

Эволюция систем оцифровки печатных изданий на примере их использования в проекте ЭБ «Научное наследие России».

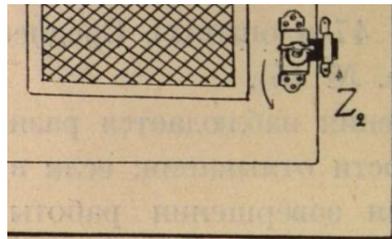
Кириллов Сергей Александрович
Электронная библиотека «Научное наследие России»

Первый этап. 2005-2007 гг. Ввод комплекса сканирования на основе монохромного планетарного сканера MINOLTA PS7000 в связке с цветным планшетным сканером Epson GT15000.



сторождений, графиты и различного рода применяемые для обработки металла, разнообразные строительные материалы и час начинают вовлекаться в строительные материалы и опоки, прекрасные гипсы, кварцевые и разнообразные шлаки металлургического прои

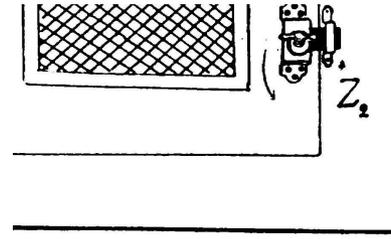
Стакан с чаем



Комбинация из двух за-
верток Z_1 и Z_2 .

обезьяна р
плохо: она
мыкания ни
(Z_2), трогает
потом зубам
контролирует
обращается
вертке (Z_1).

), с шумом полуотмыкая последнюю
ла опыта; далее обезьяна опять контр
трогает полуотомкнутую завертку Z_2
окончательно, опять трогает зубами
на секунду к верхней завертке, затем с
то замыкает, то отмыкает; к верхн
прикоснется и опять оставляет его, по
к той и к другой завертке, опери
только короткий срок. После 9 ми
обезьяна все еще возится с отомкнут
 Z_2 , не прилагая достаточного усили
хней Z_1 . Только после моего указ
вертку она обращается к ней более



Комбинация из двух за-
верток Z_1 и Z_2 .

обезьяна р
плохо: она
мыкания ни
(Z_2), трогает
потом зубам
контролирует
обращается
вертке (Z_1).

), с шумом полуотмыкая последнюю
ла опыта; далее обезьяна опять контр
трогает полуотомкнутую завертку Z_2
окончательно, опять трогает зубами
на секунду к верхней завертке, затем о
то замыкает, то отмыкает; к верхн
прикоснется и опять оставляет его, по
к той и к другой завертке, опери
только короткий срок. После 9 ми
безьяна все еще возится с отомкнут
 Z_2 , не прилагая достаточного усили
хней Z_1 . Только после моего указ
вертку она обращается к ней более

Открывание двери.

чей было установление у обезьяны двигательного-
жду актом выхода из клетки и поощрением.
тейших случаев.

льной клетки закрывается плотно вдвигающейся
пыте обезьяна засаживается внутрь клетки, и
о конца, вплотную.

ь, просовывает в нее голову и руку (табл. VII,
исовывается все больше и больше наружу, про-
туловище вытесняя вбок, по мере своего про-
зную в пазах вдоль рамы доску (табл. VII,

смотря на то, что доска вполне задвигается, не
ета, обезьяна легко справляется с отодвиганием,
другие приемы.

тся обеими руками в доску, привстает в верти-
епокло упершись ногами в пол, двигает доску ру-
оска не поддается сразу, обезьяна упирает ноги
чтобы усилить давление рук, применяя при оття-
двигая доску вбок.

обезьяна находится вне клетки и если доска не
кусывает зубами край доски (табл. VII, фиг. 1) —
ий, и едва увидит раскрывшуюся щель, как ухва-
рукой и уже более уверенно отводит доску в сто-

Открывание двери.

чей было установление у обезьяны двигательного-
жду актом выхода из клетки и поощрением.
тейших случаев.

