Что не так с текстологией и как исправить ситуацию?

Towarzystwo cybernetyczne

## Часть I. Пояснение. Зачем текстология?

*Towarzystwo cybernetyczne*

[Источник](http://propaganda-journal.net/10390.html).

Текстология это, как известно, прикладная технико-историческая дисциплина, нацеленная на критическую выработку авторского варианта текста по доступным источникам. Обращаться к этой дисциплине в этой серии публикаций мы будем не из академического интереса, а из политического. Если по своей сути ликвидации капитализма, имеющего мировое господство, будущая социальная революция имеет мировой характер, то и формирование её теоретических предпосылок не может носить местный характер. Эти теоретические предпосылки безусловно включают в себя самостоятельное и критическое мышление (ну нужно трёх слов) на лучшем достигнутом уровне. В предреволюционное время такое мышление должно оказаться достаточным для понимания общественной ситуации и для выработки победной политической линии массовых освободительных движений. Пока же таких движений нет и практика не спешит стихийно подсказывать принципы оздоровления общественной жизни, собственно теоретическая работа должна быть намного более чуткой. Важность международного теоретического взаимодействия и взаимного знакомства с главными теоретическими источниками возрастает в нашу гнилую контрреволюционную эпоху. Это возрастание важности международного взаимодействия и международного взаимодействия несомненно каждому, кто знает состояние сообществ практического материализма сейчас. Увеличение важности такого совместного (пусть даже по-преимуществу теоретического) действия несомненно относительно, скажем, 1960-х годов, когда страны народной демократии имели мощное хозяйство или 1980-х годов, когда инерция оздоровления ещё не была до конца преодолена, когда в пользу освобождения от частной собственности ещё «высказывались» практические силы, а не только разум, как сейчас. В эпоху спада всякого освободительного движения и углубления товарности во всех сферах общественной жизни особенно важным становится выработать теоретические инструменты революционизирования практики, чтобы не пропустить стихийные освободительные движения и открыть им их собственный теоретический и политический разум. Это возможно сделать только в тесной связи с мировой теоретической и политической историей. В эпоху спада освободительной активности не местные, а лишь лучшие мировые образцы могут связать всякого местного деятеля с победной для революции политической и организационной линией. Поэтому задачи теоретической подготовки мировой социальной революции необходимо включают образование связи всякого местного деятеля с лучшими мировыми теоретическими достижениями.

Подобная постановка вопроса, при ближайшем рассмотрении, сразу же наталкивается на языковой вопрос. Он имеет материальную основу в том самом факте, что капиталистическая действительность мало кому оставляет время для изучения языков иных местностей и вдумчивого изучения теоретической литературы на них. Более того, в большинстве сфер своей жизни, человечество до сих пор не имеет универсального языкового средства. Естественно, что на роль этого средства непригодны национальные языки, выработавшиеся из иных потребностей. И также естественно, что мировая языковая универсализация невозможна на стихийной основе международного товарного хозяйства и господства финансового капитала транснациональных корпораций (ТНК). Но задачи теоретического взаимодействия нужно решать уже сейчас. Ни наиболее распространённая языковая запись (китайско-японско-вьетнамско-корейская иероглифическая система), ни наиболее распылённая литературная норма (английская) для этого не подходят. Истина не решается голосованием, и потому в деле международной совместной теоретической работы никакого преимущества не будет ни за носителями иероглифической системы, не понятной соседям, ни за носителями английской литературной нормы, которая за пределами Австралии, Британии и США фактически мертва и служит только финансово-административным целям в ТНК. Никакая задача специального конструирования будущего в данном случае неуместна. Человечеству ещё предстоит стать единым в сознании подобно тому как оно стало единым в хозяйстве. Но поскольку совместная теоретическая работа нужно уже сейчас, то следует получше присмотреться к её задачам. Формирование теоретического мышления на лучшем достигнутом уровне означает, в первую голову, известность этого уровня. Эта известность может быть сформулирована как максимально лёгкое и полное приближение к источникам, демонстрирующим лучший теоретический уровень, раскрывающим принцип формирования убеждений, соответствующих законам материального мира. С точки зрения того, чего не хватает для международного теоретического взаимодействия в том случае, если негде начинают работать самообразовательные сообщества, то программа общей нужды сводиться, первоначально, к двум пунктам

**1. Облегчение критических переводов вплоть до быстрой проверки перевода по оригинальному тексту**

**2. Облегчение критики источников на языке оригинала вплоть до сравнения с подлинником первой публикации** [[1]](#footnote-21)

Почему к российскому читателю? Об основах новой текстологической работы.

Данная статья, являющаяся инструкцией-размышлением обращена к российским товарищам не случайно. Во-первых Россия в настоящий момент имеет наиболее быстро растущее сообщество теоретического и политического самообразования. По косвенным признакам можно понять, что оно ежемесячно вербует новых членов больше, чем составляют все коммунистические круги балтских стран[[2]](#footnote-22). Понятно, что стремление смыть позор безмыслия и поддержки империалистической политики в 2014 году теперь в России столь прочно, что породило самообразовательное сообщество. Это очень важная часть покаяния за восторги о крымской аннексии. К чести российских товарищей и в пользу их дальновидности нужно сказать, что начало самообразования произошло ещё до того, как российская буржуазия экспроприировала пенсионные права трудящихся России.

Если российское теоретическое и политическое самообразование достаточно укрепилось в стремлении снять с российского общества позор безмыслия (а впоследствии и позор бездействия), то в этой своей задаче оно обязано выйти на всемирный простор и сделать для себя известными теоретические сокровища других стран. Российская литературная норма является, наряду с немецкой и кастильской одним из трёх мировых языков теоретического мышления. Всюду, где мы встречаемся с полноправно оцененными лучшими образцами мировой теоретической литературы, мы встречаемся с тремя этими самоназваниями: Deutsch, русский, español. Три этих языка, хранящих наиболее концентрированные произведения революционной мысли, неизбежно должны быть освоены в ходе современного международного теоретического взаимодействия. Именно имея их ввиду, нужно для международного теоретического взаимодействия, для упрощения переводов использовать некую вспомогательную систему, которую в отношении специальной лексики проще, чем язык другого мог бы освоить и латиноамериканец, и немец, и великоросс. Система библиографических справок и аннотаций должна быть примерно равно сложной и равно удалённой от этих языковых норм, чтобы при содействии широкому охвату порождать также желание «заглянуть в соседнюю хату». Полагаем, что необходимость такого занятия в столь пострадавшей от местечкового мышления эпохе должна быть очевидной. Под требование начала диалога и межсубъектности между носителями немецкой, кастильской и российской литературной нормы вполне подходит в настоящее время язык эсперанто, автор которого в своё время проштудировал специальные филологические книги, в названия которых встречалось и „Deutsch“, и «русский», и”español".

Остаётся понять чем российское сообщество может быть полезно товарищам в других странах. Ответ может показаться весьма необычным, но главной и важнейшей международной обязанностью российского теоретического сообщества является обеспечение максимальной доступности ленинских работ. Российский революционный опыт наиболее концентрированно оказался выражен в работах Ленина, Чернышевского и Ильенкова. При этом только значительная часть работ Ильенкова доступна в систематическом хранилище в виде простого гипертекста с авторской разметкой. Что касается работ Чернышевского, то их нормализованные гипертексты создавал единственный текстолог. Планомерным и систематическим созданием гипертекста ленинских работ в России, как недавно выяснилось, никто не занимался. В результате существует несколько файловых комплектов, имитирующих пятое издание полного собрания ленинских сочинений. Однако они не имеют текстологической оценки, согласованного плана работ и представляют из себя дублированные материалы неизвестного (местами очень низкого) текстологического и оформительского качества. Создание нескольких независимых комплектов гипертекстов ленинских работ само по себе поглотило огромное количество труда, большая часть которого была растрачена весьма непроизводительно. Без установки единых текстологических критериев и разделения участков любая работа подобного масштаба обречена на сектанство, застой и угасание. Марксизм всегда подозрительно относился к индивидуальной ненужной жертвенности и фанатическому героизму. В данном случае обобществлённая скоординированная работа могла бы решить проблему создания эталонных гипертекстов ленинских произведений.

Летом 2014 года в Берлине встретились на совещании с местными текстологами представители Латинской Америки, США, Франции, Польши и некоторых других стран. Быстрое развитие коммунистического движения в США уже тогда поставило вопрос о критическом издании работ Ленина на английском языке. По замыслу американских товарищей каждый абзац английского гипертекста должен был быть связан с соответствующим абзацем ленинского оригинала, так что к его просмотру приводило простое нажатие указателя в области английского текста. Об источниках для переводов ленинских работ говорил также представитель Латинской Америки, выучивший немецкий язык хуже товарища из США. К сожалению, на том совещании выяснилось неприятное обстоятельство. Если немецкие товарищи помогли с авторскими текстами работ Маркса и Энгельса на языках оригинала (многолетний текстологический проект MEGA), то никаких материалов, пригодных для использования в качестве оригинального слоя переводных ленинских работ товарищи выявить не смогли.

Современные текстологические требования в техническом отношении столь просты, что доступны для понимания и для исполнения ученику основной школы. В отношении ленинских работ эти требования сводятся в общем к классическому для XIX века перечню. В частности гипертекст должен передавать все авторские особенности. Например, если автор использовал специальные листы рукописи, где разграфка явно связана с размерами листа (Как у Маркса в рукописях 1844 года), то гипертекст должен имитировать рамкой панно рукописного листа. Но в большинстве случаев страницы авторской рукописи не должны отмечаться в гипертексте. Также гипертекст по общему правилу не должен показывать страницы авторитетного издания, однако место смены страниц подлежит какой-то разметке. Так, место разрыва страниц 5-го издания ленинских работ безусловно стоит разметить, но совершенно глупо делать это вставляя соответствующий номер в нужное место и, тем более, прерывая абзац. В общем случае нормализованный гипертекст не должен включать никакой разметки, кроме авторских абзацев, авторских курсивов, авторских подчёркиваний, авторских разрядок, авторских надчёркиваний и пр. То есть в 2014 году в Берлине искали всего-навсего гипертексты, несущие только особенности авторского текста и аккуратные указания на смену страниц. За исключением тома XXIX, Ленин применял в большей части работ курсив, разрядку и подчёркивания. Этот «огромный арсенал текстологических приёмов», типичный для типографики своего времени в гипертексте может набрать теперь любой школьник. Неизвестно что было сложного для российских товарищей в том, чтобы создать хотя бы один полный эталонный (по пунктуации и орфографии) вариант гипертекстов ленинских работ. Но с работой требующей внимания и систематичности они до сих пор не справились, что является показателем состояния российского коммунизма. Повторим: сплошного набора простейших гипертекстов с подтверждёнными качественным литературным текстом без буквенных ошибок, и с авторской пунктуацией для ленинского письменного наследия в настоящий момент **не существует**. Фактически это означает, что серьёзный переводчик должен ориентироваться на бумажное издание. А достать бумажное российское издание для товарищей из Америки очень непросто. Объективные ограничения по трудозатратам для товарищей из США и Латинской Америки даже в случае доступности бумажной книги не обеспечивали создания распознанного нормализованного гипертекста. Одно дело переводчик, нацеленный на некоторую ленинскую работу, а другое дело массовая орфографическая выверка, явно требующая носителя языка. Несмотря на то, что через Латвию американские товарищи смогли получить нужные бумажные тома ленинских работ, едва ли такая постановка дела может быть признана удовлетворительной. Она одновременно затратна и антидемократична. Переправить 55 томов 5-го издания ленинских работ на языке оригинала в Буэнос-Айрес или Сан-Франциско одновременно дорого, трудно и не очень плодотворно. Тем не менее спустя 50 лет после формулирования принципов бузбумажной документальной работы, благодаря отставанию российского теоретического сообщества, это остаётся единственным надёжным способом работы с ленинским письменным наследием. Да и в целом отапливаемое место для 55 книг является для товарищей во многих странах проблемой, не говоря о том, что печатное издание не позволяет осуществлять быстрый и полный поиск фрагментов.

Действительными преимуществами в форме печатного издания обладают очень немногие ленинские работы, что вообще типично для классического теоретического наследия. Есть даже версия, что представление о преимуществах бумажной формы некоторых работ является чисто переходным и полуиллюзорным для нашего переходного времени. Некоторые товарищи говорят, что люди, не заставшие господство бумажных книг, определят им более узкое значение, чем кажется сейчас. Может быть, оно так и будет, подобно тому, как люди эпохи массовой фотографии стали намного выше ценить живопись. Но даже признание относительного удобства бумажных изданий для отдельных работ не означает, что бумажная форма является предпочтительной, а тем более необходимой в текстологической работе. Невозможность быстрой произвольной критической проверки перевода, быстрого поиска и необходимость хранить объёмные и тяжёлые издания это недостаток, препятствующий широкому хождению ленинских работ там, где к ним есть интерес. Это фактор угасания этого интереса под давлением обстоятельств чисто технического, а не политического или общественного плана. Но что же противопоставить хранению бумажных изданий? Попробуем выработать желательный способ обращения с ленинским письменным наследием используя хорошо понятную всем канцелярскую аллегорию.

Итак, типичная форма обращения с ленинскими работами состоит в манипуляции с книжным шкафом. Если необходима проверка переводов то к книжному шкафу издания на польском языке добавляется книжный шкаф с оригиналами. Проверка перевода в таких условиях превращается в работу с тремя раскрытыми книгами, где по страницам двух скользят пальцы, а третья является словарём. Занятие это малопродуктивное и физически утомительное. Как ускорить его?

Попробуем выкупить два комплекта польских и российских изданий ленинских работ. Из 220 типовых обложек все кроме 4 уничтожаются. Также уничтожаются переплёты, а с каждой страницы (или с соседних страниц) вырезаются на специальные карточки абзацы. Таким образом 55 томов превращаются в длинный ящик с карточками абзацев. Два таких ящика, поставленные рядом позволят последовательно сверить переводы. Заглядывать в словарь имея абзацные карточки действительно проще. Этот приём использовался переводчиками в Народной Польше, где оригинал немецкой книги мог распределяться разным переводчикам по абзацам после совещания о соблюдении единой терминологии в переводе.

Каковы проблемы использования абзацной картотеки? В общем случае абзацы должны быть пронумерованы и каким-то образом должен быть помечен тип абзаца - цитата, авторский текст, подпись под иллюстрацией, заголовок главы, заголовок работы и прочее. Каждая карточка абзаца должна содержать упоминание работы, в состав которой он выходит. Для общей работы лучше всего создать отдельную книгу учёта с указанием порядка абзацных карточек - какая после какой следует. Тогда абзацные карточки включая вставки и варианты, имеющие необычные номера, можно будет раздавать разным переводчикам и/или читателям не боясь нарушить порядок картотеки.

Отдельной книги учёта заслуживает размещение абзацев на страницах. Перечисление всех страниц всех 55 томов необходимо требует списка номеров абзацев для каждой страницы. Наконец, особый журнал учёта содержит переводческие соответствия с номером оригинального и номером переводного абзаца.

Получившаяся комнатка с длинными ящиками, несколькими толстыми журналами соответствий (и, разумеется, обслуживающим это канцеляристом) не является чем-то вдохновляющим - думает, должно быть читатель. «Что же современного в запахе книжной пыли и спорах обитающих в бумаге грибков?» Читательница может тоже представляет что-то подобное, но всё же не забывает, что это всего лишь канцелярская аллегория, которую никто не собирается реализовывать канцелярскими средствами.

Исчерпывающие методологические указания на действительную реализацию широкого и критического доступа к ленинскому письменному наследию можно выработать размышляя над книгой выдающегося советского кибернетика Анатолия Ивановича Китова[[3]](#footnote-23). Уже в те годы он уловил главное преимущество безбумажной формы как для массивов чисел, так и для массивов текстов (текстовое направление использования вычислительных машин было тогда новым). Это преимущество заключается в быстроте поиска чего-либо среди однородных элементов. Так, просмотр нескольких миллионов предложений текста на наличие контрольного слова сейчас при частоте процессора около 1 гигагерца и кодировке Уникод занимает десяток секунд. Если же тексты были предварительно разобраны на слова и был выстроен перечень ускорения, то поиск нужного слова сокращается до нескольких тысячных долей секунды.

**Быстрота поиска определяет все остальные свойства безбумажной формы любых сведений.** В частности, установление соответствий по общему признаку между разными наборами сведений уже более 30 лет является наиболее распространённой операцией в реляционных базах данных. Если в нашей канцелярской аллегории поиск страницы по абзацной карточке требовал прохода взглядом по соответствующей книге учёта, то в безбумажном виде соответствие абзацам при известном номере страницы и томе осуществляется примерно столь же быстро как поиск номера страницы для известного абзаца. В общем случае поиск соответствий *прост в обе стороны*. Для таблиц с миллионами записей, которым добавлены перечни ускорений такой поиск соответствия в любую сторону занимает всё те же тысячные доли секунды.

Использование баз данных в текстологии основывается на двух логических предпосылках:

1. Преимущество демократической обобществлённой работы над индивидуальной. *База данных используется по определению как специфическая программа для обобществления сведений.*

2. Текстология как формальное познание проходит те же этапы, что и содержательное познание. *Сначала происходит расчленение произведения на легко обозримые части, а потом его сборка без потерь из проверенных фрагментов.* Это отдалённо повторяет разбор ситуации в понятия и последующую выработку адекватной практической линии на основании верного соединения этих понятий.

Разберёмся далее как реализуются эти предпосылки.

