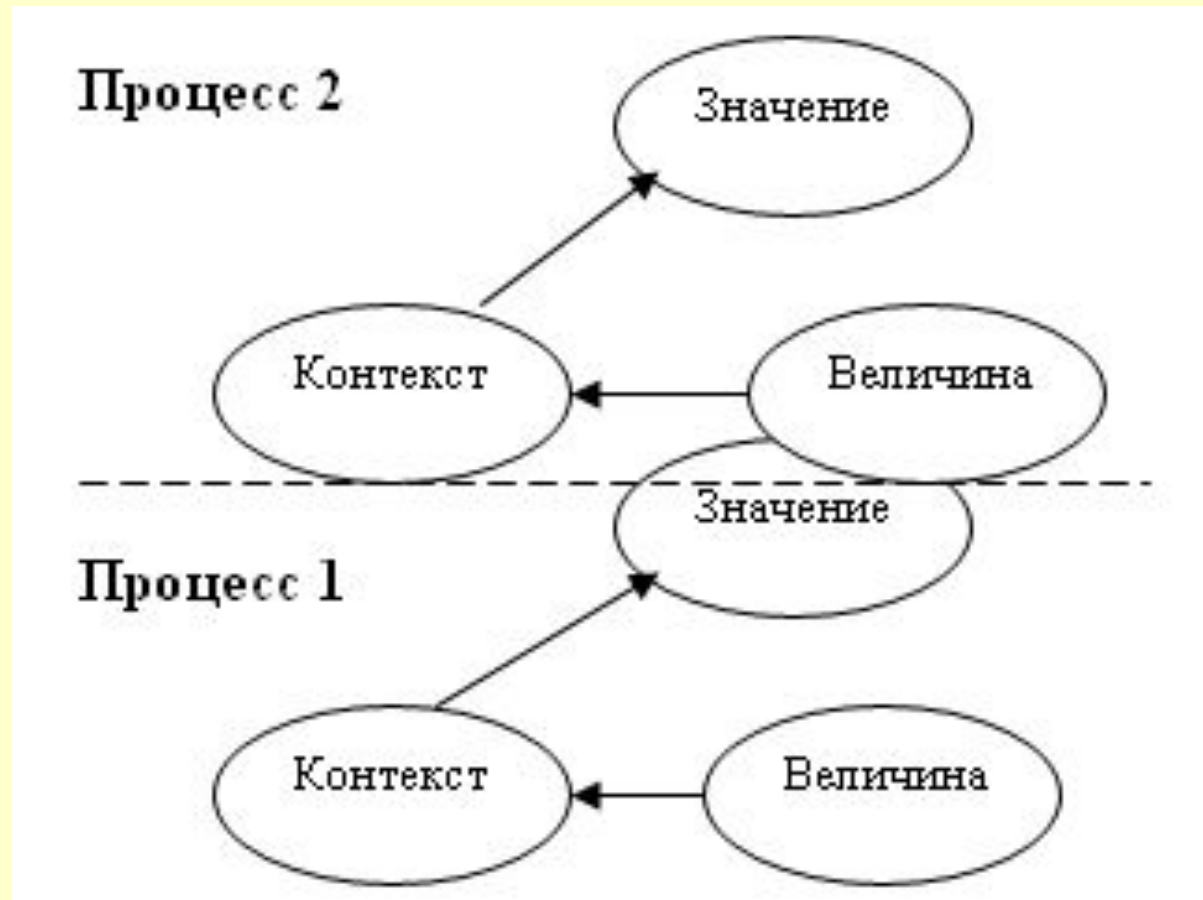


Лекция 5

**Теоретические основы документальных
БД. Теоретико-множественная модель
индексирования и поиска. Структура
процессов в абстрактной АИПС.
Примеры структур документо-
ориентированных баз данных**

Соотношение понятий «величина», «контекст» и «значение»



Индексирование реферата РЖ ВИНТИ «Информатика»

Управление библиотечным и информационным центром. Trosow Samuel E.,
Libr. Quart., 2000, № 70, 153-155

