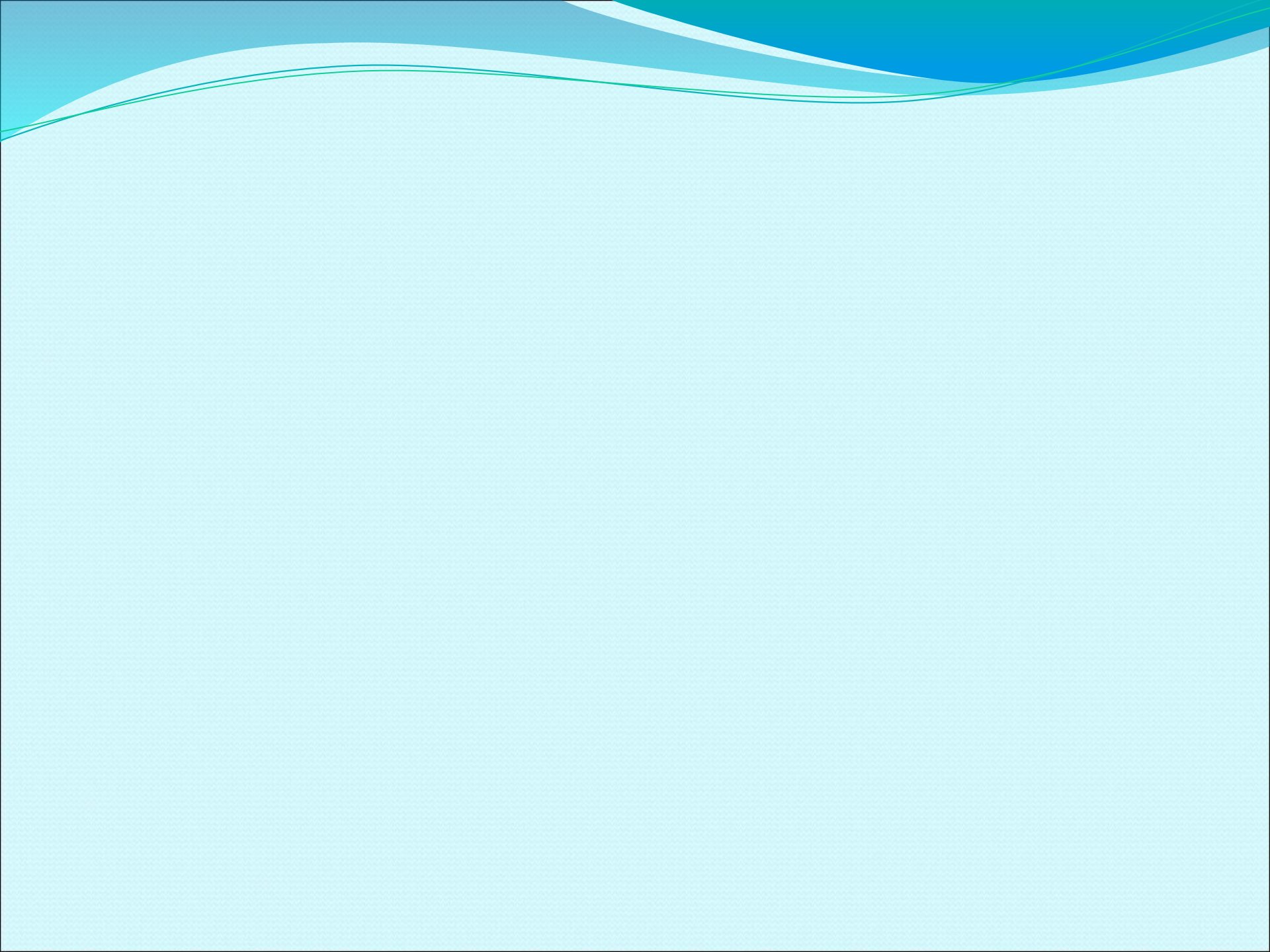
NODE:

A0

TITLE:

NUMBER:

# МОДЕЛЬ



# МОДЕЛЬ ЭЛЕКТРОННОЙ КОРРЕСПОНДЕНЦИИ



## Модель почтового ЭС

$$S(p_i) = \langle t_i, w(t_i) \rangle$$

где  $t_i$  –  $i$ -ый терм в сообщении;

$p_i$  – пространство признаков, определяющих сообщение;

$w(t_i)$  – вес термина в сообщении после удаления стоп-слов.

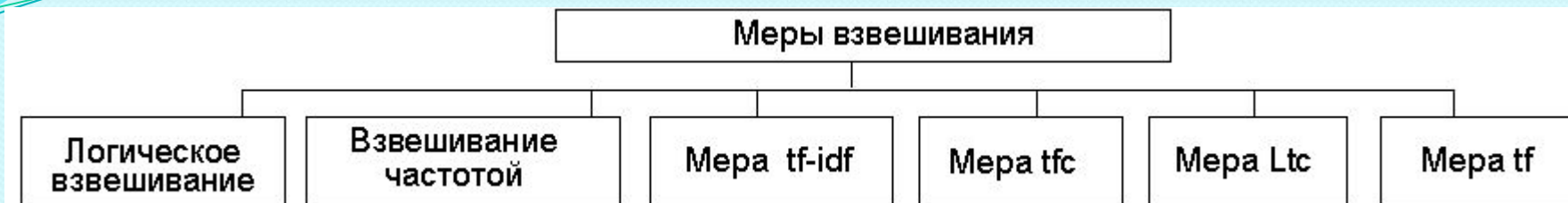
$$S_j = \begin{bmatrix} w_{1j} \\ \dots \\ w_{ij} \\ \dots \\ w_{Nj} \end{bmatrix}$$

где  $S_j$  –  $j$ -ое сообщение электронной корреспонденции;

$w_{ij}$  – вес термина  $i$  в сообщении  $j$ ;

$N$  – число термов в сообщении.

# МОДЕЛЬ ЭЛЕКТРОННОЙ КОРРЕСПОНДЕНЦИИ



$$Ltc_{ij} = \frac{\log(w_{ij} + 1) \log\left(\frac{M}{M_j}\right)}{\sqrt{\sum_{t_j=1}^N \left[ \log(w_{ij} + 1) \log\left(\frac{M}{M_j}\right) \right]^2}}$$

Матрица значимости класса  $L_k$

$$L_k = \langle T_k, w(t_j) \rangle$$

где  $T_k$  –  $k$ -ый тезаурус (класс) сообщения;  
 $w(t_j)$  – вес термина в сообщении

$$w_{it_j} = Ltc_{it_j}$$

где  $M$  – общее число сообщений в выборке;  
 $N$  – число терминов в выборке после удаления стоп-слов;  
 $M_j$  – общее число сообщений, содержащих терм  $t_j$ .  
 $w_j$  – вес термина (частота повторений) в сообщении после удаления стоп-слов.

$$L_k = \begin{bmatrix} w_{11} & w_{21} & \boxtimes & w_{j1} \\ w_{12} & w_{22} & \boxtimes & w_{j2} \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes \\ w_{1i} & w_{2i} & \boxtimes & w_{ji} \\ \boxtimes & \boxtimes & \boxtimes & \boxtimes \\ w_{1N} & w_{2N} & \boxtimes & w_{MN} \end{bmatrix} \quad \begin{matrix} j = 1, \dots, M, \\ i = 1, \dots, N, \end{matrix}$$

где  $L_k$  – модель класса  $k$  (*spam/legitim*);  
 $w_{ij}$  – вес термина  $i$  в ЭС  $j$   
 $N_j$  – число терминов в ЭС;  
 $M$  – число ЭС в классе.