## Часть II. Практикум

*Towarzystwo cybernetyczne*

[Источник](http://propaganda-journal.net/10394.html).

Той программой, которая должна облегчить текстологическую работу является управляющая программа базы данных, выступающая как своеобразный секретарь по хранению перечней, называемых таблицами. Общественной функцией управляющей программы баз данных является обобществление сведений и поддержание информационной централизации. Принцип информационной централизации в работах того же Китова расшифровывается как единственный способ обращения к некоторым сведениям, притом такой, что в результате обязательно оказывается нечто актуальное и достоверное.

От наивной реализации принципа работы с безбумажными сведениями до современных программ управления реляционными базами данных человечество прошло длинный путь. Важнейшие технические факторы преодоления частной собственности были раскрыты в работах Виктора Глушкова по хозяйственной регуляции и в математических трактатах [Эдгара Кодда](https://de.wikipedia.org/wiki/Edgar_F._Codd), которые открывают совершенно новые технические возможности при соединении с принципами Глушкова. Для российских текстологов эта специальная литература, имеющаяся в варшавских библиотеках пожалуй будет излишней. Популярная литература по реляционным базам данных тоже. Вообще бытовое сознание по поводу баз данных обладает уровнем профессионального понимания берутовской Польши. То есть сознание польских учёных, занимавшихся массивами данных и картотеками в середине 1950-х годов сейчас с некоторыми упрощениями воспроизводится каждым городским жителем любой из европейских стран, где существовали режимы народной демократии. Разумеется, ключом к этому стихийному умозрительному схватыванию является более развитая сознательная форма - понимание количественно господствующих уже как несколько десятилетий реляционных баз данных. Поэтому интереснейшую историю становления принципов обобществления сведений придётся в данном случае проигнорировать.

### Подготовка к практикуму

В настоящий момент любому желающему доступные бесплатные, хорошо проверенные и выносливые программы управления реляционными базами данных. Среди них выделяются SQLite (польск. [скули́тэ]) и PostgreSQL (польск. [постгрэ́скуль], чаще просто [по́стгрэс], тогда произношение совпадает с немецким, фламандским и чешским).

Установка этих программ довольно проста. В Ubuntu и других популярных операционных системах на основе Debian нужно набрать в поиске приложений для установки «sqlite3» или «postgresql» либо от имени администратора набрать

apt-get install sqlite3

или

apt-get install postgresql

В менее популярных операционных системах на основе Fedora или RedHat пакеты с близкими названиями нужно устанавливать от имени администратора через команду yum. Про установку на MacOS сообщить что либо затруднительно, ибо квалифицированных пользователей этой операционной системы рядом не нашлось.

Диалог с программами управления реляционными базами данными ведётся в текстовом виде. Текстовый диалог является универсальным и к нему сводится почти всё обычное взаимодействие с программой, управляющей базой данных[[4]](#footnote-27). Текст диалога составляется по особым формальным правилам, близким к одному из стандартов языка SQL. Существуют программы, которые позволяют вести диалог с базой данных не набирая текст, однако в конечном счёте они тоже формируют и посылают текст, хотя и не демонстрируют его. *Умение обращения с базой данных сводится к пониманию того, что она содержит и как выразить свои потребности в виде текста SQL*. Программа управления базой данных при своей работе выполняет функции секретаря и канцеляриста, откликаясь на специально составленные SQL-фразы.

Практикум с базами данными необходимо начать с подготовки программы, которая будет отвечать. Работу с SQLite можно начать после введения в терминале операционной системы

sqkite3 nazwa.sqlite;

В этом случае база данных будет размещена в единственном файле «nazwa.sqlite», который будет размещён в текущем каталоге (обычно это каталог пользователя). Упоминая каждый раз этот файл, можно сохранять преемственность (накопительность) обращений к базе данных. Первой командой для SQLite должна быть

PRAGMA foreign\_keys = ON;

Эта команда позднее позволит использовать ссылки между таблицами.

Аналогичную работу можно проводить на сайте https://kripken.github.io/sql.js/GUI/ , где все команды можно указывать одну за одной вместо написанных там в центре.

Для подготовки к работе программы PostgreSQL нужно выполнить несколько шагов. Первым шагом нужно создать пользователя, который будет работать с базами данных, а вторым шагом нужно создать базу данных, где этот пользователь будет полновластным хозяином. Имя пользователя для простоты работы должно совпадать и именем пользователя в операционной системе. Так, если в системе работает пользователь wojcech, то обычная настройка PostgreSQL, отождествляющая пользователей операционной системы и базы данных, требует, чтобы от имени высочайшего пользователя postgres была дана команда

CREATE USER "wojcech" LOGIN;

Эту команду нужно подавать от имени администратора операционный системы, представившись как «postgres», что можно сделать через

sudo -u postgres psql;

Команда создания базы данных будет выглядеть примерно так

CREATE DATABASE "Tekstujo internacia" OWNER "wojcech" ENCODING UTF8 LC\_COLLATE pl\_PL.UTF8;

Данная команда создаст базу данных “Tekstujo internacia”, что на эсперанто означает международное собрание текстов. Будет назначена международная многоалфавитная кодировка UTF8, позволяющая не задумываться о допустимых символах. Сортировка строк будет осуществляться по польскому алфавиту, что указано как «pl\_PL.UTF8».

Для будущей работы с UUID[[5]](#footnote-28) нужно также набрать команду

CREATE EXTENSION "uuid-ossp";

Программный блок “uuid-ossp” содержит несколько небольших программ и поставляется чаще всего вместе с PostgreSQL.

После создания, пользователя, базы данных и группы функций для UUID нужно вернуться в обычный пользовательский режим и запустить там терминал, где набрать

psql -d "Tekstujo internacia" -U "wojcech";

Результатом этой команды будет начало общения с конкретной базой данной в PostgreSQL. Это общение происходит по очень похожим правилам, что и общение с SQLite. Правила эти называются языком формальным SQL. Он произошёл к концу 1980-х годов от более ранних разработок программистов, которые были основаны на английских литературных фразах. В настоящее время эта связь ослабла как формально, так и содержательно. Товаришка, академически специализирующаяся на английской литературе и прошедшая самообразовательный курс SQL в Германии утверждала, что знание SQL и литературного английского языка (или даже просто английской лексики) это весьма отдалённые навыки. Их объединение «имеет смысл не более чем чтение оригиналов проповедей Экхарта для лучшего освоения лексики Брехта». Или, для восточного читателя, изучать английскую лексику для лучшего понимания SQL это всё равно что читать Кирилла Туровского для лучшего освоения лексики из работ Ленина. Практика освоения SQL и практика освоения литературной английской лексики это принципиально разные практики. Потому аргументом против работы с SQL не может служить отсутствие знания английской лексики, а само это знание не может быть значительным облегчением освоения SQL. С лексической стороны современный SQL представляет из себя мешанину романской и германской лексики, применение которой в обращениях для разных задач определяется строгими правилами. Различные SQL ситуации подлежат запоминанию. Для этого лучше всего завести записной листок или файл, где все команды будут записываться по мере освоения с упоминанием той потребности, когда они нужны. Полноценную актуальную документацию для PostgreSQL или SQLite на польском языке довольно сложно найти. По-настоящему выручает только самообразовательный курс немецких коммунистов, медленная наработка опыта да консультации с замечательным комплектом документации https://docs.postgresql.fr/11/ с которым познакомил приехавший в Берлин французский товарищ.

### Картотеки-таблицы

Понятие, наследующее канцелярское понятие картотеки это таблица, а понятие наследующее понятие карточки это запись или строка. В канцелярском деле карточке нередко предшествует печатная форма, создающая линии для вписывания. В работе табличных или реляционных баз данных появлению таблицы предшествует определение колонок, свойств или атрибутов (параллельная терминология).

Для международной базы данных естественно иметь список языков. Облегчение выбора языка обычно опирается на классификацию по языковым семьям. Простой понумерованный список языковых семей будет создаваться такой командой

CREATE TABLE "Lingva familioj" (  
"Kodo" int NOT NULL,  
"Nomo" varchar(32) NOT NULL  
);

В этой команде таблица (картотека) под названием «Языковые семьи» “Lingva familioj” объявляется в составе Кода “Kodo” и Имени “Nomo”. Вслед за названиями свойств приписаны их типы. Названия этих типов, некогда связанные с английскими названиями типовых полей бумажных карточек в настоящий момент очень сильно отделились как от следов канцелярского происхождения так и от какой-либо лексической нормативности, связанной с английским языком. Полная таблица стандартных типов для PostgreSQL находится по адресу https://docs.postgresql.fr/11/datatype.html , а SQLite проверку типов не производит. Некоторые основные типы присутствуют во всех распространённых табличных базах данных. Эти типы лучше всего запомнить

* int - целые числа в пределах нескольких миллиардов. Обычно используются для нумерации и указания количества.
* char(число) - знаковая группа строго из указанного количества символов. Обычно используется для разнообразных кодов. Так, например, польский почтовый индекс всегда имеет одинаковое количество символов.
* varchar(число) - знаковая группа, не длиннее указанного количества символов. Например, аннотация к картине должна уместиться в несколько строк путеводителя, где есть место только для некоторого числа знаков.
* real - дробное число.
* text - описание неограниченной длины.
* bool - да/нет. В просторечии называется флагом, поскольку важнейшие состояния произносятся как «поднят» (да) и «опущен» (нет).
* timestamptz - дата и время, такие, что всякому будет показано его собственное местное время в данный момент. Этот тип используется чтобы понять что было раньше, а что было позже и чтобы не заниматься работой с часовыми поясами.
* uuid - некое специально созданное значение, которое не должно повторяться никогда и которое имеет единственное предназначение - представлять свою строку (карточку), не будучи с ним никак связано по смыслу. Uuid это *идеальный представитель*, несущий в себе всеобщие свойства иного, но не имеющий их внутри себя. В этой роли uuid похож на номер с той разницей, что нумерация у разных карточек, описывающих разные вещи может повторяться, а правильно созданный uuid едва ли когда-либо совпадёт не только в рамках одной таблицы (картотеки), но и вообще в практике человечества.

Указанных типов вполне достаточно для того, чтобы хранить классические тексты так, как это было описано в схеме Фроша-Загорского-Радкевичюте.

Остаётся без объяснения что значит «NOT NULL» после названия и типа каждого свойства. Это указание на то, должны ли указанные свойства заполняться. В нашем случае оба свойства должны заполняться, или языковая семья без номера или языковая семья без названия это нечто бессмысленное. Если же свойство может оставаться без заполнения, то на месте «NOT NULL» не пишется ничего и сразу ставится «,» начинающая описание следующего свойства. Когда в табличных базах данных нечто не заполнено, то считается, что там находится некое специальное значение NULL, которое не является ни числом, ни датой, ни временем, ни флагом ни чем ещё другим. На польском языке есть название для NULL. Оно близко к слову «inconnu» из французской документации. Это полужаргонное выражение «нельзя знать», отсылающее к армейским порядкам периода тройного ига над Польшей. Похожими немецкими словами обычно выражают в популярной литературе смысл стройной математической концепции Эдгара Кодда, стоявшей за появлением этого необычного понятия, отражающего не заполненное поле. Пока что будет нужно подчеркнуть лишь, что для числового свойства NULL это не 0, для символьно-знакового свойства NULL это не отсутствие текста, а для флага NULL отличается от той ситуации, когда «квадратик отметок не заполнен», то есть когда bool свойство хранит «нет».

### Как заполняются карточки?

Созданная таблица “Lingva familioj” ничего не хранит. Её польза возможна лишь после внесения сведений. В нашем случае команда записывается так

INSERT INTO "Lingva familioj" ("Kodo","Nomo") VALUES (1, 'Nigera-Konga');

Данная команда указывает на добавление сведений (INSERT INTO) именно в указанную таблицу (“Lingva familioj”). Далее в скобках перечисляются свойства, которые будут заполняться. Их порядок важен, ибо после «VALUES» в скобках в точно таком же порядке должны быть написаны сведения для размещения. В примере свойство “Kodo” записывается единицей, а свойство “Nomo” текстом ‘Nigera-Konga’ (на эсперанто это указывает на Нигеро-конголезскую языковую семью). Попытка сменить порядок вызовет ошибку

INSERT INTO "Lingva familioj" ("Nomo", "Kodo") VALUES (1, 'Nigera-Konga');

Целое число запишется в строку свойства “Nomo”, а вот из строки ‘Nigera-Konga’ никак нельзя сделать целое число, чтобы оно стало номером.

Добавление сведений о других языковых семьях производится аналогично:

INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (2, 'Dravidana');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (3, 'Kreola');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (4, 'Turka');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (5, 'Aŭstroasia');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (6, 'Tupia');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (7, 'Nordokcidenta Kaŭkaza');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (8, 'Izolita lingvoj');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (9, 'Sina-Tibeta');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (10, 'Hindeŭropa');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (11, 'Nila-Sahara');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (12, 'Keĉua');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (13, 'Afra-Asia');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (14, 'Tehnika');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (15, 'Suda Kaŭkaza');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (16, 'Eskimo-Aleŭta');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (17, 'Mongola');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (18, 'Nordorienta Kaŭkaza');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (19, 'Japona');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (20, 'Ajmara');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (21, 'Korea');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (22, 'Dené-Eniseia');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (23, 'Aŭstronesia');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (24, 'Tai-Kadai');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (25, 'Uralika');  
INSERT INTO "Lingva familioj" ("Kodo", "Nomo") VALUES (26, 'Algonkia');

При работе с базой данных можно ожидать, что часто встречаться будут такие коды языковой семьи как 10 (индоевропейские), 14 (искусственные, к ним относится эсперанто), 25 (уральские, у ним относятся венгерский и финский языки), 4 (тюркские).

Код языковой семьи, хранящийся в нашей таблице позволяет однозначно получить название языковой семьи. Этот код будет указывать на языковую семью всюду, где нужно будет это указание. Поскольку наш код, являющийся условным номером всегда заполнен, то есть не может быть неизвестен (NULL) и никогда не повторяется, то его можно использовать как представителя всей карточки, которая в нашем случае содержит только название. Убыстрить поиск по номеру и указать на представительскую роль Кода в нашем случае можно назначив его *первичным ключом*. В математической теории по первичному ключу можно однозначно выйти на единственную запись-строку или убедиться, что такой строки нет. Знание ключа означает возможность знать остальное содержимое подобно номеру автомобиля или коду PESEL, которые выводят на конкретную личность.

ALTER TABLE "Lingva familioj" ADD CONSTRAINT PRIMARY KEY ( "Kodo"); -- (Данная команда не применима в SQLite)

Как именно используется первичный ключ и зачем он нужен станет лучше понятно когда таблица перестанет быть одинокой.

### Карточки языков

CREATE TABLE "Lingvoj" (  
"Nomo originala" text NOT NULL,  
"Nomo" text,  
"Familio" int NOT NULL,  
"ISO 639-1" char(2),  
"ISO 639-3" char(3) NOT NULL  
);

Итак, типовая карточка языка включает в себя обязательное самоназвание (“Nomo originala”). Как и следующее свойство оно объявлено как текст без ограничения длины. Вторым свойством является название на эсперанто (“Nomo”), однако оно сделано не обязательным из тех соображений, что пока что не собраны названия всех приводимых ниже языков. Разумеется, в комплектной базе данных с интенсивной коллективной работой это поле должно быть заполнено для всех языков и помечено обязательным, однако зафиксированная ситуация должна остаться напоминанием о скверном состоянии международного взаимодействия.

Свойство “Familio” содержит целое число, которое рассматривается как код уже знакомой языковой семьи. Не было ли смысла хранить здесь всего лишь текстовое название языковой семьи? Зачем нужно было вводить коды, выполняющие функции идеального представителя? В трактатах Эдгара Кодда можно найти следующие доводы в пользу создание независимой таблицы и перемещения кода:

1. Языковая семья есть нечто отличное от языка. Её свойства принципиально отличны от свойств отдельного языка.

2. Имя языковой семьи при хранении в перечне языков будет повторено и не поддастся простому переименованию в одном месте, нужно будет проверять все случаи, а не один.

3. Добавление некоего свойства языковой семьи в карточку языка (не только названия, а, например, распространённости) прямо сделает хранение этого сведения дублирующимся, ибо это свойство будет продублировано у языков одной и той же семьи.

4. Дублирование кода как идеального представителя не создаёт проблем изменения свойств языковой семьи, ибо код не несёт этих свойств и нет никакого смысла менять код, если что-то сменилось в карточке языковой семьи.

Нужно специально оговориться, что существование кодов предполагает возможность их осмысления. Ибо идеальный представитель, не актуализируемый в деятельности является чисто материальным бесполезным предметом, подобно закопанным античным скульптурам. Разумеется актуализация кодов является весьма утомительной, однако в современных базах данных есть развитые средства, перекладывающие на себя расшифровку разнообразных кодов. Они будут упомянуты чуть ниже.

Свойство “ISO 639-1” в карточке языка представляет из себя двухбуквенный код из соответствующего международного стандарта. Эти коды должны быть известны читателю: de, pl, uk, es, ru, lv, fr, eo. Двухбуквенный код есть не у всех языков, имеющих трёхбуквенный код, поэтому это необязательное поле. Собственно, обязательный трёхбуквенный код всегда размещён в свойстве “ISO 639-3”. Эти коды, возможно, менее популярны, но рекомендованы международной организацией по стандартам. В данном случае её лучше послушаться и привыкать к кодам типа deu, pol, rus, ukr.