льной клетки закрывается плотно вдвигающейся
пыте обезьяна засаживается внутрь клетки, и
о конца, вплотную.

ь, просовывает в нее голову и руку (табл. VII,
исовывается все больше и больше наружу, про-
туловище вытесняя вбок, по мере своего про-
зную в пазах вдоль рамы доску (табл. VII,

смотря на то, что доска вполне задвигается, не
ета, обезьяна легко справляется с отодвиганием,
другие приемы.

тся обеими руками в доску, привстает в верти-
епокло упершись ногами в пол, двигает доску ру-
оска не поддается сразу, обезьяна упирает ноги
чтобы усилить давление рук, применяя при оття-
двигая доску вбок.

обезьяна находится вне клетки и если доска не
кусывает зубами край доски (табл. VII, фиг. 1) —
ий, и едва увидит раскрывшуюся щель, как ухва-
рукой и уже более уверенно отводит доску в сто-

чей 'место', чоң 'душа'; звук к — глухой смычный, образуется путём смыкания самой задней части спинки языка с мягким нёбом; катор 'ряд', кишлак 'кишлак'. Смещение звуков к и г может привести к искажению значения слова (ср., например: қанд 'сахар' и қанда 'он сорвал', қадар 'количество' и қадар 'печаль'). Звук г — звонкий шелевой, образуется путём сближения самой задней части языка с мягким нёбом; гайр 'прочий', бог 'сад' богайрат 'энергичный'. Звук х — глухой шелевой, образуется путём сближения стенок глотки (фаринкса) в нижней её части. Звук х напоминает звук һ в немецком языке в словах habe, heute. Смещение звука х с русским х может привести к искажению значения слова (ср., например: хам 'также' и хам 'наклонённый', хар 'каждый' и хар 'осёл').

Звук х несколько отличается от соответствующего русского звука х, так как образуется глубже (так же, как и звук г, но без участия голоса).

Звук л на слух воспринимается как средний между русским твёрдым л и мягким ль: лола 'тюльпан', баланд 'высокий'. При образовании таджикского л передняя часть спинки языка смыкается с твёрдым нёбом, а края языка опущены.

Звонкие согласные б, д, г в отличие от соответствующих русских согласных в конце слова оглушаются незначительно: буд 'был' (а не бут), раг 'жила' (а не рак).

Остальные согласные таджикского языка не имеют существенных отличий от соответствующих звуков русского языка.

Ударение

В таджикском языке в отличие от русского ударение закреплённое и обычно падает на последний слог слова: хона́ 'дом', ола́м 'человек', талаба́ 'учащийся', хонахо́ 'дома́', корга́р 'рабочий', давлатма́нд 'богатый'.

В словах, заимствованных из русского и других языков, ударение может падать на разные слоги, например: тра́ктор, инсти́тут.

Ударение в таджикском языке не падает на следующие морфологические показатели, служебные частицы и слова, примыкающие к предшествующему слову и пишущиеся с ним слитно: 1) изафет: гу́ли сурх 'красный цветок', макта́би миёна́ 'средняя школа'; 2) суффикс -е, выражающий неопределённость: ола́ме 'какой-то человек'; 3) послелог -ро ман ин китобро́ хондам 'я прочитал эту книгу'; 4) местоименные суффиксы -ам, ат, -ан, -амон, -атон, -амон: китобам 'моя книга', китобхо́ам 'мои книги'. В сочетании с местоимением худ 'сам', худам 'я сам', на местоименные суффиксы: худаг 'ты сам'; ударение падает; 5) формы глагола-связки -ам, -и, -аст, -ем, -ед, -анд: ман талаба́ам 'я ученик', ту талаба́и 'ты ученик', онхо талабагонанд 'они ученики'; 6) соединительный союз -у // -ю: китобу дафта́р китба́ и тетради́, оча́ю бача́ 'мать и ребёнок'.