Рецензируемая книга. (R. D. Stueart, B. B. Moran. Library and information center management, 5th ed. Englewood, Colo.: Libraries Unlimited Inc., 1998, xxv, 509 p.), вышедшая уже в 5-м издании, давно стала стандартным учебником по курсу управления в рамках библиотечной и информационной науки, охват материала в котором расширялся с каждым очередным изданием. В предисловии отмечается, что значительные изменения в окружающих условиях, вызванные внутренними и внешними факторами, требуют более систематического подхода к обзору функций в условиях организации. Технологии, политическая, экономическая и социальная среда указываются как наиболее мощные силы для изменений. При подготовке данного издания авторы решили сохранить классическую структуру, включающую планирование, организацию, подбор персонала, определение направлений, координацию и финансирование. Они считают такой подход продуктивным и не признают постоянной критики данной структуры со стороны современных теоретиков по организации. Помимо отдельных глав, посвященных перечисленным вопросам, авторы уделяют большое внимание теории организации, связи социологической теории и ее применений в организационном анализе, различным уровням анализа, на которых могут изучаться организационные явления.

библиотеки
информационные центры
управление
функции руководителей
рецензии
США

Теоретико-множественное описание информационных потоков

D – универсальный словарь

l_i – совокупность лексических единиц документа

L_0 – универсальный поток (массив):

$$L_0 = \{l_1, \dots, l_k, \dots, l_{n_0}\}, |L_0| = n_0, l_k \subset D \text{ для всех } k$$

Линейное представление теоретико-множественного образа документа

$$l_k = \begin{pmatrix} b_{1k} \\ \vdots \\ b_{ik} \\ \vdots \\ b_{Dk} \end{pmatrix} \text{ - поисковый образ документа (ПОД)}$$

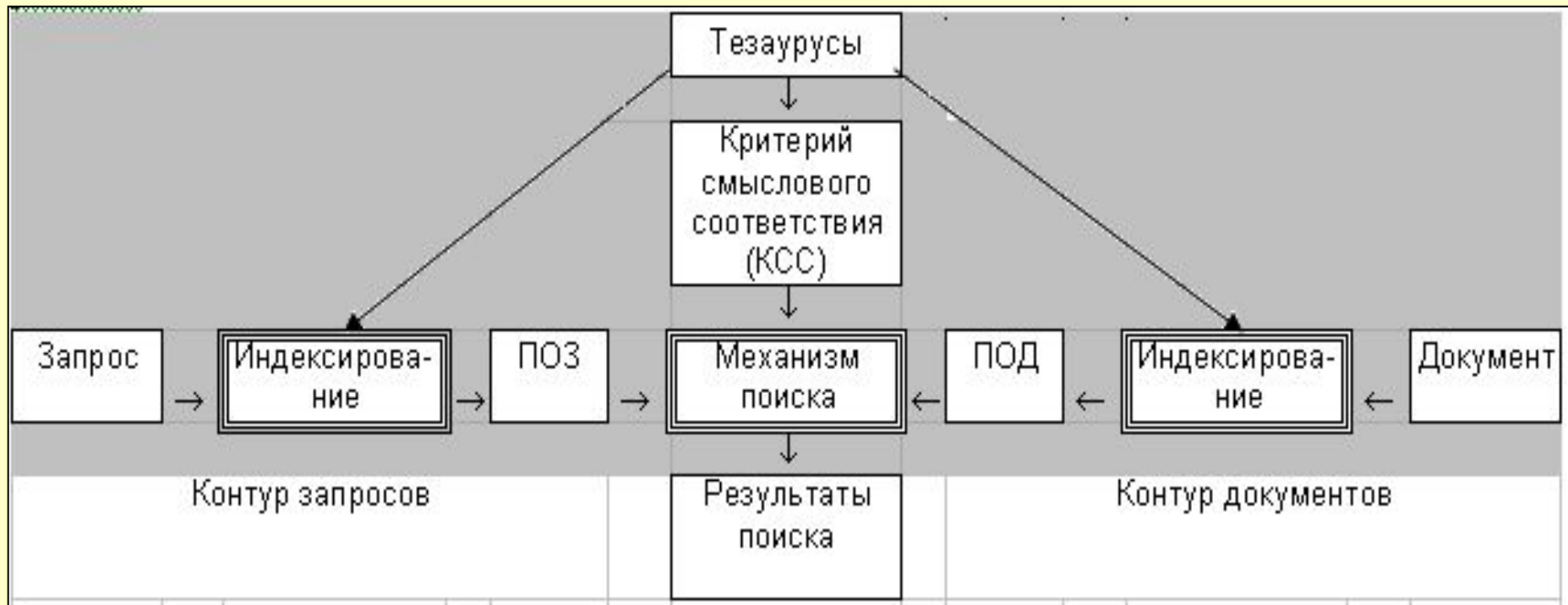
$$b_{ik} = \begin{cases} 1, & \text{если } i\text{-й термин входит в } k\text{-й документ} \\ 0, & \text{если не входит} \end{cases}$$