В PostgreSQL будет весьма полезно объявить трёхбуквенные коды опознающий свойством, то есть обязательным и выводящим на единственную запись/строку/карточку.

ALTER TABLE "Lingva familioj" ADD CONSTRAINT "Lingvoj\_pkey" PRIMARY KEY ("ISO 639-3"); -- (Данная команда не применима в SQLite)

После этой команды ввести несколько строк с повторяющимся трёхбуквенным кодом будет невозможно. Попытка сделать это будет приводить к сообщению об ошибке. В SQLite существует подобный механизм, но его несколько сложнее привести в действие. Командный порядок «ALTER TABLE … ADD CONSTRAINT» в SQLite не работает[[6]](#footnote-32).

Кроме единственной записи с языковым кодом нужно гарантировать, что код языковой семьи формально правильный, то есть относится к списку кодов языковых семей. В данном случае формальной фразой связаны таблица “Lingvoj”, её поле “Familio” с таблицей “Lingva familioj” и её полем “Kodo”. Команда выглядит так:

ALTER TABLE "Lingvoj" ADD CONSTRAINT "Lingvoj\_Lingva\_familioj\_FK" FOREIGN KEY ("Familio") REFERENCES "Lingva familioj"("Kodo"); -- \*(Данная команда не применима в SQLite)\*

Поскольку командный порядок «ALTER TABLE … ADD CONSTRAINT» в SQLite не работает, для этой программы существует косвенный способ реализации ссылок[[7]](#footnote-33).

После создания предпосылок добавление языковых карточек можно сделать уже известным способом

INSERT INTO "Lingvoj" ("Nomo originala","Nomo","Familio","ISO 639-1","ISO 639-3") VALUES ('Esperanto','Esperanto',14,'eo','epo');

Однако есть способ упрощения массовой вставки. После директивы «VALUES» можно через запятую указывать несколько групп в скобках, где каждая группа превратится во вставленную строку/запись/карточку. Порядок в любом случае должен совпадать с тем, который указан в скобках до директивы «VALUES»:

INSERT INTO "Lingvoj" ("Nomo originala","Nomo","Familio","ISO 639-1","ISO 639-3") VALUES  
('Deutsch','Germana',10,'de','deu')  
,('Русский','Rusa',10,'ru','rus')  
,('Español','Hispana',10,'es','spa')  
,('Українська','Ukraina',10,'uk','ukr')  
,('język polski, polszczyzna','Polona',10,'pl','pol')  
,('latine, lingua latina','Latina',10,'la','lat');

Полный перечень языков, близкий к перечню Международной Организации по стандартам помещается в приложение.

### Абзацы — основа системы текстологических работ

Причины по которым абзацы (таблица абзацев) были выбраны основой хранилища классических текстов не стоит здесь обсуждать. Это было сделано при обосновании базы данных по схеме Фроша-Загорского-Радкевичюте.

Таблица абзацев в первом приближении выглядит так

CREATE TABLE "Paragrafoj" (  
"Kodo" uuid NOT NULL PRIMARY KEY,  
"Laboro" uuid NOT NULL,  
"Etikedo" varchar(32) NOT NULL,  
"Klaso" varchar(64) NULL,  
"Interno" text NULL  
);

Абзац имеет собственный код(“Kodo”), принадлежность к работе (код другой таблицы в “Laboro”), маркировку для гипертекста “Etikedo”, класс для гипертекста “Klaso” и, главное, текстовое содержание “Interno”. Пока что принципиально важно, что без каталога работ абзацы хранить бессмысленно. Следовательно нужно подготовить каталог работ:

CREATE TABLE "Laboroj" (  
"Kodo" uuid NOT NULL PRIMARY KEY,  
"Nomo originalo - lingvo" char(3) NOT NULL,  
"Publikacio" date NOT NULL,  
FOREIGN KEY("Nomo originalo - lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);

Устроен каталог работ просто, он состоит из кода работы, кода языка на котором было дано название и даты публикации. Отдельно уже при создании каталога работ предусмотрено, что язык можно выбрать только из списка языков. Директива

FOREIGN KEY("Nomo originalo - lingvo") REFERENCES "Lingvoj"("ISO 639-3")

расшифровывается так, что свойство “Nomo originalo - lingvo” должно иметь только то значение, которое уже хранится в таблице “Lingvoj” и именно в её свойстве “ISO 639-3”.

В карточке работы специально не содержится никаких названий. Принцип равноправия языков требует лишь указания языка оригинального названия. Но и оно указывается в общем перечне в другой таблице

CREATE TABLE "Laboroj - nomoj" (  
"Laboro" uuid NOT NULL,  
"Lingvo" char(3) NOT NULL,  
"Nomo" text NOT NULL  
);

В каталоге имён письменных произведений указывается код произведения, трёхбуквенный код языка и само название на указанном языке.

В системе международного общения произведения письменности получают переводы. Для элементарного примера нужно иметь возможность выразить происхождение перевода

CREATE TABLE "Laboroj - originaloj" (  
"Laboro" uuid NOT NULL,  
"Originalo" uuid NOT NULL,  
"Lingvo" char(3) NOT NULL,  
FOREIGN KEY("Lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);

Эта таблица содержит код работы(“Laboro”), код её языка“Lingvo” и код оригинальной работы “Originalo”. С появлением таблицы “Laboroj - originaloj” можно создать пример международного текстологического хранилища, ибо появилась минимальная замкнутая система таблиц[[8]](#footnote-35).

### Пример заполнения системы таблиц

В SQLite[[9]](#footnote-37)

INSERT INTO "Laboroj" ("Kodo", "Nomo originalo - lingvo","Publikacio") VALUES (lower(hex(randomblob(16)))), 'fra', '1888');

В PostgreSQL

INSERT INTO "Laboroj" ("Kodo", "Nomo originalo - lingvo","Publikacio") VALUES (uuid\_generate\_v4(), 'fra', '1888-01-01');

Эти команды заполнили код произведения письменности. Нам нужно будет знать его для указания названия и переводов, для упоминания в каждом абзаце произведения. Узнать наш код может команда

SELECT "Kodo", "Nomo originalo - lingvo","Publikacio" FROM "Laboroj";

Эта команда выведет всё содержимое всех свойств из таблицы работ. На одной из строк (в нашем случае на единственной) будет нужный код. Допустим, он равен 488097c8-3f9c-4ecf-9d1d-64701ab9764c , тогда читатель должен будет применить свою непонятную последовательность вместо упоминаний «488097c8-3f9c-4ecf-9d1d-64701ab9764c».

Вносим переводы и оригинальное название (апостроф внутри цитаты удваивается, чтобы не быть концом цитаты).

INSERT INTO "Laboroj - nomoj" ("Laboro", "Lingvo", "Nomo") VALUES  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'fra', 'L''Internationale'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'epo', 'La Internacio (himno)'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'deu', 'Die Internationale'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'rus', 'Интернационал (гимн)'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'spa', 'La Internacional'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'ukr', 'Інтернаціонал (гімн)'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'pol', 'Międzynarodówka (pieśń)'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'bel', 'Інтэрнацыянал (гімн)'),  
('488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'zho', '国际歌');

В международной текстологической базе данных разумный порядок наименования включает обязательные наименования на языке оригинала и на эсперанто, рекомендуемые наименования на мировых языках теоретического мышления (читатель мог заметить последовательность ‘deu’ - ‘rus’ - ‘spa’), а затем дополнительные наименования на доступных читателю и других известных языках, например, языках ООН.

Создадим польский перевод, опубликованный в 1920 году.

В SQLite

INSERT INTO "Laboroj" ("Kodo", "Nomo originalo - lingvo","Publikacio") VALUES (lower(hex(randomblob(16)))), 'pol', '1920');

В PostgreSQL

INSERT INTO "Laboroj" ("Kodo", "Nomo originalo - lingvo","Publikacio") VALUES (uuid\_generate\_v4(), 'pol', '1920');

Пускай эта команда SELECT “Kodo”, “Nomo originalo - lingvo”,“Publikacio” FROM “Laboroj”; показала нам новый код c01122df-18e4-4a78-a446-fbf7b8f2949b, тогда можно добавить точно такую же систему переводов, хотя в общем случае перевод может впервые появиться под названием, которое будет переводится на язык оригинала иначе, чем оригинальное название и, следовательно, подобного повтора не будет.

INSERT INTO "Laboroj - nomoj" ("Laboro", "Lingvo", "Nomo") VALUES  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'pol', 'Międzynarodówka (pieśń)'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'epo', 'La Internacio (himno)'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'deu', 'Die Internationale'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'rus', 'Интернационал (гимн)'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'spa', 'La Internacional'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'ukr', 'Інтернаціонал (гімн)'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'bel', 'Інтэрнацыянал (гімн)'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'fra', 'L''Internationale'),  
('c01122df-18e4-4a78-a446-fbf7b8f2949b', 'zho', '国际歌');

В этой команде для подчёркивания регулярного порядка первой была указана карточка польского названия, однако для базы данных это не важно, ибо она хранит все языковые названия наравне.

Польский текст является переводом французского

INSERT INTO "Laboroj - originaloj" ("Laboro", "Lingvo", "Originalo") VALUES ('c01122df-18e4-4a78-a446-fbf7b8f2949b, 'pol', '488097c8-3f9c-4ecf-9d1d-64701ab9764c');

Как видим, подобные сложные указания формируются довольно лаконично. Когда у нас есть два указания на письменные произведения, уместно заполнить их содержимое.

Вместо \*\*\* здесь и далее нужно подставить функцию создающую UUID: для SQLite lower(hex(randomblob(16)))), а для PostgreSQL uuid\_generate\_v4().

INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, '488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'p', NULL, 'Debout ! l''âme du prolétaire  
Travailleurs, groupons-nous enfin.  
Debout ! les damnés de la terre !  
Debout ! les forçats de la faim !');

В данном случае как абзац во французский текст «Интернационала» внесён обычный гипертекстовый абзац (‘p’) без особого класса оформления (NULL), состоящий из четырёх стихотворных строк, разделённых гипертекстовым знаком переноса строки. Апострофы внутри текста удвоены, чтобы программа управления базой данных не считала это место концом цитаты.

Аналогично добавляется другой абзац:

INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, '488097c8-3f9c-4ecf-9d1d-64701ab9764c', 'p', NULL, 'Pour vaincre la misère et l''ombre  
Foule esclave, debout ! debout !  
C''est nous le droit, c''est nous le nombre :  
Nous qui n''étions rien, soyons tout :');

Совершенно подобным образом добавляется польский текст

INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, 'c01122df-18e4-4a78-a446-fbf7b8f2949b', 'p', NULL, 'Wyklęty, powstań ludu ziemi!  
Powstańcie, których dręczy głód!  
Myśl nowa blaski promiennemi  
Dziś wiedzie nas na bój, na trud.');  
  
INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, 'c01122df-18e4-4a78-a446-fbf7b8f2949b', 'p', NULL, 'Przeszłości ślad dłoń nasza zmiata;  
Przed ciosem niechaj tyran drży!  
Ruszymy z posad bryłę świata,  
Dziś niczem - jutro wszystkiem my!');

### Связь абзацев и указание переводов — основа международного общения

В созданной системе хранения абзацев каждый отнесён к произведению письменности, а это произведение может оказаться переводом оригинала (или другого перевода, такое тоже бывает). Однако принципиально не это, а то, что порядок следования абзацев никак не указывается. Точно также абзацы перевода никак не связаны с абзацами оригинала. Для этих целей создаются две таблицы.

1. Таблица связей абзацев. “Antaŭa” - предшествующий, “Sekva” - последующий

CREATE TABLE "Paragrafoj - ĉenoj" (  
"Antaŭa" uuid NOT NULL,  
"Sekva" uuid NOT NULL  
);

2. Таблица переводов абзацев, полностью аналогичная таблице переводов работ.

CREATE TABLE "Paragrafoj - originaloj" (  
"Paragrafo" uuid NOT NULL,  
"Lingvo" char(3) NOT NULL,  
"Originalo" uuid NOT NULL  
);

### Инструкция-упражнение

С помощью команды SELECT “Kodo”, “Laboro”, “Etikedo”, “Klaso”, “Interno” FROM “Paragrafoj”; выводится весь каталог абзацев.

1. В таблицу “Paragrafoj - ĉenoj” добавляется указание на то, что второй французский абзац следует за первым (коды размещаются в “Antaŭa” и предшествующий, “Sekva”).

2. В таблицу “Paragrafoj - originaloj” вносится код первого польского абзаца с указанием на язык перевода и с указанием первого французского абзаца как оригинала.

3. В таблицу “Paragrafoj - originaloj” вносится код второго польского абзаца с указанием на язык перевода и с указанием второго французского абзаца как оригинала.

### Выводы в конце

Совместная международная текстологическая база данных должна поощрить минимальное знание иных языков, необходимое для переводов названий и отождествления абзацев. Других средств решения задачи активизации международного теоретического общения до сих пор не было предложено в противовес нашему предложению. Разве может быть это общение организовано иначе, если оно рассматривается как систематическое и планомерное международное теоретическое взаимодействие. Думается, что в доступности оригинальных и подвергнутых выверке (текстологической критике) произведений заинтересованы в освободительном движении все. На этапе сообществ самообразования суть ведущейся работы наилучшим образом располагает к основательной текстологической проработке. Формирование агитационных и политических органов освободительного движения автоматически ставит иные задачи и оттесняет текстологические работы по обеспеченности кадрами. Однако успешная политическая работа принципиально невозможна, если ещё на этапе самообразования не было заложено знакомство с мировой теоретической литературой. Местные идеологии, тем более патриотической направленности или изоляционистские ещё нигде не приводили к хозяйственному и культурному процветанию и надёжному ходу от товарного хозяйства с принципами господства и подчинения к нетоварному хозяйству с принципом сотрудничества. Обращение к российскому читателю обусловлено в первую голову тем, что более никто не сможет формировать эталонную выверенную базу данных по ленинскому письменному наследию. Это обусловлено как тем, что в других странах нет должного количества носителей языка Ленина так и тем, что критическая выверка по лучшим известным изданиям не может быть делом узкой группы людей. Так отсутствие оригинальной абзацной структуры ленинских работ тормозит (хотя и не останавливает) поабзацную привязку немецкого и румынского варианта некоторых текстов. Подготовка эталонного поабзацного гипертекстового издания ленинских работ рассматривается как почётная обязанность российского самообразовательного сообщества. Ибо тем самым в нём во-первых будут выработаны здоровые силы, неспособные восторгаться империалистической политикой и во-вторых тем самым товарищам из других стран будет облегчена возможность сопоставить свои переводы ленинских работ с оригиналом с перспективной новых переводов или критических исследований уже существующих переводов.

### Послесловие

Переводческий коллектив имеет разрешение на перевод и публикацию других частей конспекта немецких самообразовательных курсов по базам данных, адаптированного к раскрытию смысла базы данных по схеме Фроша-Загорского-Радкевичюте. Поэтому ***мы запрашиваем отклик заинтересованных читателей о том нужно ли продолжение, в какой форме оно должно быть и каковы недостатки настоящего изложения.***

Практическое значение описанной базы данных состоит в принципиальной возможности[[10]](#footnote-42) создания лучших выверенных и многоязычных эталонных текстов классических письменных произведений, в повышении языковой культуры в обход популярных транзитных языков и в улучшении взаимной известности теоретических работ представителей разных народов.