Ударение не падает также на соединительный союз ҳам 'также', 'тоже': ман ҳам меравам я то́же пойду́; на подчинительный союз ки 'что': У гуфт, ки имру́з дарс мешавад он сказал, что сегодня занятия состоятся, и на вопросительную частицу -ми: падарат меояд-ми? 'придёт ли твой отец?'

В личных глагольных формах с приставкой ме — главное ударение падает на последний слог формы, а добавочное — на приставку: меравам. добавочное — на последний слог формы: бихона́д — пусть он читает.

Частица на-, выражающая отрицание при глаголах, принимает основное ударение, добавочное падает на последний слог глагольной формы: намегира́м 'я не возьму'. Глаголы простого прошедшего времени имеют параллельное произношение либо с ударением на основе, либо с ударением на окончании: рафта́м || рафта́м 'я пошёл'. В русских и интернациональных словах, вошедших в словарный состав таджикского языка, а также в именах собственных и географических названиях обычно сохраняется ударение, принятое в русском языке: па́ртия, респу́блика, авиаци́я; Ле́нин, Давра́на, Во́ра, А́нна; Во́лга, Но́вгород.

В словах заимствованных из русского языка, давно вошедших в разговорную речь, ударение падает на последний слог, например: самово́р, карто́шка.

МОРФОЛОГИЯ

Части речи

По основному значению, грамматическим формам и роли в предложении слова в таджикском языке, как и в русском, делятся на две основные группы: 1) самостоятельные или знаменательные слова, которые служат в предложении его членами, и 2) служебные слова, которые показывают отношения между самостоятельными словами или отдельными предложениями и не являются членами предложения. К служебным словам относятся предлоги, союзы, частицы. Некоторые служебные слова, например, частицы, могут придавать предложениям различные оттенки. Самостоятельные слова по значению, синтаксическим и морфологическим признакам делятся на следующие разряды: именные части речи — имя существительное, прилагательное, числительное, местоимение — и глаголы, наречие, междометие.

Имена существительные, прилагательные, числительные и местоимения в отличие от русского языка не имеют грамматического рода, падежей и соответствующих окончаний.

В таджикском языке именные части речи имеют следующие грамматические показатели: а) множественного числа (суффиксы -он, -хо); б) единичности и неопределённости (суффикс -е). Для связи именных частей речи в предложении используются изафет, предлоги и послелоги.

Однако не все перечисленные грамматические показатели и средства свойственны всем именным частям речи в равной степени. Прилагательные, числительные и большинство местоимений не принимают показателя множественного числа, например: девори баланд 'высокая стена', деворҳои баланд 'высокие стены' или хона 'этот дом', или хонахо 'эти дома'. Большинство местоимений не принимают изафета и т. п.