Универсальный массив в линейном представлении:

$$L_0 = \begin{pmatrix} b_{11} b_{12} \cdots b_{1n_0} \\ \cdots \cdots \cdots \\ b_{i1} b_{i2} \cdots b_{in_0} \\ \cdots \cdots \cdots \\ b_{D1} b_{D2} \cdots b_{Dn_0} \end{pmatrix} \text{ - матрица «термин-документ»}$$

$$F_i = \sum_{k=1}^{n_0} b_{ik} \text{ - частотная характеристика термина}$$

Схема процессов в абстрактной АИПС



Критерий смыслового соответствия (КСС)

КСС есть пара:

$$\text{КСС} = \langle \beta^*, \kappa^e \rangle$$

где β^* есть мера *формальной релевантности*, или мера близости (способ исчисления близости) поисковых образов документа и запроса,

κ^e - пороговое значение меры близости, при превышении которого документ признается формально релевантным соответствующему запросу.

Меры формальной релевантности

- скалярное произведение векторов ПОЗ и ПОД:

$$\beta_{mk}^* = \sum_{i=1}^D a_{im} \times b_{ik}$$

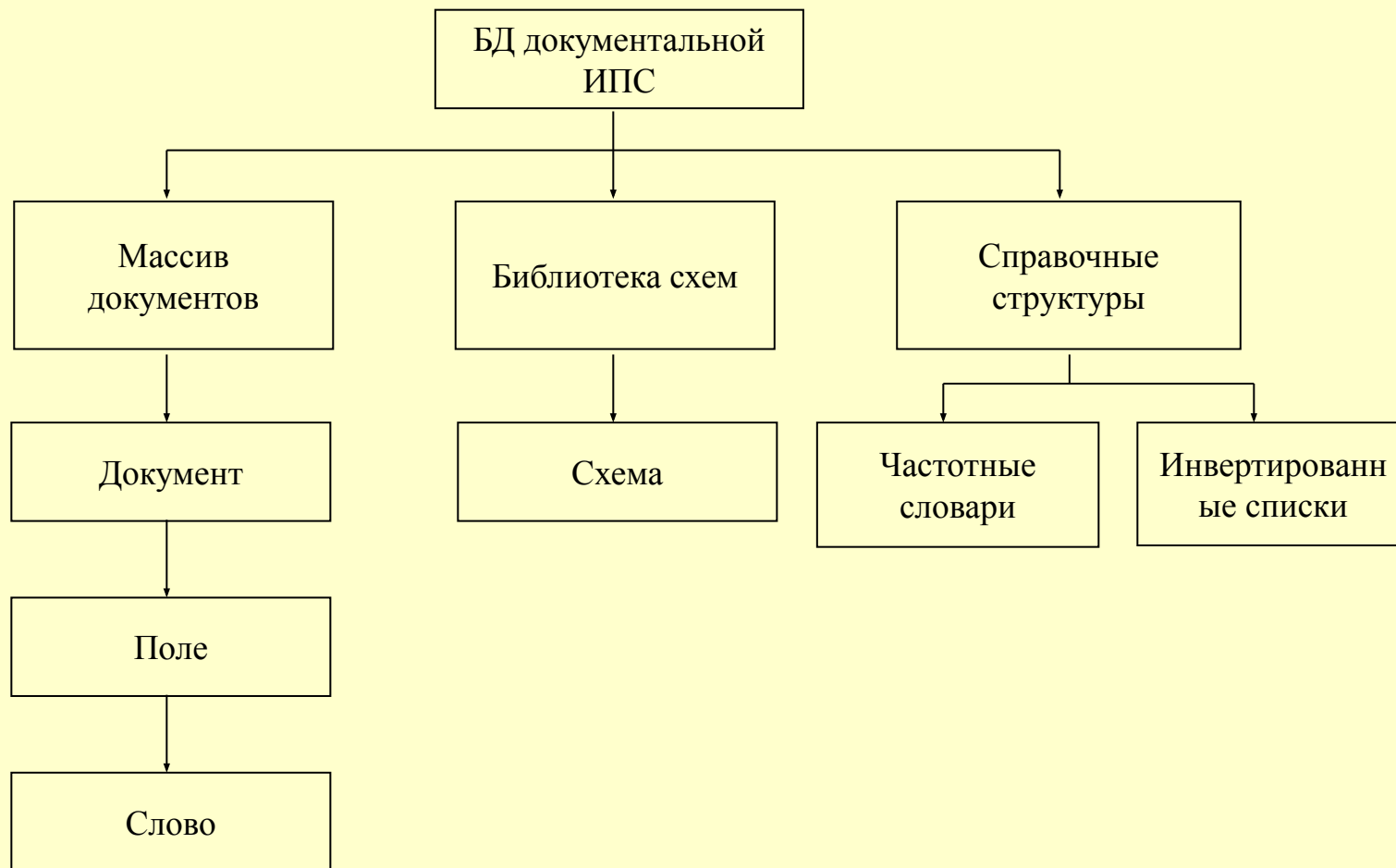
- мера Танзимото (нормированное скалярное произведение):