### Вместо эпилога

(команды требуют подстановки функции, создающей UUID)

INSERT INTO "Laboroj" ("Kodo", "Nomo originalo - lingvo","Publikacio") VALUES ('62e75f87-978b-4d9e-b587-bedcc2d23598', 'rus', NULL);  
  
INSERT INTO "Laboroj - nomoj" ("Laboro", "Lingvo", "Nomo") VALUES  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'rus', 'Интернационал (гимн)'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'epo', 'La Internacio (himno)'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'deu', 'Die Internationale'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'spa', 'La Internacional'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'fra', 'L''Internationale'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'ukr', 'Інтернаціонал (гімн)'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'pol', 'Międzynarodówka (pieśń)'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'bel', 'Інтэрнацыянал (гімн)'),  
('62e75f87-978b-4d9e-b587-bedcc2d23598', 'zho', '国际歌');  
  
INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, '62e75f87-978b-4d9e-b587-bedcc2d23598', 'p', NULL, 'Вставай, проклятьем заклеймённый,  
Голодный, угнетённый люд!  
Наш разум - кратер раскалённый,  
Потоки лавы мир зальют.');  
  
INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, '62e75f87-978b-4d9e-b587-bedcc2d23598', 'p', NULL, 'Сбивая прошлого оковы,  
Рабы восстанут, а затем  
Мир будет изменён в основе:  
Теперь ничто - мы станем всем!');  
  
INSERT INTO "Laboroj" ("Kodo", "Nomo originalo - lingvo","Publikacio") VALUES ('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'rus', NULL);  
  
INSERT INTO "Laboroj - nomoj" ("Laboro", "Lingvo", "Nomo") VALUES  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'ukr', 'Інтернаціонал (гімн)'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'epo', 'La Internacio (himno)'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'deu', 'Die Internationale'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'rus', 'Интернационал (гимн)'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'spa', 'La Internacional'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'fra', 'L''Internationale'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'pol', 'Międzynarodówka (pieśń)'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'bel', 'Інтэрнацыянал (гімн)'),  
('30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'zho', '国际歌');  
  
INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, '30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'p', NULL, 'Повстаньте, гнані і голодні  
Робітники усіх країв,  
Як y вулкановій безодні  
B серцях y нас клекоче гнів.');  
  
INSERT INTO "Paragrafoj" ("Kodo", "Laboro", "Etikedo", "Klaso", "Interno") VALUES (\*\*\*, '30ee70fa-2ff2-4ff5-b8ef-f99378272909', 'p', NULL, 'Ми всіх катів зітрем на порох  
Повстань же, військо злидарів  
Bce, що забрав наш лютий ворог  
Щоб повернути, час наспів.');

### Приложение 1

**Языковой справочник, близкий к ISO 639,** за исключением некоторых языков, использовавшихся в примерах ранее.

INSERT INTO "Lingvoj" ("Nomo originala","Nomo","Familio","ISO 639-1","ISO 639-3") VALUES  
('Afaraf',NULL,13,'aa','aar')  
,('аҧсуа бызшәа, аҧсшәа',NULL,7,'ab','abk')  
,('Afrikaans',NULL,10,'af','afr')  
,('Akan',NULL,1,'ak','aka')  
,('አማርኛ',NULL,13,'am','amh')  
,('العربية',NULL,13,'ar','ara')  
,('aragonés',NULL,10,'an','arg')  
,('অসমীয়া',NULL,10,'as','asm')  
,('авар мацӀ, магӀарул мацӀ',NULL,18,'av','ava')  
,('avesta',NULL,10,'ae','ave')  
,('aymar aru',NULL,20,'ay','aym')  
,('azərbaycan dili',NULL,4,'az','aze')  
,('башҡорт теле',NULL,4,'ba','bak')  
,('bamanankan',NULL,1,'bm','bam')  
,('беларуская мова',NULL,10,'be','bel')  
,('বাংলা',NULL,10,'bn','ben')  
,('Bislama',NULL,3,'bi','bis')  
,('བོད་ཡིག',NULL,9,'bo','bod')  
,('bosanski jezik',NULL,10,'bs','bos')  
,('brezhoneg',NULL,10,'br','bre')  
,('български език',NULL,10,'bg','bul')  
,('català, valencià',NULL,10,'ca','cat')  
,('čeština, český jazyk',NULL,10,'cs','ces')  
,('Chamoru',NULL,23,'ch','cha')  
,('нохчийн мотт',NULL,18,'ce','che')  
,('ѩзыкъ словѣньскъ',NULL,10,'cu','chu')  
,('чӑваш чӗлхи',NULL,4,'cv','chv')  
,('Kernewek',NULL,10,'kw','cor')  
,('corsu, lingua corsa',NULL,10,'co','cos')  
,('ᓀᐦᐃᔭᐍᐏᐣ',NULL,26,'cr','cre')  
,('Cymraeg',NULL,10,'cy','cym')  
,('dansk',NULL,10,'da','dan')  
,('ދިވެހި',NULL,10,'dv','div')  
,('རྫོང་ཁ',NULL,9,'dz','dzo')  
,('eesti, eesti keel',NULL,25,'et','est')  
,('euskara, euskera',NULL,8,'eu','eus')  
,('Eʋegbe',NULL,1,'ee','ewe')  
,('føroyskt',NULL,10,'fo','fao')  
,('فارسی',NULL,10,'fa','fas')  
,('vosa Vakaviti',NULL,23,'fj','fij')  
,('suomi, suomen kieli',NULL,25,'fi','fin')  
,('Frysk',NULL,10,'fy','fry')  
,('Fulfulde, Pulaar, Pular',NULL,1,'ff','ful')  
,('Gàidhlig',NULL,10,'gd','gla')  
,('Gaeilge',NULL,10,'ga','gle')  
,('Galego',NULL,10,'gl','glg')  
,('Gaelg, Gailck',NULL,10,'gv','glv')  
,('Avañe''ẽ',NULL,6,'gn','grn')  
,('ગુજરાતી',NULL,10,'gu','guj')  
,('Kreyòl ayisyen',NULL,3,'ht','hat')  
,('(Hausa) هَوُسَ',NULL,13,'ha','hau')  
,('עברית',NULL,13,'he','heb')  
,('Otjiherero',NULL,1,'hz','her')  
,('हिन्दी, हिंदी',NULL,10,'hi','hin')  
,('Hiri Motu',NULL,23,'ho','hmo')  
,('hrvatski jezik',NULL,10,'hr','hrv')  
,('magyar',NULL,25,'hu','hun')  
,('Հայերեն',NULL,10,'hy','hye')  
,('Asụsụ Igbo',NULL,1,'ig','ibo')  
,('Ido',NULL,14,'io','ido')  
,('ꆈꌠ꒿ Nuosuhxop',NULL,9,'ii','iii')  
,('ᐃᓄᒃᑎᑐᑦ',NULL,16,'iu','iku')  
,('Interlingue/Occidental',NULL,14,'ie','ile')  
,('Interlingua',NULL,14,'ia','ina')  
,('Bahasa Indonesia',NULL,23,'id','ind')  
,('Iñupiaq, Iñupiatun',NULL,16,'ik','ipk')  
,('Íslenska',NULL,10,'is','isl')  
,('Italiano',NULL,10,'it','ita')  
,('ꦧꦱꦗꦮ, Basa Jawa',NULL,23,'jv','jav')  
,('日本語 (にほんご)',NULL,19,'ja','jpn')  
,('kalaallisut, kalaallit oqaasii',NULL,16,'kl','kal')  
,('ಕನ್ನಡ',NULL,2,'kn','kan')  
,('कश्मीरी, كشميري‎',NULL,10,'ks','kas')  
,('ქართული',NULL,15,'ka','kat')  
,('Kanuri',NULL,11,'kr','kau')  
,('қазақ тілі',NULL,4,'kk','kaz')  
,('ខ្មែរ, ខេមរភាសា, ភាសាខ្មែរ',NULL,5,'km','khm')  
,('Gĩkũyũ',NULL,1,'ki','kik')  
,('Ikinyarwanda',NULL,1,'rw','kin')  
,('Кыргызча, Кыргыз тили',NULL,4,'ky','kir')  
,('коми кыв',NULL,25,'kv','kom')  
,('Kikongo',NULL,1,'kg','kon')  
,('한국어',NULL,21,'ko','kor')  
,('Kuanyama',NULL,1,'kj','kua')  
,('Kurdî, کوردی‎',NULL,10,'ku','kur')  
,('ພາສາລາວ',NULL,24,'lo','lao')  
,('latviešu valoda',NULL,10,'lv','lav')  
,('Limburgs',NULL,10,'li','lim')  
,('Lingála',NULL,1,'ln','lin')  
,('lietuvių kalba',NULL,10,'lt','lit')  
,('Lëtzebuergesch',NULL,10,'lb','ltz')  
,('Kiluba',NULL,1,'lu','lub')  
,('Luganda',NULL,1,'lg','lug')  
,('Kajin M̧ajeļ',NULL,23,'mh','mah')  
,('മലയാളം',NULL,2,'ml','mal')  
,('मराठी',NULL,10,'mr','mar')  
,('македонски јазик',NULL,10,'mk','mkd')  
,('fiteny malagasy',NULL,23,'mg','mlg')  
,('Malti',NULL,13,'mt','mlt')  
,('Монгол хэл',NULL,17,'mn','mon')  
,('te reo Māori',NULL,23,'mi','mri')  
,('Bahasa Melayu, بهاس ملايو‎',NULL,23,'ms','msa')  
,('ဗမာစာ',NULL,9,'my','mya')  
,('Dorerin Naoero',NULL,23,'na','nau')  
,('Diné bizaad',NULL,22,'nv','nav')  
,('isiNdebele',NULL,1,'nr','nbl')  
,('isiNdebele',NULL,1,'nd','nde')  
,('Owambo',NULL,1,'ng','ndo')  
,('नेपाली',NULL,10,'ne','nep')  
,('Nederlands, Vlaams',NULL,10,'nl','nld')  
,('Norsk Nynorsk',NULL,10,'nn','nno')  
,('Norsk Bokmål',NULL,10,'nb','nob')  
,('Norsk',NULL,10,'no','nor')  
,('chiCheŵa, chinyanja',NULL,1,'ny','nya')  
,('occitan, lenga d''òc',NULL,10,'oc','oci')  
,('ᐊᓂᔑᓈᐯᒧᐎᓐ',NULL,26,'oj','oji')  
,('ଓଡ଼ିଆ',NULL,10,'or','ori')  
,('Afaan Oromoo',NULL,13,'om','orm')  
,('ирон æвзаг',NULL,10,'os','oss')  
,('ਪੰਜਾਬੀ',NULL,10,'pa','pan')  
,('पाऴि',NULL,10,'pi','pli')  
,('Português',NULL,10,'pt','por')  
,('پښتو',NULL,10,'ps','pus')  
,('ελληνικά','Greka',10,'el','ell')  
,('English','Angla',10,'en','eng')  
,('français, langue française','Franca',10,'fr','fra')  
,('Runa Simi, Kichwa',NULL,12,'qu','que')  
,('Rumantsch Grischun',NULL,10,'rm','roh')  
,('Română',NULL,10,'ro','ron')  
,('Ikirundi',NULL,1,'rn','run')  
,('yângâ tî sängö',NULL,3,'sg','sag')  
,('संस्कृतम्',NULL,10,'sa','san')  
,('සිංහල',NULL,10,'si','sin')  
,('Slovenčina, Slovenský Jazyk',NULL,10,'sk','slk')  
,('Slovenski Jezik, Slovenščina',NULL,10,'sl','slv')  
,('Davvisámegiella',NULL,25,'se','sme')  
,('gagana fa''a Samoa',NULL,23,'sm','smo')  
,('chiShona',NULL,1,'sn','sna')  
,('सिन्धी, سنڌي، سندھی‎',NULL,10,'sd','snd')  
,('Soomaaliga, af Soomaali',NULL,13,'so','som')  
,('Sesotho',NULL,1,'st','sot')  
,('Shqip',NULL,10,'sq','sqi')  
,('sardu',NULL,10,'sc','srd')  
,('српски језик',NULL,10,'sr','srp')  
,('SiSwati',NULL,1,'ss','ssw')  
,('Basa Sunda',NULL,23,'su','sun')  
,('Kiswahili',NULL,1,'sw','swa')  
,('Svenska',NULL,10,'sv','swe')  
,('Reo Tahiti',NULL,23,'ty','tah')  
,('தமிழ்',NULL,2,'ta','tam')  
,('татар теле, tatar tele',NULL,4,'tt','tat')  
,('తెలుగు',NULL,2,'te','tel')  
,('тоҷикӣ, toçikī, تاجیکی‎',NULL,10,'tg','tgk')  
,('Wikang Tagalog',NULL,23,'tl','tgl')  
,('ไทย',NULL,24,'th','tha')  
,('ትግርኛ',NULL,13,'ti','tir')  
,('Faka Tonga',NULL,23,'to','ton')  
,('Setswana',NULL,1,'tn','tsn')  
,('Xitsonga',NULL,1,'ts','tso')  
,('Türkmen, Түркмен',NULL,4,'tk','tuk')  
,('Türkçe',NULL,4,'tr','tur')  
,('Twi',NULL,1,'tw','twi')  
,('ئۇيغۇرچە‎, Uyghurche',NULL,4,'ug','uig')  
,('اردو',NULL,10,'ur','urd')  
,('Oʻzbek, Ўзбек, أۇزبېك‎',NULL,4,'uz','uzb')  
,('Tshivenḓa',NULL,1,'ve','ven')  
,('Tiếng Việt',NULL,5,'vi','vie')  
,('Volapük',NULL,14,'vo','vol')  
,('Walon',NULL,10,'wa','wln')  
,('Wollof',NULL,1,'wo','wol')  
,('isiXhosa',NULL,1,'xh','xho')  
,('ייִדיש',NULL,10,'yi','yid')  
,('Yorùbá',NULL,1,'yo','yor')  
,('Saɯ cueŋƅ, Saw cuengh',NULL,24,'za','zha')  
,('中文 (Zhōngwén), 汉语, 漢語',NULL,9,'zh','zho')  
,('isiZulu',NULL,1,'zu','zul');

## Часть III. Ответы на вопросы товарищей

*Towarzystwo cybernetyczne*

[Источник](http://propaganda-journal.net/10402.html).

Пока что нельзя сказать, что очерки «**Что не так с текстологией и как исправить ситуацию?**» вызывали большой интерес и активный практический отклик. Среди тех, кто прислал какой-либо отклик наиболее частой реакцией было недоумение. «Люди практического склада» недоумевали, что очерки слишком уводят в сторону программирования, программисты недоумевали, что нет никаких сведений по работе схемы Фроша-Загорского-Радкевичюте с общепринятыми языками программирования. Текстологи недоумевали, что нет никакого разбора исторических критериев установления авторского текста для ленинских сочинений, а техники недоумевали, что полностью проигнорирована проблема настройки управляющей программы базы данных. Наконец, просто внимательные читатели недоумевали в отношении того, чем отличается схема Фроша-Загорского-Радкевичюте от схемы Фроша-Загорского. Попробуем разрушить это и другие недоумения.

### Вопрос 1

**Что такое схема (в терминологии «схема Фроша-Загорского-Радкевичюте»)?**

Ответ: Схема есть то, что так называется (Die Scheme) в управляющих программах баз данных PostgreSQL и Oracle. Точнее, это совокупность таблиц (хранимых картотек) и правил их взаимной связи (это то, что в языке SQL помещается вокруг слова REFERENCE).

### Вопрос 2

**Чем отличается схема Фроша-Загорского-Радкевичюте от схемы Фроша-Загорского?**

Ответ: Схема Фроша-Загорского (нем. FS-Scheme, пол. FZ-schemat) является частью схемы Фроша-Загорского-Радкевичюте (нем. FSR-Scheme, пол. FZR-schemat). Схема Фроша-Загорского не предусматривает таблицы, отвечающий за хранение графических координат абзаца на фотокопии и таблицы, отвечающей за хранение геометрического описания фотокопий страниц. Иными словами, схема Фроша-Загорского имеет своим содержанием только тексты безотносительно к фотокопиям оригинальных изданий.

### Вопрос 3

**Как выглядит схема Фроша-Загорского-Радкевичюте?**

Ответ: Полноценной и актуальной диаграммы не существует. Однако товарищем Фрошем была реализована база данных, выходящая за пределы FZ-схемы, но не имеющая некоторых таблиц FZR-схемы. Для получения чертежа базы данных мы обратились к товарищу Фрошу. Он любезно согласился на публикацию диаграммы таблиц и связей, созданной по его просьбе товарищами при помощи специальной программы-графопостроителя.

Публикуемая диаграмма отражает попытку введения (в ответ на замечание товаришки Радкевичюте) свойств, содержащих время последнего изменения и название пользователя-редактора. Предложения товаришки Радкевичюте реализованы в данной схеме не в полной мере, также отсутствует связь работ с книгами[[11]](#footnote-50)

Места крепления связей не соотнесены со свойствами, то есть конкретное свойство для связи нужно выбирать из имеющихся в таблице по контексту, не полагаясь на линии. Много где названия свойств, участвующих в связи, выделены жирным шрифтом со стороны источника информации и никак не выделены со стороны приёмника (где хранится ссылка). Тем не менее, по смыслу должно быть понятно, что из одних таблиц уходит в другие. Так факт авторского участия, очевидно, содержит упоминание автора и его произведения.