Глаголы сравнительно с глаголами русского языка отличаются

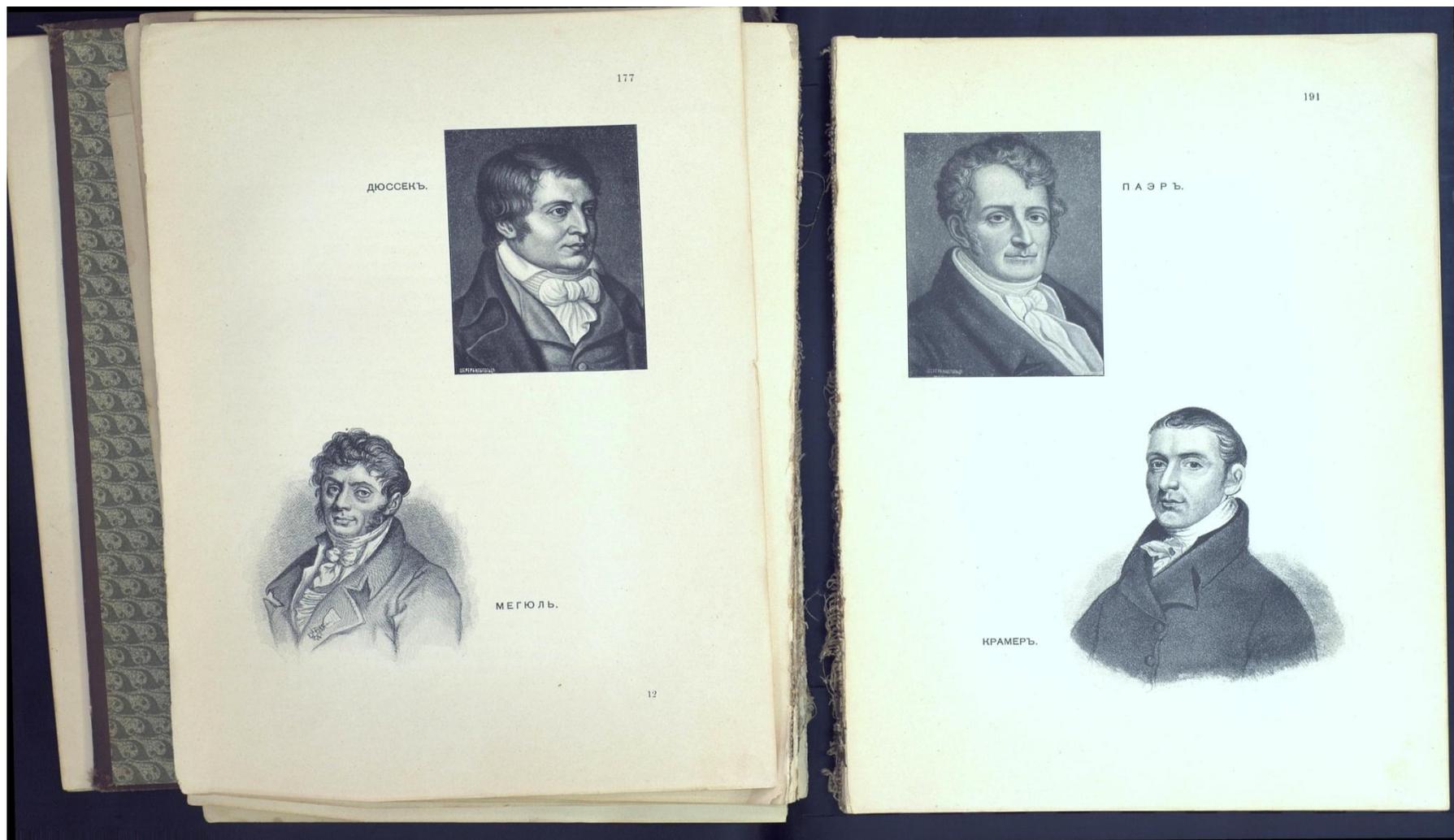
Цветной планшетный сканер Epson GT-15000

- При оцифровке старых библиотечных фондов часто возникают ситуации, с которыми сканеры, использующие алгоритмы бинаризации, справиться не могут или справляются неудовлетворительно.
- Для решения этих проблем в комплекс сканирования был включен цветной планшетный сканер Epson GT-15000



Недостатки планшетного сканера Epson GT-15000.

Недостатки Epson GT-15000 свойственны всем планшетным сканерам: это небольшая площадь оцифровки (до формата А3) и плохая эргономика, которая приводит к механическому повреждению книги оператором.



Второй этап. Модернизация сканера Minolta PS7000. 2007-2008 г.

- Одним из решений, повышающим как сохранность книг, так и производительность стала модернизация сканера Minolta PS7000 модулем PWB-J, который позволяет сохранять изображения в градациях серого.
- Однако цветные изображения все равно приходилось сканировать на Epson GT-15000.
- Кроме этого все чаще стала обсуждаться тема безопасного освещения при сканировании.

