

Проект «Архив ММК» (на уровне эскиза)

1. Ситуация

- В настоящее время худо-бедно описан только архив **ГД**. По **моим** грубым оценкам (сделанным по описи) в архиве ГД порядка **100 тыс. стр.** Большинство из них — машинописные, но, как правило, плохого качества.
- Опись архива составлена в весьма свободном жанре: нет единого языка описания, единого алгоритма, отсутствует важная информация и т.д. Единицей описания в ней является "папка", а не "единичный" (элементарный) текст. Использовать данную опись для компьютерного Архива непосредственно (без существенных затрат по адаптации) невозможно.
- Существует база данных, разработанная группой М.Украинского. Но, к сожалению, она разработана не **for Windows**. Поэтому я смею утверждать, что эта база данных "мертворожденная".
- Согласно информации, полученной от К.Зендрикова, существует 80 дискет (100 Мб) с разного рода текстами, но без описи. Большинство текстов на этих дискетах в формате "dds", следовательно, их объем порядка **50 тыс. стр.** Весьма вероятно, однако, что не все тексты относятся к Архиву ММК, и могут быть дубли.
- На компьютере у Н.Крюковой хранятся тексты (и рисунки) опубликованных работ ГП, вошедших в "Избранные труды".

2. Представление о цели

В данный момент я не имею понятия о том, как функционирует современный компьютерный архив. Однако у меня есть самое общее представление о *первом приближении* к цели. Я полагаю, что даже не вполне совершенный, но способный к развитию '**мини-архив**', **содержащий 20-25 тыс. стр. и организованный достаточно мощной СУБД**, представляет собой значительный вклад в дело включения наследия ММК в культурный оборот.

И такое первое приближение (при бюджете 1.500 \$ в месяц) можно получить за весьма короткий срок — полтора года.

Затем можно будет поставить вопрос о втором приближении (с учетом того, что через полтора года ситуация наверняка потребует коррективов) и т.д.

3. Общая стратегия

3.1. Фаза порождения

- Архив порождается как исходная "клеточка", способная к рациональному **развитию** — **начальная (исходная) версия Архива**.
- Это порождение осуществляется уже сейчас — в ситуации отсутствия представления о конечном, развитом виде продукта: он будет уточняться по ходу работы.
- Необходимая и достаточная структура Архива в момент порождения включает два компонента: 1) элементарную БД "**Опись Архива**", созданную на основе СУБД **for Windows** (например, Paradox for Windows) и 2) **массив файлов Архива**, читаемых в редакторе **for Windows** (например, MS Word 6.0).
- Поля таблицы БД описывают не папки Архива, а **элементарные тексты** — единицы хранения.
- Массив файлов Архива включает в себя лишь те файлы, которые прошли процедуру сличения с оригиналом — процедуру **сертификации**.

3.2. Фаза запуска по минимуму

- Все существующие файлы (не прошедшие процедуру сертификации) соотносятся с Описью Архива и маркируются в ней.
- Запускается процедура сертификации промаркированных файлов.
- Запускается процедура поставки текстов для ввода их в компьютер с последующим возвратом их в архив ГД, а также процедура пополнения архива ГД.
- Запускается процедура ввода текстов с помощью сканера (по возможности без затрат на приобретение оборудования).

- Запускается процедура ввода текстов путем непосредственной набивки.
- Отрабатывается процедура создания Библиотеки схем.
- Отрабатывается порядок (условия) пользования Архивом.

3.3. Фаза отработки и оценки технологий

- Уточняются временные и стоимостные затраты для всех запущенных технологических цепочек.
- Собирается информация о других — более эффективных — технологиях.

3.4. Фаза проектирования развития

Эта фаза предполагает реализацию всех предыдущих и зависит от полученных результатов. Кроме того, она предполагает получение информации о современных архивах и библиотеках.

4. Тактические предложения

4.1. Фаза порождения

Эта фаза предполагает в данной версии следующие виды работ:

- Разработка формуляра (пустографки) для описания единицы хранения в Архиве.
- Разработка таксономической структуры полей БД (в частности определение перечня видов текста, перечня тем и т.д.).
- Разработка пакета программ для БД "Опись Архива".
- Подготовка существующей описи (описи имени Р.Максудова) для втягивания в БД (по минимуму) — в той мере, в какой это возможно без обращения к самим папкам документов.
- Заполнение БД.
- Описание в БД файлов, прошедших сертификацию. Таковыми на настоящий момент являются только файлы тома "Избранные труды". Они и составят содержание массива файлов исходной версии Архива на фазе его порождения.
- Загрузка компьютера исходной версией Архива.
- Оценка исходной версии.

Реализация всех перечисленных видов работ может быть организована мною. В этом случае общая стоимость этой фазы составит 1200 \$, а время реализации — 2.5 месяца.