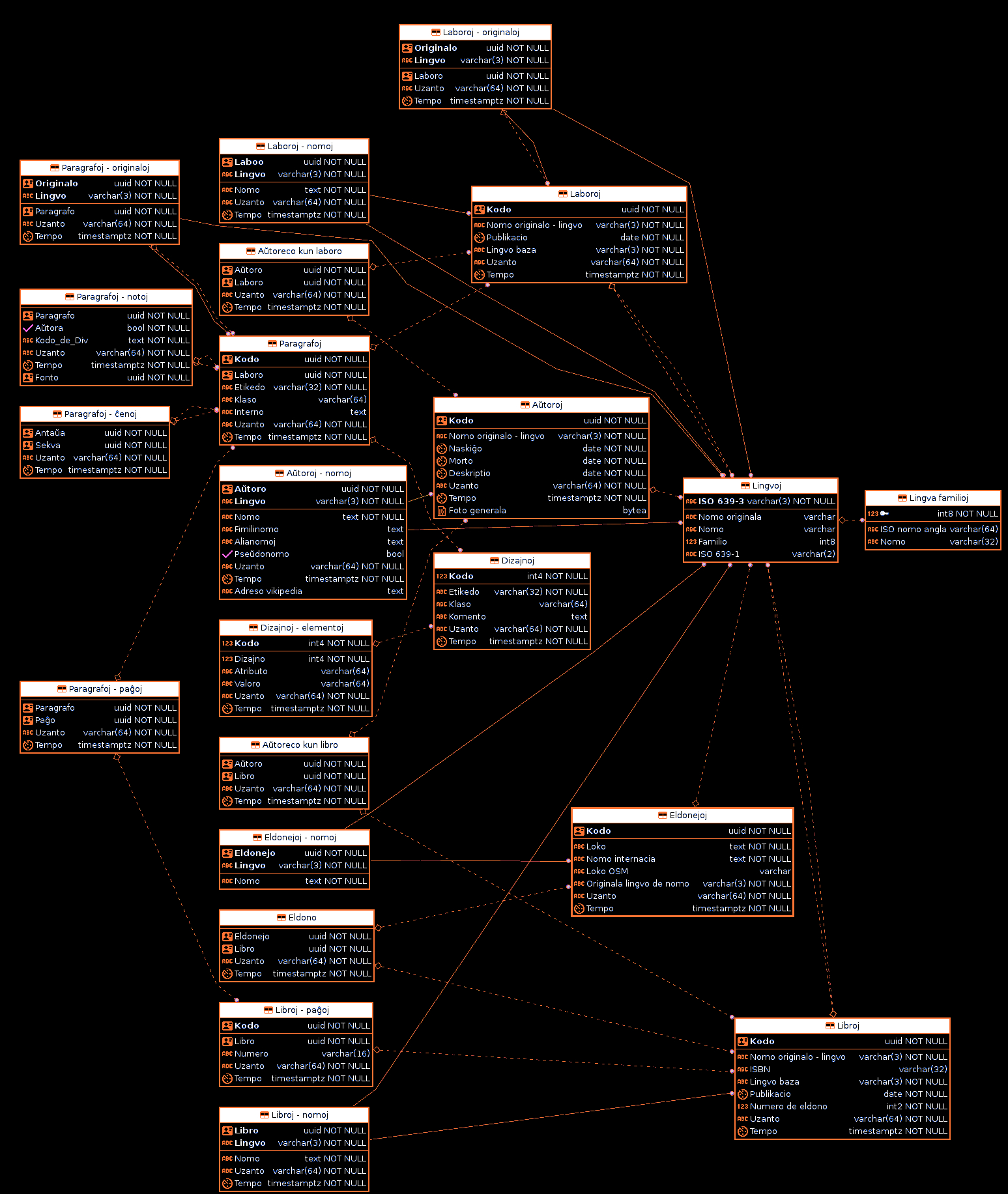


Диаграмма базы данных

Примерные переводы названий таблиц с эсперанто (по рядам слева направо, в каждом ряду сверху вниз)

(первый ряд)

1. Абзац-оригинал (какой абзац является основой перевода)
2. Абзац-примечание (к какому абзацу является примечанием этот абзац)
3. Абзац-последовательность (какой абзац следует за каким в тексте оригинала)
4. Абзац-страница (к каким страницам относится данный абзац)

(второй ряд)

1. Работы-названия (оригинал и переводы названий работ)
2. Авторство работ (какой автор принял участие в написании какой работы)
3. Абзацы (основная картотека текстов абзацев)
4. Авторы-имя (оригинальные и переводные имена авторов)
5. Правила оформления (содержимое специальных правил CSS, определяющих оформление всех текстов)
6. Авторство книг (какой автор указан на титульных листах какой книги)
7. Названия издательств (переводные и оригинальные)
8. Издание (участие издательства в публикации книги)
9. Книги-страницы (перечень страниц с номером и указанием на книгу)
10. Книги-названия (переводные и оригинальные)

(третий ряд)

1. Оригиналы произведений
2. Произведения
3. Авторы
4. Заголовки правил оформления (к каким абзацам каких классов применяются элементы правил)

(четвёртый ряд)

1. Языки
2. Издательства
3. Книги

(пятый ряд)

1. Языковые семьи

### Вопрос 4

**Порождает ли работа по схеме Фроша-Загорского какие-либо информационно-технические зависимости?**

Ответ: Жить в обществе и быть свободным от общества нельзя. Вопрос содержательно переносится в характер этого влияния.

С юридической точки зрения правовые системы всех стабильных государств современности гарантируют свободу использования, распространения и доработки под свои нужды программ, управляющих базами данных - SQLite и PostgreSQL. Подобно ближайшим аналогам - программам Firebird, MariaDB, MySQL - названные находятся в той или иной форме общественной собственности: всякому гарантируется присвоение, но исключается любая попытка приватизации исключительных прав. SQLite, PostgreSQL и Firebird производятся коммунистическим способом - международным комитетом программистов-добровольцев, зарабатывающих иным способом. Не существует запретов на создание своих модификаций названных программ и не существует никаких платёжных обязательств в отношении авторов или распространителей этих программ. Конкретно SQLite и PostgreSQL созданы на языке программирования Си, который известен с 1970-х годов и может создавать программы для большого числа разных моделей процессоров. Формальный язык SQL в той части, которая нужна для схемы Фроша-Загорского-Радкевичюте, поддерживается весьма большим количеством платных и бесплатных программ управления базами данных, производимых как капиталистическим так и коммунистическим способом. Бесплатные производимые коммунистическим способом программы SQLite и PostgreSQL вследствие создания на языке Си почти не ограничены в выборе модели процессора. Таким образом, как информационная, так и техническая зависимость при работе по схеме Фроша-Загорского-Радкевичюте не больше, чем при работе с безбумажной формой текста в виде гипертекста. Например, форматы гипертекста стандартизированы в США в документах, многие из которых не имеют переводов с английского языка. В противоположность, схема Фроша-Загорского-Радкевичюте предлагается в наглядном виде на языке эсперанто для облегчения работы с ней и описания её сути на языках Европы.

### Вопрос 5

**Что было источником вдохновения для схемы Фроша-Загорского-Радкевичюте?**

С этим вопросом мы обратились к товарищу Фрошу.

«Источником вдохновения было конкретное бумажное издание, которое показал мне один товарищ, связанный с помощью польским и украинским товарищам. Это трёхязычное издание «Іван Франко Franko Iwan Иван Франко Зів’яле листя Zwiędłe liście Увядшие листья» (Львів, Каменяр, 2003). Это было примерно в 2007 году. Через день мы с тем товарищем синхронно направили друг другу письма, что хорошо бы иметь в таком виде работы Ленина. Через несколько месяцев выяснилась невозможность сотрудничества с какими-либо носителями языка Ленина. В 2013 году заезжая литовка дала контакты польских текстологов, состоящих теперь в SMP . Лишь наши самообразовательные курсы по базам данных активизировали подготовку принципов FSR-схемы в 2015 году. Довольно легко убедиться, что названное издание отлично раскладывается в базу данных, созданную по FSR-схеме».

### Вопрос 6

**Что такое uuid, какую роль он играет и для чего применяется.**

UUID[[12]](#footnote-55) (эта аббревиатура чаще пишется заглавными буквами) - это технический стандарт, предназначенный для создания особого бессодержательного представителя *иного*. UUID как представитель *иного* аналогичен номеру (документа, шкафа, заводского изделия, тепловоза, кабинета), однако не выполняет очевидное правило о том, что больший номер выдаётся позднее. В общем сравнение UUID имеет смысл только с результатом равен и не равен, но едва ли имеет какое-либо значение результат больше или меньше. От случайных чисел, нужной длины, получаемых разными способами, UUID отличается тем, что обеспечивает не столько случайность, сколько неповторимость. Вероятность создать одинаковые UUID на разных машинах должна быть крайне низкой. При создании миллиардов UUID ежесекундно повторение должно гарантированно произойти не чаще одного раза в несколько миллиардов лет. По просьбе товарища Фроша несколько немецких математиков подтвердили такую оценку повторимости UUID для алгоритма uuid\_generate\_v4(), заложенного в модуль uuid-ossp внутри PosgtgreSQL.

В отличие от номеров, где целое число может быть почтовым индексом, номером швейной машины, номером квартиры, телефонным кодом абонента и т. д., UUID однозначно указывает на то, что он означает. Скажем, неизвестно к чему относящийся UUID можно проверить по всем каталогам конкретной базы данных и выяснить к чему он относится, если все элементы какого-либо рода снабжены представляющими их кодами UUID. В этом случае в одном из каталогов данный UUID совпадёт и укажет на все другие свойства представляемого им объекта.

### Вопрос 7

**Чем вызвана странная форма записи UUID?**

Ответ: Если взять образец формата «a0eebc99-9c0b-4ef8-bb6d-6bb9bd380a11», то такая форма записи вызвана программистскими особенностями формирования UUID. В одной из отделённых четвёрок зафиксирован способ получения UUID и она обычно не является случайной. Других подробностей узнать не удалось, ибо не нашлись сведения на немецком, польском или французском языке по истории программирования касательно получения UUID.

Возможно, товарищам помогут сведения, что в SQLite полноценный UUID в текстовом виде можно создать директивой

select '{' || hex( randomblob(4)) || '-' || hex( randomblob(2))  
|| '-' || '4' || substr( hex( randomblob(2)), 2) || '-'  
|| substr('AB89', 1 + (abs(random()) % 4) , 1) ||  
substr(hex(randomblob(2)), 2) || '-' || hex(randomblob(6)) ||  
'}';

### Вопрос 8

**Как UUID может хранить тип обозначаемого объекта, если это бессодержательный и неповторимый набор кодов?**

Ответ: UUID как идеальный феномен ничем не отличается от других аналогичных феноменов по принципу функционирования. Так статуи Венеры не повторяют физическое тело, а воспроизводят его геометрию и фактуру в той мере, в какой это было значимо для античных мастеров. Так надпись «ptak» не хранит ничего связанного с признаками группы живых существ - птиц. С реальными птицами эта надпись не имеет никаких существенных общих свойств. Разве что надпись, как и птица, - это нечто протяжённое. Однако всякий, знающий польский язык, без труда проводит связь от экранных или типографских знаков «ptak» до реальных живых существ. Подобно надписи, UUID бесполезен для того, кто не умеет его применять по месту. Как знаки надписи, так и UUID ничего не содержат для того, кто не знает ни одного упоминания. Но как только появляется словарь польского языка или база данных с нужными каталогами, так для всякого человека обретают смысл надпись в случае со словарём и UUID в случае с базой данных. В каждом случае надпись и UUID, будучи по форме беесодержательными, несут в себе общественно важное содержимое некоего иного. Притом это содержание раскрывается всегда только для действующего по свои целям, связанного с обществом человека и только в процессе действия. Без осмысления надписи и без приказаний базе данных проверить наличие UUID оба феномена не имеют общественного значения и содержательных функций.

### Вопрос 9

**Что произошло с книжным комплектом 5-го издания ПСС Ленина в США?**

Ответ: Мы обратились за разъяснениями к товарищам, организовавшим текстологическое совещание. Они сообщили, что убеждения гостя из Сан-Франциско не были известны - «он не предоставил никаких публикаций или текстов для прочтения. Латвийские товарищи предложили рассматривать гостя как политического авантюриста, а получение книжного комплекта как коммерческую операцию. Гость довольно хорошо разбирался в текстологии. Даже лучше, чем в немецком языке, но не демонстрировал желания держать контакт. В конце 2015 года попытки написать на указанный адрес электронной почты стали приводить к ошибке «неправильный или несуществующий получатель»».

Исходя из имеющихся у нас сообщений, нет смысла предполагать, что какие-либо самообразовательные сообщества в США приступили к текстологическим работам по произведениям Ленина. Критическое переиздание или комплектование книг, называемых у белорусов двухмоўнікамі, при существующем состоянии гипертекстов оригиналов исключается. Вероятнее всего, книжный комплект попал в академическую среду на началах обыкновенной товарной спекуляции. По опыту некоторых стран можно вывести, что направления текстологической работы связаны с направлениями теоретической работы через несколько опосредствующих элементов, то есть прямого соответствия не существует. Далеко не всегда уровень самообразовательной работы прямо сказывается на уровне текстологической работы. Единство этих уровней является чертой некоторой части центральной Европы и ряда провинций Индии, тогда как в большинстве остальных стран может быть найдена как текстология без связи с самообразованием, так и самообразование без связи с текстологическими работами.

### Вопрос 10

**Каков ожидаемый результат текстологической работы с использованием FZR-схемы?**

Ответ: Результат непредсказуем, он ограничен творческими догадками тех людей, который будут работать с базой данных. Названные принципы построения базы данных потенциально пригодны и для автоматического создания прототипов одноязычных бумажных изданий, и для создания академических двойных и тройных (двухмоўнік, трохмоўнік) параллельных текстов. Автоматическое создание разных форматов электронных книг едва ли стоит упоминать как очевидное применение. Лёгкое улучшение перевода через доступность оригинала и лёгкое поабзацное создание перевода без отвлечения на проблемы оформления издания также должно быть упомянуто.

### Вопрос 11

**Существует ли реализация схемы Фроша-Загорского-Радкевичюте?**

Ответ: Да. Авторы схемы согласились признать, что её реализует (за исключением двух языковых таблиц, описанных ранее) некоторая апробированная группа SQLкоманд для PostgreSQL (помещается в приложение №1). Предпосылкой правильного выполнения этих команд является наличие заполненного перечня языков и языковых семей, как было указано в предыдущей статье настоящего цикла.

### Вопрос 12

**Каковы ближайшие задачи по развитию FZR-схемы?**

Ответ: Основательное комментирование приведённых выше команд на трёх мировых языках теоретического мышления. Сбор откликов.

### Вопрос 13

**Существуют ли образцы заполнения баз данных созданных по FZR-сехме?**

Ответ: Нет, за исключением образцов из предыдущей части. Пока что отсутствуют образцы описания работы через оригинальную книгу, выпущенную издательством и имеющую авторство.

### Вопрос 14

**Нельзя ли понимать FZR-сехму иначе, чем группу таблиц базы данных с указанными связями?**

Ответ: Нет. Только понимание группы таблиц как связанных в базе данных позволяет легко обращаться с результатами всех видов текстологической работы. Если же понимать FZR-схему как справочную основу для какой-либо одной программы или функции, то будет непонятен эффект объединения всех видов текстологического труда. FZR-сехма задумывалась *не для удовлетворения текстологических потребностей, а для обеспечения самообразовательных потребностей и международного теоретического взаимодействия*. В настоящий момент FZ-сехма охватывает (не ограничиваясь указанными) такие виды текстологического труда как:

1. Проверка достоверности (верификация) по оригиналу.
2. Корректура результатов полуавтоматического формирования текста по фотокопии.
3. Формирование параллельных многоязычных изданий.
4. Создание новых переводов.
5. Перепроверка переводов.
6. Поиск контекста цитат.
7. Поиск и перепроверка библиографических ссылок на оригинальное бумажное издание.

### Вопрос 15

**Как произвести занесение конкретного тома ленинских работ (например,** http://uaio.ru/vil/33.htm**) в базу данных, реализующую FZR-схему?**

Ответ: Сводная рекомендация будет опубликована отдельной статьёй. Если коротко, то многое решает анализ текстологических свойств конкретного источника, после чего вырабатывается группа SQL команд, порядок которых можно считать типовым, ибо он не зависит от многих особенностей источника.

### Вопрос 16

**Каков принцип формирования сложной разметки текстов, например, отображения ленинского конспекта книги Фейербаха из XXIX тома?**

Ответ: Руководящий текстологический принцип: в таблицу абзацев не попадет ни одного элемента, который не происходил бы прямо из авторской рукописи. Так, если автор отчеркнул тремя чертами слева строку с каким-то предложением, то это должно быть воспроизведено тут же в гипертекстовой форме. Безусловно, это относится к таким простым приёмам как курсив, жирный шрифт, подчёркивание, надчёркивание, написание индексов. Если же издатели отметили все абзацы цитат шрифтом несколько меньшего размера, но авторская рукопись не даёт примеров такого приёма, то абзацы цитат подлежат пометке «», что даст в таблице абзацев Etikedo = p Klaso = Citato. В этом случае правила оформления цитат должны быть прописаны в двух таблицах, название которых включает «Dizajnoj». В одной таблице указывается применимость правил (элемент p и классом Citato), в другой его код связывается с отдельными правилами оформления.

### Вопрос 17

**Существуют ли текстологические проблемы, связанные с гипертекстовой формой представления?**

Ответ: Да. В XXIX томе 5-го издания ленинских сочинений может быть найдено несколько приёмов, которые до сих пор не имеют соответствий в стандарте гипертекста HTML5. Немецкие товарищи из текстологической группы mlwerke.de указали на фигурные скобки и некоторые типы надчёрквиваний. Полноценной текстологической экспертизы оригинала XXIX тома до сих пор никем не было осуществлено. Однако несомненно, что все авторские приёмы оформления текста (за исключением выделения заглавий и, может быть, надписей на полях) не подлежат выносу в CSS и подлежат хранению в составе гипертекста.

### Вопрос 18

**Значительны ли текстологические проблемы, связанные с гипертекстовой формой представления?**

Ответ: Нет. За исключением XXIX тома и нескольких подобных работ, всё остальное с точки зрения текстологических работ и текстологической экспертизы не требует знаний за пределами курса основной школы.

### Вопрос 19

**Как FZR-схема позволяет хранить иллюстрации?**

Ответ: Авторские иллюстрации хранятся в гипертекстовом виде по месту как отдельный абзац или в составе абзаца, что определятся источником. Редакционные иллюстрации, титульные и технические листы не описаны в FZR-схеме. Однако возможно создать отдельную таблицу с привязкой к бумажному изданию для хранения подобных элементов. Заинтересованные текстологи приглашаются для обсуждения.

### Вопрос 20

**Какова правовая форма FZR-схемы и FZ-схемы?**

Ответ: Авторы определили запрет на коммерческое использование подобной структуры базы данных до успешных испытаний на массовой работе. После того, как будет подтверждено, что базы данных, созданные по названным принципам, позволяют сохранять текстологическую работу без утраты авторского замысла и полноты бумажного издания-прототипа, ожидается передача всех SQL-команд на создание базы данных в общественное достояние.

### Вопрос 21

**Какова правовая форма текстологического результата в базах данных, созданных по FZR-схеме или по FZ-схеме?**

Ответ: В главном это определяется теми, кто контролирует базу данных. Авторы FZR-схемы рекомендуют устанавливать режим ODBL лицензии или лицензии творческих коммун CC-BY-SA.