Проблема безопасного освещения при сканировании. Рекомендации НИЦКД РГБ.

- «Можно констатировать, что сканирование документов не безвредный процесс для бумажной основы документов» - делают осторожный вывод сотрудники НИЦКД РГБ
- «Оцифровывание следует рассматривать, как стратегию сохранности документов, посредством увеличения доступа к нему в другом формате. Это означает, что цифровая копия должна быть такого качества, чтобы не было необходимости изготавливать повторную копию для нужд микрофильмирования или факсимильного переиздания документов».
- В качестве практической рекомендации НИЦКД РГБ предлагает ограничить оцифровку каждого библиотечного документа «не более одного–трех раз»

Третий этап. 2008-2010 гг. Внедрение цветных широкоформатных книжных сканеров.

- Для продолжения работ по формированию ЭБ «Научное наследие России» потребовалось новое оборудование, позволяющее:
- а) сканировать книги с минимальным освещением;
- б) проводить цветное сканирование книг;
- в) сканировать ветхие издания, не допускающие разворот на 180 градусов;
- г) сканировать материалы большого формата.

Комплекс планетарного цветного сканирования ЭЛАР ПланСкан С2-ЦА-600



Комплекс цветного сканирования ЭЛАР ПауэрСкан Д14000.



Проблемы сканирования при минимальном уровне освещенности.

- **Опыт эксплуатации показал, что сканирование с минимальным освещением требует выделения специально оборудованных помещений, в которых можно поддерживать одинаковый уровень освещенности (не менее 300 Люкс) в течение всей рабочей смены.**
- **Поскольку выделить отдельное помещение для каждого сканера не представляется возможным, в 2011 году будет проводиться, в частности, плановая модернизация систем подсветки.**

Эволюция систем оцифровки в проекте ЭБ «Научное наследие России»

- **2005-2007 гг.** Ввод комплекса сканирования на основе монохромного планетарного сканера MINOLTA PS7000 в связке с цветным планшетным сканером Epson GP1500.
- **2007-2008 гг.** Для повышения качества и безопасности сканирования проводилась модернизация сканера MINOLTA PS7000, в частности, модулем PWB-J, позволившая производить сканирование печатных материалов, как в черно-белом изображении, так и в градациях серого.
- **2008-2010 гг.** Начался переход на цветные широкоформатные книжные сканеры. В качестве базового варианта был принят комплекс цветного сканирования на основе сканера ПланСкан С2-ЦА-600. Для сканирования печатных материалов особо больших размеров до формата А0 был внедрен один комплекс цветного сканирования на основе сканера Пауэрскан Д14000 А0.

Переход на цветные широкоформатные книжные сканеры позволил:

улучшить качество сканирования;

снизить трудоемкость выполнения работ по получению сканов с различных типов печатных изданий, т.е. мы получили производственную схему, по которой одну книгу сканирует один оператор на одном сканере;

повысить безопасность сканирования за счет исключения механического повреждения оригиналов при сканировании на планшетном сканере и физического повреждения за счет снижения мощности освещения сканируемого оригинала.

Эволюция систем оцифровки в проекте ЭБ «Научное наследие России»

- **2011 г. и планы на будущее.** Главная задача настоящего периода - это постепенный переход на цветные широкоформатные книжные сканеры. Базовым вариантом остается комплекс цветного сканирования на основе сканеров серии ПланСкан С2-ЦА. Увеличивается количество сканеров, производится модернизация отдельных элементов комплекса, разрабатываются новые алгоритмы работы оператора, изменяется технология сканирования. Для решения задач мобильного сканирования среднеформатных изданий будет введен комплекс цветного сканирования ПланСкан А2-ВС-В.

Иллюзия изменений градаций серого цвета создавалась методом разбивки изображения в зависимости от его тональности на различные по размеру, но однородные по цвету, штриховые и точечные элементы.