4.2. Фаза запуска по минимуму

В рамках этой фазы необходимо (и, если не оговорено иное, то в той последовательности, в какой это перечислено ниже):

- Организовать работу над созданием Глоссария. (Могу взять это на себя.) Эта работа может вестись параллельно с работами на фазе порождения.
- Создать опись всех файлов (не прошедших сертификацию). Эта работа включает перевод файлов в формат *WinWord*, объединение файлов, составляющих один "единичный" текст, и создание отдельного файла-таблицы, в котором соотнесены имена файлов с названиями текстов в них. (Я умею это делать со скоростью 12 файлов в час. Следовательно, если не будет предложена более эффективная технология, опись 1500 файлов — такова моя очень грубая оценка их общего количества — потребует 125 часов и 250 \$.) Создание описи файлов может быть начато в любой момент — оно не требует завершения фазы порождения.
- Соотнести эту опись с Описью Архива (БД) и заполнить соответствующие графы (поля) последней именами файлов. Часть этой работы можно сделать с помощью программы, а затем оставшиеся файлы придется соотносить вручную.
- Найти оригиналы текстов для тех файлов, содержание которых не удалось соотнести с Описью Архива.
- Снять ксерокопию с этих текстов и пополнить архив ГД.
- Внести описание этих текстов в Опись Архива.
- Открыть доступ к архиву ПЩ и приступить к его описанию по соответствующему формуляру. Для этого нужен очень аккуратный исполнитель. (Оценить объем работы можно только начав эту работу.)
- Начать сличение описей архивов ГД и ПЩ. После сличения пополнить Опись Архива.
- Начать процесс сертификации файлов. (Я умею это делать с помощью менее квалифицированного помощника со скоростью 15 стр. в час, что составляет 0.2 \$ за страницу.)
- Определить, какие тексты из архива ГД должны быть введены в компьютер в первую очередь.
- Составить Перечень безусловных требований к исполнителям, осуществляющим пополнение Архива.

- Найти подходящего кандидата на место "курьера", отвечающего за циркуляцию текстов, за связь с библиотеками, за ксерокопирование текстов и т.п.
- Начать ввод текстов с помощью планшетного сканера и системы распознавания *Fine Reader*. Я полагаю, что на этой фазе нет нужды в аренде помещения. Считаю более целесообразной установку компьютера и сканера непосредственно у исполнителя (если их нельзя арендовать, то покупать нужно, на мой взгляд, не суперсервер — вопрос о его покупке (как и вопрос об аренде помещения) должен решаться лишь на фазе проектирования развития — и не суперсканер, а "стандартный" пентиум (или 486) со средней постоянной и повышенной оперативной памятью (1200 \$) и HP ScanJet (500 \$). (Я умею сканировать тексты из архива ГД — как правило, это плохие копии — со скоростью примерно 5 стр. в час. Скорость небольшая, но зато нет нужды в последующей сертификации. И я не стал бы сходу отвергать такое медленное сканирование без сопоставительных оценок. Стоимость страницы в этом случае составит 0.4 \$.) На данной фазе для сканерного ввода вполне достаточно двух человек. При этом работу второго обеспечу я — на моем компьютере, который желательнее усилить 4 Мб оперативной памяти (150 \$). В этом варианте на сканерный ввод потребуется 400 \$ в месяц (1000 стр.).
- Начать ввод опубликованных текстов с помощью сканера с оригиналов. Оригиналы должен поставлять "курьер". Скорость ввода этих текстов раза в два выше (но не любых). Особое место среди этих текстов занимают тексты на иностранном языке. На них нужно обратить специальное внимание (имея в виду *Internet*).
- Начать ввод текстов путем набивки, тщательно отобрав исполнителей. Стоимость страницы текста должна быть специально определена (предположительно это будет 0.4 \$.). Стандартной страницей, чтобы избежать разночтений, предлагаю считать страницу, содержащую 1600 *characters*. На данной фазе объем вводимой информации должен быть небольшим — не более, чем на 400 \$ в месяц (1000 стр.).
- Начать разработку технологии пополнения Библиотеки схем стандартными и нестандартными схемами. Для этого прежде всего нужно "дотянуть" схемы "Избранных трудов", а затем на их базе создать "конструктор".
- Найти кандидата на место "менеджера-контролера", отвечающего за качество работы всех исполнителей. Этот кандидат должен свято верить, что **архивная работа и халтура — вещи несовместные**. А это означает, что любые выявленные проявления халтуры наказываются не штрафом, а немедленным разрывом контракта. Может оказаться, что места "курьера" и "менеджера-контролера" целесообразно занимать одному человеку.

Затраты на этой фазе определяются, с одной стороны, бюджетом Фонда, с другой — временем, необходимым для завершения всех оценок, выбора и отработки технологий: фаза запуска по минимуму заканчивается одновременно с окончанием фазы отработки и оценки технологий.

4.3. Фаза отработки и оценки технологий

Для реализации этой фазы потребуется специальный человек. Длительность фазы во многом будет определяться его профессионализмом.

4.4. Фаза проектирования развития

Считаю принципиально неверным вложение больших средств — как в "железки", так и в СУБД — до того, как полностью доопределится ситуация. Не следует действовать по схеме: "думать некогда — доставать надо". Такой большой проект, как Архив, требует постепенного развертывания. Поэтому и на фазе проектирования развития речь должна идти только о первом приближении, или, другими словами, только о "мини-архиве". Суперсерверы и другие супержелезки имеет смысл обсуждать, на мой взгляд, лишь на втором витке проекта — при втором приближении к цели.