### Вопрос 22

**Какой этап автоматизации и работ по программированию проходит сейчас FZR-схема?**

Ответ: По причине отсутствия людей с должной квалификацией в программировании в настоящий момент средства автоматизации загрузки текстов в базу данных полностью отсутствуют. В виде группы SQL-директив существует формирователь одноязычного гипертекстового прототипа любой работы. В виде программы на языке Java существует средство занесения стилей оформления текстов CSS в базу данных. Несколько функций были автоматизированы для терминала операционной системы в виде sh файлов с применением https://pl.wikipedia.org/wiki/Zenity и стандартного клиента psql. Однако эти черновые программы не выдерживают критики с точки зрения скорости, многоязычности и безопасности. В настоящий момент над созданием комплекта программ для FZR-схемы не работает ни один программист и эргономист, более того, графических интерфейсов FZR-схемы до сих пор не существует. Эта схема базы данных была создана на основании профессиональных знаний по математике (теория баз данных), кибернетике и текстологии, соображения общего программирования не рассматривались.

### Вопрос 23

**Какой объём хранения информации может потребовать база данных по схеме Фроша-Загорского для всего 5-го издания ленинских сочинений?**

Ответ: Нормативный гипертекст в международной кодировке уникод для каждого тома занимает оценочно 4-6 мегабайт. При прямой оценке это 275 мегабайт текстовых данных. Добавляя использование кодов uuid, занимающих пятую часть от объёма текста (очень пессимистическая оценка) получаем 330 мегабайт. Книжные описания и описания работ можно оценить в 2-3 мегабайта. Итого огромный текстологический проект по созданию гипертекстов ленинских работ может уместиться по пессимистической оценке в 512 мегабайт пространства хранения. В тройной копии без архивирования это 1,5 гигабайта, что равно объёму 2 CD дисков. Фактически целесообразны 2 независимых CD копии. Затраты на накопители не превышают 5 евро для дорогих моделей. Проблема текстологического качества (упорядоченности и проверенности базы данных) и обобществления труда в данном случае намного превосходит по сложности любые технические трудности, что ясно из опыта международного текстологического проекта MEGA (издание оригиналов сочинений Маркса и Энгельса).

### Вопрос 24

**Существуют ли публичные базы данных, созданные по схемам FZ или FRZ?**

Ответ: Нет. Существует испытательная модель в одном из городов Бранденбурга и готовится аналогичная испытательная модель в Варшаве. Для публичной работы обе модели не предназначены и публичных адресов не имеют. Все существенные команды по созданию таких моделей были опубликованы ранее и всякий внимательный читатель (даже совершенно незнакомый с базами данных до этой серии статей) может создать у себя аналогичную испытательную модель.

### Приложение 1

**Директивы SQL по созданию базы данных для текстологической работы**

/\* TEKSTUJO INTERNACIO ĜENERALA PROEKTO DE DATUMBAZO \*/  
  
CREATE TABLE "Eldonejoj" (  
"Kodo" uuid NOT NULL DEFAULT uuid\_generate\_v4(),  
"Loko" text NOT NULL,  
"Nomo internacia" text NOT NULL,  
"Loko OSM" varchar NULL,  
"Originala lingvo de nomo" varchar(3) NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Eldonejoj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Eldonejoj\_Lingvoj\_FK" FOREIGN KEY ("Nomo originalo - lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
   
CREATE TABLE "Eldonejoj - nomoj" (  
"Eldonejo" uuid NOT NULL,  
"Lingvo" varchar(3) NOT NULL,  
"Nomo" text NOT NULL,  
CONSTRAINT "Eldonejoj\_nomoj\_pk" PRIMARY KEY ("Eldonejo","Lingvo"),  
CONSTRAINT "Eldonejoj\_nomoj\_Eldonejoj\_FK" FOREIGN KEY ("Eldonejo") REFERENCES "Eldonejoj"("Kodo"),  
CONSTRAINT "Eldonejoj\_nomoj\_Lingvoj\_FK" FOREIGN KEY ("Lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
  
CREATE TABLE "Libroj" (  
"Kodo" uuid NOT NULL DEFAULT uuid\_generate\_v4(),  
"Nomo originalo - lingvo" varchar(3) NOT NULL,  
"ISBN" varchar(32) NULL,  
"Publikacio" date NOT NULL,  
"Numero de eldono" int2 NOT NULL DEFAULT 1,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Libroj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Libroj\_Lingvoj\_FK" FOREIGN KEY ("Nomo originalo - lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
  
CREATE TABLE "Libroj - nomoj" (  
"Libro" uuid NOT NULL,  
"Lingvo" varchar(3) NOT NULL,  
"Nomo" text NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Libroj\_nomoj\_pk" PRIMARY KEY ("Libro","Lingvo"),  
CONSTRAINT "Libroj\_nomoj\_Libroj\_FK" FOREIGN KEY ("Libro") REFERENCES "Libroj"("Kodo"),  
CONSTRAINT "Libroj\_nomoj\_Lingvoj\_FK" FOREIGN KEY ("Lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
  
CREATE TABLE "Libroj - paĝoj" (  
"Kodo" uuid NOT NULL DEFAULT uuid\_generate\_v4(),  
"Libro" uuid NOT NULL,  
"Numero" varchar(16) NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Libroj\_paĝoj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Libroj\_paĝoj\_Libroj\_FK" FOREIGN KEY ("Libro") REFERENCES "Libroj"("Kodo")  
);  
  
CREATE TABLE "Aŭtoroj" (  
"Kodo" uuid NOT NULL DEFAULT uuid\_generate\_v4(),  
"Nomo originalo - lingvo" varchar(3) NOT NULL,  
"Naskiĝo" date NOT NULL,  
"Morto" date NOT NULL,  
"Deskripcio" text NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
"Foto generala" bytea NULL,  
CONSTRAINT "Aŭtoroj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Aŭtoroj\_Lingvoj\_FK" FOREIGN KEY ("Nomo originalo - lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
  
CREATE TABLE "Aŭtoroj - nomoj" (  
"Aŭtoro" uuid NOT NULL,  
"Lingvo" varchar(3) NOT NULL,  
"Nomo" text NOT NULL,  
"Familinomo" text NULL,  
"Alianomoj" text NULL,  
"Pseŭdonomo" bool,  
"Adreso vikipedia" text,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Aŭtoroj\_nomoj\_pk" PRIMARY KEY ("Aŭtoro","Lingvo"),  
CONSTRAINT "Aŭtoroj\_nomoj\_Aŭtoroj\_FK" FOREIGN KEY ("Aŭtoro") REFERENCES "Aŭtoroj"("Kodo"),  
CONSTRAINT "Aŭtoroj\_nomoj\_Lingvoj\_FK" FOREIGN KEY ("Lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
  
CREATE TABLE "Eldono" (  
"Eldonejo" uuid NOT NULL,  
"Libro" uuid NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Eldono\_Eldonejoj\_FK" FOREIGN KEY ("Eldonejo") REFERENCES "Eldonejoj"("Kodo"),  
CONSTRAINT "Eldono\_Libroj\_FK" FOREIGN KEY ("Libro") REFERENCES "Libroj"("Kodo")  
);  
  
CREATE TABLE "Aŭtoreco kun libro" (  
"Aŭtoro" uuid NOT NULL,  
"Libro" uuid NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Eldono\_Aŭtoroj\_FK" FOREIGN KEY ("Aŭtoro") REFERENCES "Aŭtoroj"("Kodo"),  
CONSTRAINT "Eldono\_Libroj\_FK" FOREIGN KEY ("Libro") REFERENCES "Libroj"("Kodo")  
);  
  
CREATE TABLE "Laboroj" (  
"Kodo" uuid NOT NULL DEFAULT uuid\_generate\_v4(),  
"Nomo originalo - lingvo" varchar(3) NOT NULL,  
"Publikacio" date NOT NULL,  
"Libro" uuid,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Laboroj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Laboroj\_Lingvoj\_FK" FOREIGN KEY ("Nomo originalo - lingvo") REFERENCES "Lingvoj"("ISO 639-3"),  
CONSTRAINT "Laboroj\_Libroj\_FK" FOREIGN KEY ("Libro") REFERENCES "Libroj"("Kodo")  
);  
  
CREATE TABLE "Laboroj - nomoj" (  
"Laboro" uuid NOT NULL,  
"Lingvo" varchar(3) NOT NULL,  
"Nomo" text NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Laboroj\_nomoj\_pk" PRIMARY KEY ("Laboo","Lingvo"),  
CONSTRAINT "Laboroj\_nomoj\_Laboroj\_FK" FOREIGN KEY ("Laboro") REFERENCES "Laboroj"("Kodo"),  
CONSTRAINT "Laboroj\_nomoj\_Lingvoj\_FK" FOREIGN KEY ("Lingvo") REFERENCES "Lingvoj"("ISO 639-3")  
);  
  
CREATE TABLE "Laboroj - originaloj" (  
"Laboro" uuid NOT NULL,  
"Originalo" uuid NOT NULL,  
"Lingvo" varchar(3) NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Laboroj\_originaloj\_pk" PRIMARY KEY ("Lingvo", "Originalo"),  
CONSTRAINT "Laboroj\_originaloj\_Lingvoj\_FK" FOREIGN KEY ("Lingvo") REFERENCES "Lingvoj"("ISO 639-3"),  
CONSTRAINT "Laboroj\_originaloj\_Laboroj\_FK" FOREIGN KEY ("Laboro") REFERENCES "Laboroj"("Kodo"),  
CONSTRAINT "Laboroj\_originaloj\_Laboroj\_FK\_1" FOREIGN KEY ("Originalo") REFERENCES "Laboroj"("Kodo")  
);  
  
CREATE TABLE "Aŭtoreco kun laboro" (  
"Aŭtoro" uuid NOT NULL,  
"Laboro" uuid NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Eldono\_Aŭtoroj\_FK" FOREIGN KEY ("Aŭtoro") REFERENCES "Aŭtoroj"("Kodo"),  
CONSTRAINT "Eldono\_Laboroj\_FK" FOREIGN KEY ("Laboro") REFERENCES "Laboroj"("Kodo")  
);  
  
/\* CSS \*/  
CREATE TABLE "Dizajnoj" (  
"Kodo" serial NOT NULL,  
"Etikedo" varchar(32) NOT NULL,  
"Klaso" varchar(64),  
"Komento" text,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Dizajnoj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Dizajnoj\_uq" UNIQUE ("Etikedo", "Klaso")  
);  
  
CREATE TABLE "Dizajnoj - elementoj" (  
"Kodo" serial NOT NULL,  
"Dizajno" int NOT NULL,  
"Atributo" varchar(64),  
"Valoro" varchar(64),  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Dizajnoj\_elementoj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Dizajnoj\_elementoj\_Dizajnoj\_FK" FOREIGN KEY ("Dizajno") REFERENCES "Dizajnoj"("Kodo")  
);  
  
/\* PARAGRAFOJ \*/  
CREATE TABLE "Paragrafoj" (  
"Kodo" uuid NOT NULL DEFAULT uuid\_generate\_v4(),  
"Laboro" uuid NOT NULL,  
"Etikedo" varchar(32) NOT NULL,  
"Klaso" varchar(64) NULL,  
"Interno" text NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Paragrafoj\_pk" PRIMARY KEY ("Kodo"),  
CONSTRAINT "Paragrafoj\_Dizajnoj\_FK" FOREIGN KEY ("Etikedo", "Klaso") REFERENCES "Dizajnoj"("Etikedo", "Klaso"),  
CONSTRAINT "Paragrafoj\_Laboroj\_FK" FOREIGN KEY ("Laboro") REFERENCES "Laboroj"("Kodo")  
);  
  
CREATE TABLE "Paragrafoj - ĉenoj" (  
"Antaŭa" uuid NOT NULL,  
"Sekva" uuid NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Paragrafoj\_ĉenoj\_Paragrafoj\_FK" FOREIGN KEY ("Antaŭa") REFERENCES "Paragrafoj"("Kodo"),  
CONSTRAINT "Paragrafoj\_ĉenoj\_Paragrafoj\_FK\_1" FOREIGN KEY ("Sekva") REFERENCES "Paragrafoj"("Kodo")  
);  
  
CREATE TABLE "Paragrafoj - notoj" (  
"Paragrafo" uuid NOT NULL,  
"Aŭtora" bool NOT NULL,  
"Kodo\_de\_Div" text NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
"Fonto" uuid NOT NULL,  
CONSTRAINT "Paragrafoj\_notoj\_Paragrafoj\_FK" FOREIGN KEY ("Paragrafo") REFERENCES "Paragrafoj"("Kodo"),  
CONSTRAINT "Paragrafoj\_notoj\_Paragrafoj\_FK1" FOREIGN KEY ("Fonto") REFERENCES "Paragrafoj"("Kodo")  
);  
  
CREATE TABLE "Paragrafoj - originaloj" (  
"Paragrafo" uuid NOT NULL,  
"Originalo" uuid NOT NULL,  
"Lingvo" varchar(3) NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Paragrafoj\_originaloj\_pk" PRIMARY KEY ("Lingvo", "Originalo"),  
CONSTRAINT "Paragrafoj\_originaloj\_Lingvoj\_FK" FOREIGN KEY ("Lingvo") REFERENCES "Lingvoj"("ISO 639-3"),  
CONSTRAINT "Paragrafoj\_originaloj\_Paragrafoj\_FK" FOREIGN KEY ("Paragrafo") REFERENCES "Paragrafoj"("Kodo"),  
CONSTRAINT "Paragrafoj\_originaloj\_Paragrafoj\_FK\_1" FOREIGN KEY ("Originalo") REFERENCES "Paragrafoj"("Kodo")  
);  
  
CREATE TABLE "Paragrafoj - paĝoj" (  
"Paragrafo" uuid NOT NULL,  
"Paĝo" uuid NOT NULL,  
"Uzanto" varchar(64) NOT NULL DEFAULT user,  
"Tempo" timestamptz NOT NULL DEFAULT now(),  
CONSTRAINT "Paragrafoj\_paĝoj\_Libroj\_paĝoj\_FK" FOREIGN KEY ("Paĝo") REFERENCES "Libroj - paĝoj"("Kodo"),  
CONSTRAINT "Paragrafoj\_paĝoj\_Paragrafoj\_FK" FOREIGN KEY ("Paragrafo") REFERENCES "Paragrafoj"("Kodo")  
);

## Часть IV. Пример текстологического процесса

*Towarzystwo cybernetyczne*

[Источник](http://propaganda-journal.net/10412.html).

**Введение**

Этот очерк помимо нескольких вопросов, пришедших от разных товарищей, посвящён разбору конкретного (притом весьма типичного) текстологического процесса над источником, указанным в вопросе №15 из предыдущей части этой серии. Повторим этот вопрос.

### Вопрос 15

**Как произвести занесение конкретного тома ленинских работ (например,** http://uaio.ru/vil/33.htm**) в базу данных, реализующую FZR-схему?**

Полноценный ответ на этот вопрос требует, во-первых, детального раскрытия текстологических свойств конкретного источника, и, во-вторых, формирования группы SQL команд, порядок которых можно считать типовым. Читатель, не знакомый со сферой текстологической оценки, должен будет уловить преимущественно порядок заполнения сведений в SQL. Этот принцип более важен, чем текстологическая оценка; хотя бы потому, что порядок формирования базы данных не сильно отличается для источников в разных формах и на разных языках, а, значит, будет полезен также для понимания принципов международной текстологической работы.

#### Пример анализа текстологических свойств источника

Источником в нашем случае выступает гипертекст с публичным адресом http://uaio.ru/vil/33.htm. Какие особенности источника нужно рассмотреть в первую очередь? К сожалению, на этот вопрос нет никакого стандартного ответа. Текстология ориентируется в отношении произведений Нового Времени на содержательную полноту, понимаемую как точную передачу авторских особенностей. Если речь идёт об академическом гипертексте, то должна быть соблюдена и формальная полнота, проявляющаяся в заимствовании аппарата из наиболее авторитетного издания а также во вспомогательной разметке страниц. Посмотрим, как обстоит дело с полнотой источника.

С точки зрения формальной полноты представленный источник был подвергнут сравнению с польским изданием «Włodzimierz Lenin Dzieła wszystkie, tom 33» ([в фотокопии](http://marksizm.edu.pl/wydawnictwa/klasyka-mysli-marksistowskiej/wlodzimierz-i-lenin/dziela-wszystkie-tom-33/)). Сравниваемый российский источник, поскольку он не является фотокопией, имеет более высокое качество. С точки зрения состава текстов русский и польский варианты ничем не отличаются. Аппарат 5-го издания ленинских сочинений на языке оригинала в рассматриваемом источнике заимствован. Также воспроизведена страничная структура. Однако, что касается авторского текста, то содержательная полнота принесена в жертву формальной. Авторский текст сильно разбавлен страничными отметками с линейками, что разрывает цитаты и нарушает поиск словосочетаний-цитат, если они попадают на границу страниц.

#### Оценка титульной части

Проведём конкретный анализ гипертекста. Его титульная часть выглядит так:

Ленин ПСС издание 5 том 33

В титульной части можно обнаружить языковую метку «**content=“ru”**», что позволяет добраться до стандартного трёхбуквенного кода через каталог языков. Это безусловное достоинство.

Рядом есть ссылка на таблицу оформления «». В действительности таблица эта весьма скромная и отнюдь не полностью охватывает способы оформления, заимствуемые из печатного издания. Однако никакое значительное искажение авторского текста из этого не происходит.