$$\beta_{mk}^* = \frac{\sum_{i=1}^D a_{im} \times b_{ik}}{\left(\sum_{i=1}^D a_{im} + \sum_{i=1}^D b_{ik} - \sum_{i=1}^D a_{im} \times b_{ik} \right)}$$

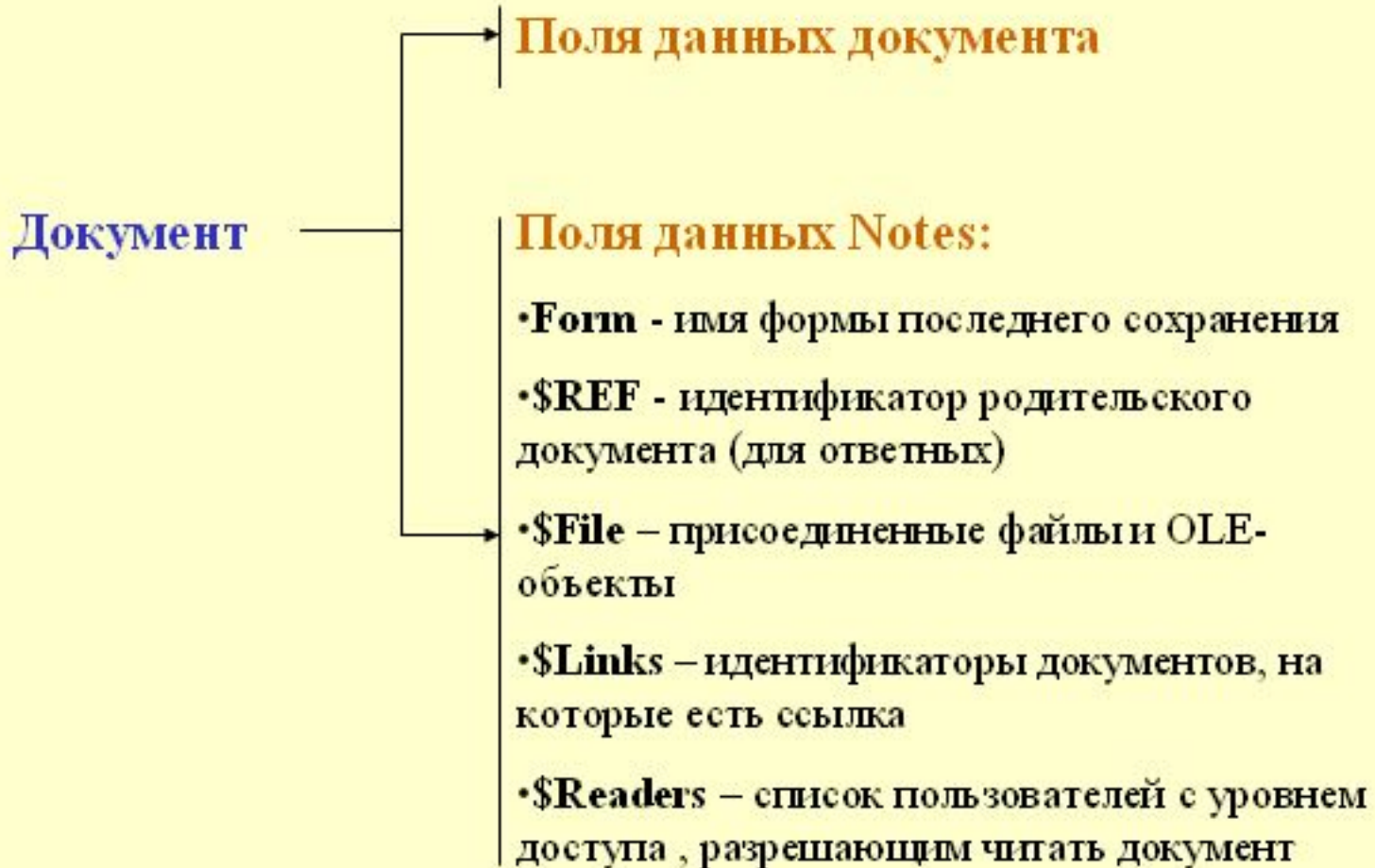
- мера косинуса угла векторов ПОЗ и ПОД:

$$\beta_{mk}^* = \frac{\sum_{i=1}^D a_{im} \times b_{ik}}{\sqrt{\left(\sum_{i=1}^D a_{im}^2 \right) \times \left(\sum_{i=1}^D b_{ik}^2 \right)}}$$

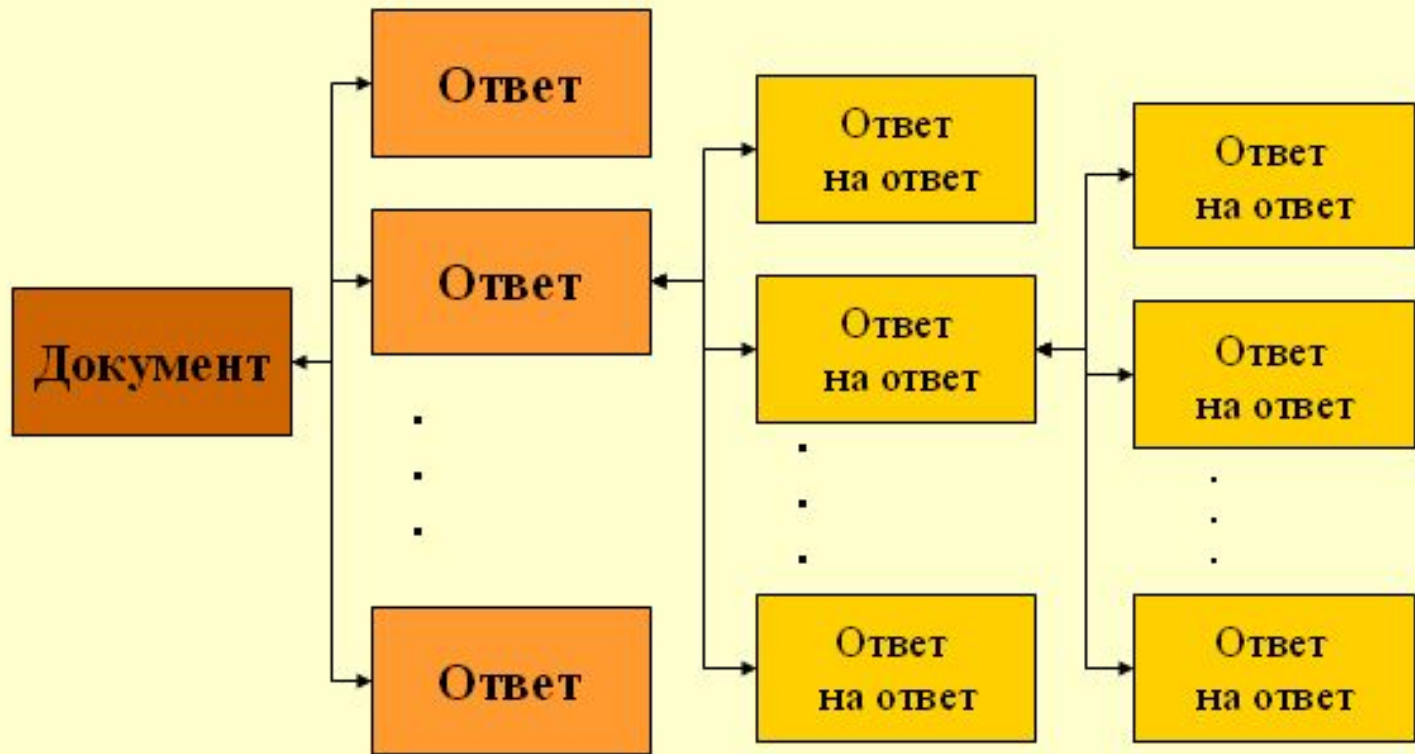
Логическая структура БД документальной ИПС IRBIS