Достоинством является вынесение названия документа в подпись гипертекста: «**Ленин ПСС издание 5 том 33**».

Несомненным недостатком является применение шовинистической системы кодирования «**charset=windows-1251"**». Немецкие текстологи не позднее 2002 года доказали, что эта кодировка непригодна для воспроизведения ленинских сочинений. Своей ненужной национальной ограниченностью и архаичностью она лишает нас возможности изучать, например, ленинский конспект одного французского популяризатора наследия Гегеля. Нужных символов для французских слов в этой кодировке нет. Нет в ней двойного восклицательного знака и знака «½». Точнее не просто нет, а не было, нет и никогда не будет, как и в других архаически-шовинистических кодировках типа №1251 (латышско-эстонская?) или №1250 (польско-чешско-словацкая). Разумеется, архаичные и шовинистические кодировки для воспроизведения классических работ - это не российская особенность, но здесь хочется попросить российских товарищей не уподобляться ни польским текстологам 2001 года, ни чешским текстологам 2003 года, которые с применением архаической и шовинистической кодировки №1250 воспроизвели некоторые ленинские сочинения по адресам https://www.marxists.org/cestina/lenin/1917/041917t.html и https://www.marxists.org/polski/lenin/1917/07/10zryw.htm. С кучей архаических шовинистических кодировок не только невозможно полноценно воспроизводить классические тексты (авторы которых не признавали ограничения местным набором букв и знаков), но и нельзя координировать международные работы без изрядной квалификации в программировании. Международная кодировка Уникод ныне уничтожила для языков Европы проблему недоступных символов, которая осталась актуальной лишь для некоторых редких алфавитов в Индии. В отношении популярных алфавитов и популярной пунктуации теперь Уникодом каждому гарантируется возможность прочтения оригинала без искажений. Для ссылки между разными языковыми вариантами ленинских сочинений или для параллельного текста всякие иные варианты неприменимы в принципе.

#### Оценка образца внутренностей

Разбираемый образец охватывает несколько первых страниц. Все остальные построены на основе тех же самых особенностей. Приводимому ниже коду гипертекста точно соответствуют первые авторские страницы по адресу http://uaio.ru/vil/33.htm#s1.

1  
  
ГОСУДАРСТВО И РЕВОЛЮЦИЯ  
   
УЧЕНИЕ МАРКСИЗМА О ГОСУДАРСТВЕ И ЗАДАЧИ ПРОЛЕТАРИАТА В РЕВОЛЮЦИИ 1  
  
Написано в августе - сентябре  
1917 г.; § 3 главы II - ранее  
17 декабря 1918 г.  
   
Напечатано в 1918 г.  
в Петрограде отдельной книгой  
издательством «Жизнь и знание»  
  
Печатается по рукописи, сверенной с текстом книги, изданной в 1919 г.  
в Москве - Петрограде издательством «Коммунист»  
  
2  
  
Первая страница рукописи В. И. Ленина «Государство и революция». - Август - сентябрь 1917 г.  
  
Уменьшено  
  
3  
  
ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ  
   
Вопрос о государстве приобретает в настоящее время \*\*\*\* Передовые страны превращаются - мы говорим о «тыле» их - в военно-каторжные тюрьмы для рабочих.  
  
Неслыханные ужасы и бедствия затягивающейся войны делают положение масс невыносимым, усиливают возмущение их. Явно нарастает международная пролетарская революция. Вопрос об отношении ее к государству приобретает практическое значение.  
  
Накопленные десятилетиями сравнительно мирного развития элементы \*\*\*\* ибо большинство так называемых великих держав давно эксплуатирует и порабощает  
  
4  
  
В. И. ЛЕНИН  
  
целый ряд мелких и слабых народностей. А империалистская война \*\*\*\* ига капитала, в ближайшем будущем.  
  
Автор

Итак, попробуем разобраться, как ленинский авторский текст связан с цитируемой разметкой.

Начало авторской страницы №1 отмечено как , а начало авторской страницы №2 как . Сам факт наличия постраничных ссылок, являющихся пробелами образует весьма удобную базу для справочного цитирования бумажного оригинала и сверки фотокопий (если они будут когда-либо загружены при работе по варианту Радкевичюте). Список всех страниц книги (из предисловия и авторских) подлежит занесению в таблицу “Libroj - paĝoj”. Предварительно должна быть описана с указанием авторства и издательства вся книга. Упоминание самого тома будет во всякой записи о странице происходить как присутствие UUID всей книги в каждой соответствующей записи из таблицы “Libroj - paĝoj”.

Место смены страниц передаётся хуже всего при работе с имитациями бумажных изданий. Так за отметкой о смене страницы, которая уже известно читателю, стоит повторение номера страницы прямо в гипертексте «2». Эта цифра, делающая смену страниц явной, означает, что абзац, прерванный страницей, никогда не будет найден как цельная цитата в какой-либо поисковой системе. Если же страница разрывает переносом слово, то и оно не будет найдено среди окружающего авторского текста. Присмотримся ещё раз к одному уже цитировавшемуся фрагменту:

... порабощает  
  
4  
  
В. И. ЛЕНИН  
  
целый ряд ...

Имитация бумажного издания, выражающаяся в центрированной надписи « В. И. ЛЕНИН» вблизи чётного номера авторской страницы и названия работы вблизи нечётного номера страницы абсолютно излишня. Она не выражает ни авторский текст, ни необходимый справочный аппарат, а потому подобные надписи должны быть безжалостно проигнорированы при составлении базы данных. Однако преобразование на стыке страниц доставляет немало хлопот. Товарищами было задано несколько вопросов.

**Вопрос: Как представить абзац на стыке страниц?**

Ответ: В таблице “Paragrafoj - paĝoj” код абзаца упоминается два раза, рядом с кодом одной страницы и рядом с кодом следующей страницы.

**Вопрос: Как хранить в схеме Фроша-Загорского-Радкевичюте место разрыва страниц?**

Ответ: FZR-схема оставляет этот вопрос открытым. Заседание Товарищества выявило несколько возможных способов решения проблемы.

1. Сохранение гиперссылок с номером страницы на первую букву или первый пробел новой страницы. Цитированный фрагмент внутри одной из записей таблицы абзацев будет выглядеть так:

… порабощает целый ряд …

Недостатком этого способа является избыточность номеров страниц, которые, помимо таблицы “Libroj - paĝoj”, повторяются в хранимом гипертексте. При условии того, что авторский текст не меняется такие гиперссылки принципиально допустимы, хотя похожи на избыточные.

2. Внедрение непечатного символа, который игнорируют алгоритмы поиска. В таком случае алгоритм формирования эталонного гипертекста обнаруживает этот символ и заменяет его на наименьший номер страницы, который извлекается для абзаца, следующего за данным. Таблица Уникода содержит довольно много непечатных символов, которые игнорируются популярными поисковиками. Выбор конкретного универсального символа должен быть согласован со знатоками этой таблицы (эта задач пока что не решена).

3. Полное уничтожение в хранимом абзаце следов смены страницы. Цитированный фрагмент внутри одной из записей таблицы абзацев будет выглядеть так:

… порабощает целый ряд …

Однако при этом в таблице “Paragrafoj - paĝoj”создаётся необязательное свойство для указания позиции первого символа страницы в данном абзаце. У абзацев из середины страницы это свойство оставляется без заполнения.

**Описание общего алгоритма разложения источника текста в базу данных**

1. Ручным способом вводятся в разные таблицы сведения об издательстве, авторе и книге. Помимо времени жизни и расположения издательства это обязательные названия и имена на эсперанто и рекомендуемые на мировых языках теоретического мышления - deu, rus, esp. В результате база данных даёт UUID конкретной бумажной книги, который будет использоваться далее во всех уместных случаях.

2. Для известного кода книги формируется список страниц, повторяющий бумажное издание. Номерами страниц заполняется исключительно таблица “Libroj - paĝoj”. Каждая страница получает свой UUID. В каждой записи о странице указана к какой книге она относится.

3. По довольно сложным признакам определяется начало авторской работы. Эти признаки абсолютно нестабильны и сильно зависят от источника гипертекста. **Авторская работа понимается как блок авторского текста с авторскими предисловиями и послесловиями**. В нашем случае признаком начала работы будет разметка: «ГОСУДАРСТВО И РЕВОЛЮЦИЯ». Главы и другие единицы внутри авторской работы не получают никакого специального выражения, отличающегося от того, что уже есть в большинстве гипретекстов-источников. Схема Загорского-Фроша-Радкевичюте не хранит внутренних оглавлений работ, хотя и не затрудняет их полностью автоматическое составление.

4. Получив название работы, программа для размещения в базе данных должна потребовать обязательного перевода названия работы и её подзаголовков на эсперанто и, в случае с ленинскими сочинениями, обязательного перевода названия на немецкий и испанский языки. Это необходимо для того, чтобы позднее товарищи могли найти основу для сверки или составления переводов на эти мировые языки теоретического мышления. После занесения названий работы, её код из таблицы “Laboroj” будет сопровождать каждый составляющий её абзац.

5. Разбор абзацев и подтверждение их качества это наиболее трудоёмкая работа. Определитель абзаца должен уметь относить абзац к уже зарегистрированной ранее странице. Если же абзац примыкает к границе страниц, то пользователь должен иметь возможность вручную определить продолжается ли абзац на новой странице. Если абзац связан с двумя страницами, то этот факт тоже должен получить отражение. Если же на новой станице начинается новый абзац, то особенностей работы алгоритма не возникает.

6. В гипертексте абзаца идёт поиск ссылок на авторские и редакционные примечания. Их содержимое также помещается в таблицу абзацев как независимая цепочка абзацев. По результатам связывания в таблице “Paragrafoj - notoj” остаётся связь абзаца и первого абзаца в составе примечания к данному абзацу авторского текста. Примечания должны вручную размечаться текстологом как авторские или редакционные. Для этого в схеме Фроша-Загорского предусмотрено специальное свойство.

#### О результативности текстологической базы данных и её форме

Понимание текстологической работы вокруг схемы Фроша-Загорского-Радкевичюте именно на уровне базы данных принципиально именно потому, что результативность базы данных может принимать самые разнообразные формы. Помимо полуавтоматической подготовки бумажных изданий для печати это ещё поиск цитат, библиографических ссылок, облегчение новых переводов и наиболее лёгкая проверка качества и покрытия существующих, автоматический лингвистический и грамматический анализ и многое другое. Все эти разные виды библиографической и лингвистической работы значат весьма мало ровно до тех пор, пока не определена и не реализована сборка ***авторской эталонной формы гипертекста***.

Как примерно выглядит эта форма? Должно быть понятно, что пользователь базы данных желает иметь дело либо с отдельной работой, либо с отдельной книгой. Поскольку библиотеку гипертекстов, вероятнее всего, реализуют для веб-сайта, то верхний блок со своими правилами оформления относится к сайту. Ниже начинается блок заголовков бумажного издания (если запрашиваемая работа имеет бумажную форму). Ещё ниже располагается блок заголовков конкретной работы. Ниже этого блока авторский текст со всеми особенностями оформления. Оформление блоков заголовков для книги и работы происходит по особым правилам и не подчиняется текстологической таблице стилей. Это легко понять, если вспомнить, что автор в общем случае не определяет особенности заголовка бумажного издания.

*Как оформляются блоки заголовка бумажного издания и отдельной работы?*

Заголовок книги начинается с алфавитного перечня авторов языком оригинала. Затем этот перечень повторяется на эсперанто другим цветом или уменьшенным шрифтом. Ниже идут поочередно перечни авторов на мировых языках теоретического мышления и на языке пользователя. Если языком оригинала является эсперанто, то текста со стилем особого названия на эсперанто не будет, если же работа создана на одном из мировых языков теоретического мышления (а для работ Ленина это справедливо), то соответствующий блок будет состоять из двух, а не трёх строк. Наконец, если пользователь предпочитает эсперанто или один из мировых языков теоретического мышления, то блок списка авторов на пользовательском языке, очевидно, не появляется.

За списком авторов следует блок разноязычных названий бумажного издания, где надписи следуют в том же самом порядке (оригинал->эсперанто->мировые языки теоретического мышления->пользовательский язык). Ниже названий указывается год издания и название издательства в том же самом языковом порядке для нескольких языковых вариантов.

Таким образом для бумажных изданий и названий работ устанавливается обязательный перевод на эсперанто и три указанных ранее мировых языка теоретического мышления. Для нескольких тысяч работ Ленина это выглядит непосильной задачей, но иного способа обеспечить быстрый поиск для международного сотрудничества в области текстологии и теории не существует. К сожалению современное теоретическое положение России таково, что изучение языка Ленина не является особенно популярным. Однако это дело может продвинуться, если в Германии или в США заинтересованные товарищи будут иметь как базу для новых переводов лучший академический гипертекст работ Ленина и Чернышевского. В таком случае, имея нечто с несомненным теоретическим значением для понимания современной общественной борьбы, товарищи из других стран будут иметь наибольшие побуждения к проникновению в оригинальные тексты и составлению переводов, а, значит, и к изучению языка Чернышевского и Ленина.

#### Как собирается блок авторского гипертекста?

Блок авторского гипертекста - это самый очевидный и самый ожидаемый результат работы текстологической базы данных. Этот блок всегда основан на структуре документа на языке оригинала. В общем случае пользователь указывает код документа на любом доступном языке и небольшой список языков для формирования параллельного текста. Даже при такой задаче формирования параллельной таблицы абзацев алгоритм оказывается довольно прост. Его основные этапы будут аннотированы ниже.

1. По коду работы определить работу-оригинал, являющуюся авторитетной основой для переводов (из таблицы оригиналов произведений). Может оказаться, что пользователь уже указал оригинальную работу.

2. Определить первый абзац оригинальной работы (из таблицы последовательности абзацев). Признаком первого абзаца является то, что он указан рядом с выбранным кодом оригинальной работы, но ни разу не упомянут как следующий за каким-либо (то есть отсутствует среди значений свойства Sekva). От первых абзацев многоабзацных примечаний первый абзац авторского текста отличается тем, что для него невозможно найти (в таблице абзацев-примечаний) исходного абзаца. То есть первый абзац авторского текста не является примечанием для какого-либо другого абзаца.

3. Для каждого абзаца оригинального текста собирается группа переводных абзацев в соответствии с указанным пользователем перечнем языков.

4. Для абзаца оригинального текста ищется следующий абзац и потом добавляется его текст и набор его переводов. Это повторяется до завершения авторского оригинального текста.

5. Формируется блок примечаний. Это формирование само по себе сложно для понимания тому, кто никогда не видел многоязычных книг. Однако интуитивно должно быть понятно, что по каждой группе абзацев-переводов собирается перечень примечаний, которые идут в ряд. Для авторских примечаний должно быть особо отмечено отсутствие их переводов, тогда как редакционные примечания без соответствия на других языках можно никак не обозначать. Если хоть к одному абзацу из полученных переводов есть примечание, то оно образует самостоятельный ряд в блоке примечаний. Примечания к другому абзацу или другой группе абзацев-переводов образуют другой ряд.

Вопрос о том, отделяется ли блок редакционных примечаний от блока авторских примечаний является спорным. В общем случае авторские примечания удобнее дать отдельным блоком с полным набором переводных соответствий, тогда как редакционные примечания нередко будут выглядеть разреженной таблицей в силу уникальности языковых условий и необходимых исторических пояснений.

Предложенный принцип разделения авторских и редакционных примечаний не имеет проверенного международного значения. Он подлежит уточнению и попыткам опровержения, поскольку пока что это всего лишь вывод из польской издательской традиции, в частности, применённый в сочинениях Ленина под редакцией Ярослава Ладоша.