Структура документа Notes



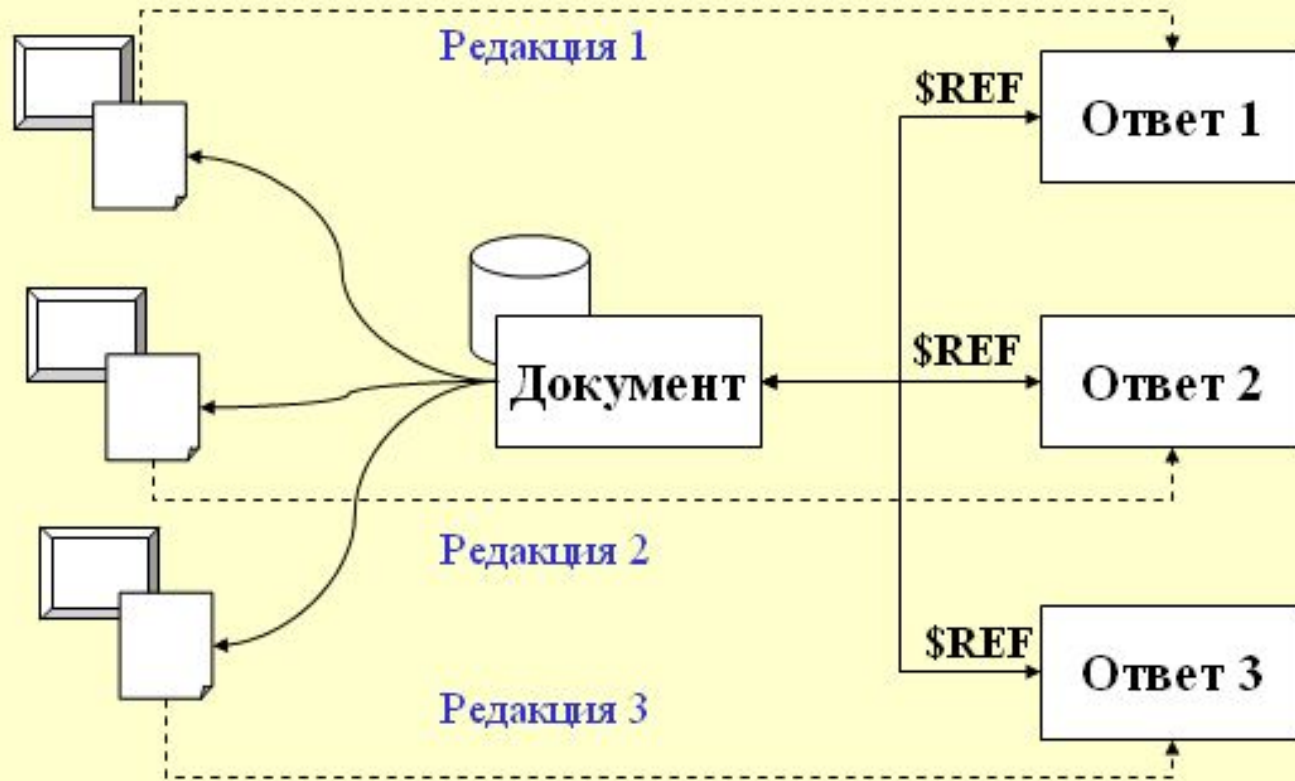
Связи документов Notes



$\$REF(\text{Документ})$

$\$REF(\text{Ответ/Ответ на ответ})$

Связи документов Notes



Новые версии становятся ответами – сам документ не изменяется

Связи документов Notes

НОВЫЕ ВЕРСИИ СТАНОВЯТСЯ ОСНОВНЫМИ