Предложенный алгоритм для критической оценки полезно проследить по диаграмме связей таблиц в FZR-схеме:

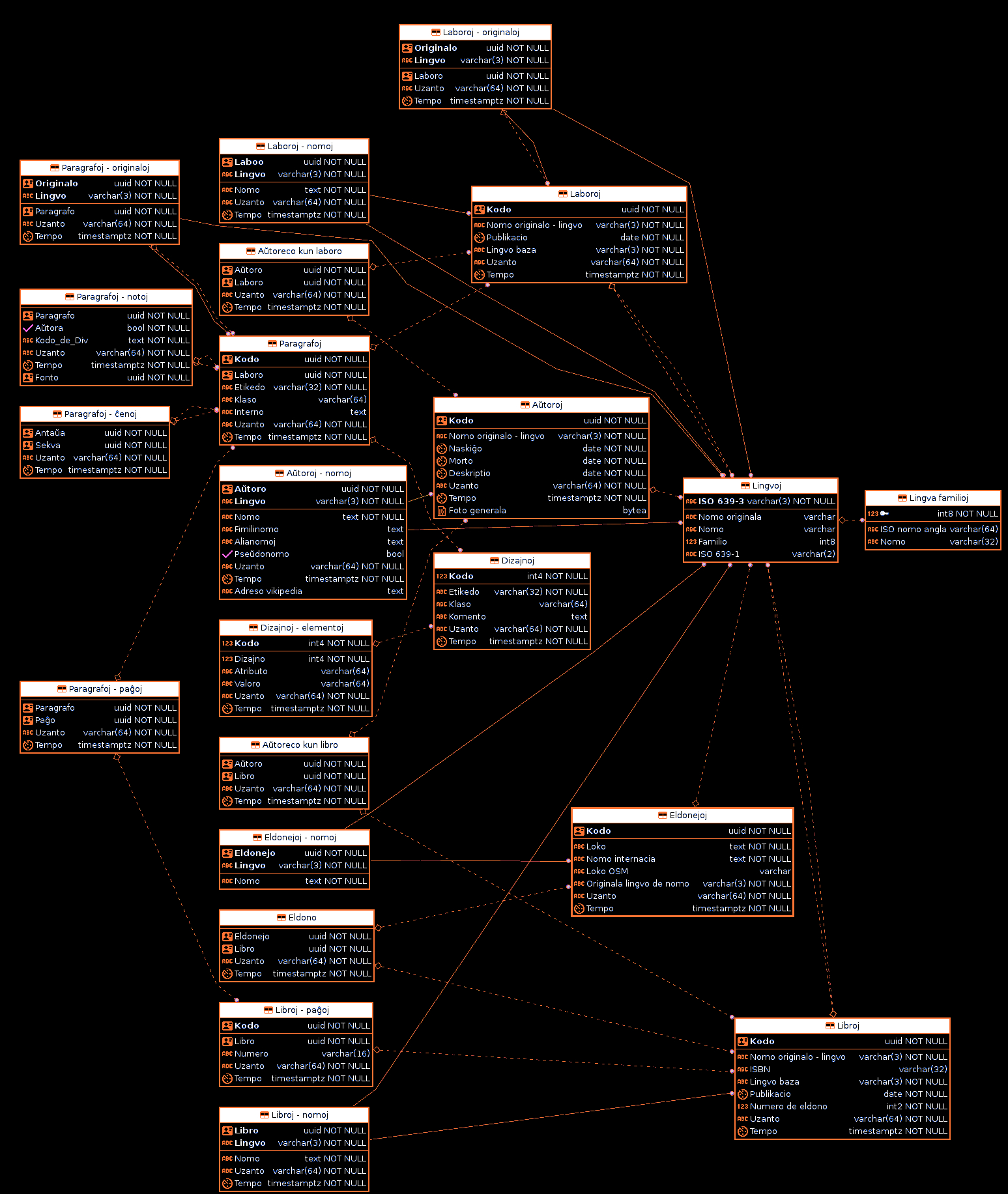


Диаграмма базы данных

### Вопрос 25

**Чем вызвано применение UUID вместо обычных кодов-номеров во всех изменяемых таблицах?**

Ответ: Создавать и наполнять базу данных по FZ-схеме могут независимо один от одного разные люди, разные текстологические и/или самообразовательные группы. Отсутствие нумерации целыми числами облегчает объединение таких независимо наполняемых баз данных в единое хранилище общего результата, если для этого создадутся необходимые условия. Или, что тоже самое, устраняется значительная часть программистской работы по объединению баз данных так, что из задач общего программирования остаётся только поиск дубликатов в перечнях авторов/издательств/изданий/работ и прочих.

### Вопрос 26

**Чем обусловлен выбор PostgreSQL среди прочих программ, управляющих коллективными реляционными базами данных?**

Ответ: Это решение было принято на мартовском (2016 г.) консилиуме в Берлине и соответствует местным техническим традициям. Отвечает т. Фрош:

«Рациональное обоснование, выдвинутое именно для PostgreSQL, состоит в наличии системы уведомлений, о которой сообщил небезызвестный depesz. В наших немецких условиях существует некоторая неприязнь программистского и кибернетического сообщества, которая определила направление на минимизацию программистского труда при работе с текстологической базой данных. Система сообщений, скрывающаяся за командами [LISTEN](https://docs.postgresql.fr/11/sql-listen.html) и [NOTIFY](https://docs.postgresql.fr/11/sql-notify.html), позволяет значительно сократить вероятность конфликтов и ненамеренной перезаписи чужого результата обработки абзаца. Так, коды абзацев, с которыми ведётся работа могут рассылаться всем работающим в данный момент текстологам с тем, чтобы предупреждать о необходимости согласовать редактирование именно данных абзацев, поскольку кто-то ещё может создавать новый проверенный вариант текста. С точки зрения программирования система уведомлений это незаменимый способ инициировать событием в базе данных какие-либо уведомления или сообщения пользователю».

### Вопрос 27

**Чем обусловлен выбор языка эсперанто для названия объектов FZ-схемы?**

Ответ: Это решение было принято на мартовском (2016 г.) консилиуме в Берлине. Отвечает т. Фрош:

«В наших немецких условиях существует некоторая неприязнь программистского и кибернетического сообщества. Я полагаю, что в целом сфера баз данных намного более демократична, чем сфера программирования, поскольку больше ориентирована на предметный смысл и меньше на технические особенности. Именно это на Консилиуме определило выбор названий объектов проектируемой текстологической схемы (ныне FS-схемы) на языке эсперанто. ***Преимущественными потребителями содержимого текстологической базы данных должны быть не программисты, а участники самообразования, то есть текстологи-любители***. В различных графических программах-клиентах баз данных, переводимых обычно на родной язык, появляется необходимость отображать название таблиц и их свойств. Идея иметь несколько языковых версий или несколько языковых оболочек для базы данных не показалась здравой. Но понятность текстологам-любителям была необходима. На этом держали veto все иностранные участники Консилиума - из Латинской Америки, Франции, Польши, Чехии и Болгарии. Как не самая удачная (из-за лексического европоцентризма), но, безусловно, понятная и распространённая вспомогательная система был выбран язык эсперанто, обеспечивающий максимальную понятность, равную сложность и равную простоту для носителей трёх мировых языков теоретического мышления.

Эсперанто был выбран на Консилиуме (не получив ни одного veto) исходя из фактических мировых языков теоретического мышления и возможного добавления англоязычных и китайскоязычных товарищей. Эсперанто это отличная вспомогательная языковая система для наиболее быстрого изучения. Хотя лексика эсперанто далека от любой литературной китайской нормы, этот язык, несомненно, осваивается местными товарищами легче, чем любая другая сильно европеизированная лексическая система».

**Вопрос 28: Как решается вопрос пользовательских надписей в текстологических программах, работающих с FZR-схемой? Где должны быть собраны названия библиографических свойств, надписи об отсутствии перевода и пр.?**

Ответ: Самым логичным способом будет разместить комплекты переводов в специальной таблице баз данных «Mesaĝoj de programoj». Внутри таблицы разместить свойства «Kodo» тмпа uuid, «Lingvo» типа char(3) и «Teksto» типа text. Ключом таблицы объявляется UUID код сообщения вместе с трёхбуквенным кодом языка. То есть зная код сообщения и код языка можно найти текст. Суть использования таблицы весьма проста. При необходимости составлять сообщения для программ создаётся новый UUID, который записывается в свойство «Kodo» как минимум четырежды и связывается с текстами на эсперанто и трёх мировых языках теоретического мышления. Программисту, таким образом, нужно знать UUID сообщения, чтобы пользователю была дана надпись на запрошенном языке или на одном из имеющихся языков. Однако принципиально, что таблица «Mesaĝoj de programoj» никак не связана с текстологической работой, она носит независимый и справочный характер, облегчая составление и наладку многоязычных программ для программистов.

### Вопрос 29

**Почему в публикациях мало упомянуты проблемы текстологического разбора в базу данных письменного наследия Николая Чернышевского?**

Ответ: Главным образом потому, что из 15 томов наиболее авторитетного бумажного издания [едва ли больше двух томов покрыты приемлемым результатом распознания](http://ngchernyshevsky.ru/works/16/), не говоря о гипертексте, хотя бы и ненормальном. Таким образом, в отношении наследия Чернышевского российские текстологи освобождены разве что от утомительной необходимости делать фотокопии и получать первый вариант автоматического распознания. Остальные же текстологические работы ждут своих исполнителей. Сделать предстоит немало: сверить пунктуацию и исправить ошибки распознания, сверстать распознанные блоки текста в отдельные работы и, вместе с академическим аппаратом, создать каждой работе гипертекст, пригодный для публикации в базе данных. Все названные работы выходят за пределы того, что доступно западнее Днепра и Нарвы как по языковым причинам, так и по причинам трудоёмкости. Работа с письменным наследием Ленина, которое в целом уже неплохо нормализовано в текстологическом смысле, должна по указанным причинам (в сравнении с нормализацией произведений Чернышевского) рассматриваться как проба. Но так же, как важнейший участок получения лучшего теоретического наследия России всеми внешними для её самообразовательных сообществ товарищами.

### Вопрос 30

**Каковы легальные условия для текстологической работы над ленинским наследием?**

Ответ: В континентальной Европе нам неизвестны случаи полицейских преследований за какую-либо текстологическую работу вообще и в отношении ленинского письменного наследия в частности. В некоторых штатах Индии, охваченных активной текстологической, переводческой и самообразовательной работой, работа с произведениями Ленина условно нелегальна. Очевидно, что местные товарищи должны будут обеспечить особые условия своей базе данных в том случае, если будут иметь желание и возможность вносить вклад. В континентальной Европе до сих пор допустима не только публикация результата текстологических работ в базе данных, но и гласная работа в ней, если иметь ввиду перспективу размещения письменного наследия Ленина и Чернышевского.

## Часть V. Об организации текстологической работы

*Towarzystwo cybernetyczne*

[Источник](http://propaganda-journal.net/10421.html).

Некоторое время назад к проблемам текстологии и, в частности к международному европейскому опыту текстологических работ, стали проявлять интерес не только переводчики, любители даталогии[[13]](#footnote-95), участники самообразования по теории марксизма и/или базам данных, но и практикующие текстологи. С нами связались участники текстологической группы, работающей над ресурсом [многоязычной библиотеки классической теоретической, научной и чувственной (художественной) литературы](http://sokrat.online/). Поскольку товарищи уже столкнулись с теми же самыми проблемами, что и польские текстологи начала 2000-х годов, то нам был передан меморандум[[14]](#footnote-97), затрагивающий некоторые вопросы организации и способа проведения текстологических работ. Мы считаем необходимым публично прокомментировать этот документ из восьми параграфов и предать гласности некоторые места, не содержащие специальной терминологии и/или технических жаргонизмов.

В §4 меморандума превосходно сформулированы упоминаемые в широком международном и историческом контексте главным мысли целого ряда статей по текстологии, которые были опубликованы ранее на «Пропаганде»:

… *возникает необходимость проведения международной текстологической работы, что и нашло отражение в* [историко-логическом изложении Фроша - Загорского](http://propaganda-journal.net/10319.html) *и* [технологическом проекте Фроша - Загорского - Радкевичюте](http://propaganda-journal.net/10402.html)*, из которого мы уяснили:*

*во-первых,* ***попытки создания текстологического фонда предпринимались и не раз, но не были доведены до ума, поскольку это требует международной совместно-разделённой деятельности****;*

*во-вторых, их* ***исполнение не отвечало как техническим требованиям, так и требованиям текстологии, на преодоление чего и направлена*** [техническая часть проекта Фроша - Загорского - Радкевичюте](http://propaganda-journal.net/10402.html) *…*

Вслед за этим изложением, выражается обеспокоенность

… *в-третьих, что существенно, остаётся открытым вопрос и о критерии подбора литературы для организации работы в теоретическом сообществе, особенно что касается пропедевтики и дидактики для овладения теоретическим способом мышления, как инструментом революционной практики - диалектикой, логикой и теорией познания.*

Что касается вопросов дидактики и пропедевтики, то они принципиально были оставлены за пределами текстологического проекта. Впрочем, должно быть понятно, что хорошо исполненный международный текстологический проект должен быть пригоден именно для этих целей в первую очередь и, лишь после этого, для издательских и академических целей. Каковы должны быть критерии начального этапа? Не будет ошибкой указать, что жизненно важно в самом начале, чтобы состоялась ***международная*** ***совместно-разделённая*** ***деятельность*** (как это удачно названо в меморандуме). Каковы перспективы такой деятельности? Нужно сказать, что едва ли будет полезен громоздкий текстологический план товарища Петра Стрэнбского, который [был опубликован](http://propaganda-journal.net/10316.html) другим польским текстологом - Домиником Ярошкевичем. Громоздкость этого плана сыграла в долгосрочной перспективе очень плохую роль, поскольку отсутствие сильного пропедевтического (популярного и завлекающего) компонента фактически остановило текстологическую работу в Польше спустя 12 лет после её триумфального начала[[15]](#footnote-100).

Некоторая близость популярных, как бы сказали ранее, издательских планов несомненно должна сохраняться в наше время[[16]](#footnote-101) для украинских, польских и российских условий исходя из общего полуимпериалистического положения Польши и России. В особенности небольшое отличие польской, российский и украинской пропедевтики и дидактики происходит из общего упадка промышленности и общего политического подавления со стороны транснациональных корпораций, связанных с политическим аппаратом США, которое особенно ярко проявляется на Украине. Эта страна тем менее отличается в пропедевтическом и дидактическом плане от соседей, чем более демонстрирует возможное будущее как польского, так и российского общества в случае продолжения господства всем хорошо известных упадочных тенденций материальной и духовной жизни.

Словом, не является проблемой понимание того, что для начала международной совместно-разделённой деятельности должна совпасть некоторая пропедевтическая потребность. Естественно ожидать, что в базу данных по FZ(R)-схеме сначала должно попасть нечто более-менее нормализованное и небольшое. Рискнём выдвинуть гипотезу, что разобраться в работе базы данных поможет размещение статьи Марека Яна Семека «Почему Сократ?»[[17]](#footnote-103), которая уже имеется в оригинальной аномальной (PDF) и российской нормализованной (html) версии. Полный трёхъязычный цикл с базой данных включает в себя выработку авторской карточки, книжной карточки и карточки статьи с обязательным переводом названий на эсперанто и мировые языки теоретического мышления. Затем на примере размещения вновь создаваемого украинского перевода можно будет понять принципы облегчения переводческой работы, которые реализованы в FZ-схеме.

Выраженное в меморандуме стремление текстологов-практиков к пониманию принципов FZR-схемы базы данных довольно показательно и весьма удобно для того, чтобы ещё раз противопоставить ясное изложение нашей позиции различным возражениям, которые не перестают поступать даже после многократного повторения основополагающих принципов международного текстологического проекта Фроша-Загорского-Радкевичюте. Итак, §5 меморандума сообщает нам:

*§5.* [На Сократе](http://sokrat.online/)*, вначале работы мы акцентировали работу над переводами текстов и у нас не было представления о технической стороне дела, ни как работать с текстами, ни как оформить сайт - библиотеку, в лучшем случае она была дилетантской. Максимум что мы достигли - это книги в формате качественного ПДФ формата, образец «Капитал» и вообще* [работы Маркса в украинском переводе](http://sokrat.online/files-view-20.html)*. Позже мы убедились, что так вначале были сделаны книги и в* [библиотеке Гутенберг](https://uk.wikipedia.org/wiki/Проект_«Гутенберг»)*.* [Полемика Фроша - Загорского](http://propaganda-journal.net/10319.html) *поставила ребром вопрос о подготовке книг для фонда, которые отвечают требованиям текстологии - оригиналу, в формате html, в крайнем случае, для нас.*

Если обращаться к международному текстологическому опыту, то сложно не заметить, что именно в названном направлении развивалась текстологическая работа в Варшаве и в Берлине. Ещё около 12 лет назад собственноручно стандартизованные PDF документы готовила текстологическая команда под руководством Петра Стрэнбского. Эта работа, однако, не получила должного технического развития, но подобные же проблемы вынуждены были решать немецкие товарищи. Так, несколько лет назад взамен PDF форм стали публиковать базовый гипертекстовый вариант издатели известного немецкого журнала [RotFuschs](http://www.rotfuchs.net/). Необходимость универсализации текстовой формы классических произведений в уникодовый[[18]](#footnote-107) гипертекст западнее Буга не только признаётся в настоящий момент почти всеми текстологами, но и стихийно реализуется по мере сил уже несколько лет. Проблемы наших товарищей на этом пути указаны в §6 меморандума.

*§6. …поскольку нам необходимо готовить тексты в ворде и делать переводы на украинский, остановились на следующем варианте подготовки гипертекстов . При этом мы делаем два формата - оставляем ПДФ как образец оригинала для сверки и делаем гипертекст в формате html. <…> нам необходимо понимать, соответствует ли наша работа требованиям текстологии для подготовки текстов для хранилища? Если нет то как это должно выглядеть, с учётом того, что нам необходимо готовить тексты для работы переводчиков.*

Первое замечание касается того, что в западноевропейской текстологии фактическим стандартом примерно с 2001-2003 года является не Word (имеющий нестабильный и избыточный выход гипертекста), а редактор из пакета OpenOffice/LibreOffice под названием Writer. Этот редактор имеет полную поддержку ведению таблицы стилей в каждом подготавливаемом документе. В том числе поэтому он обеспечивает качественный выход гипертекста при условии, что выполняется нормальный текстологический цикл:

1. Восстановление авторских абзацев и пунктуации
2. Нейтрализация текста (Ctrl+M) то есть снятие стилей оформления
3. Восстановление *авторской* разметки (курсивы, подчёркивания, разрядки и пр.)
4. Отнесение абзаца к стилю оформления. Обычно это происходит через отнесение абзаца к авторскому типу (обычный текст, заголовок некоторого уровня, авторское примечание и пр.) или ко вспомогательному типу (редакционное примечание, сноска, цитата, подпись под иллюстрацией и пр.). При этом вспомогательные типы подлежат выносу в единую таблицу стилей текстологического хранилища (таблицы FZ-схемы со словом «Dizajnoj») и должны называться на языке эсперанто. Перечень вспомогательных типов пока что не утверждался и мы приглашаем всех заинтересованных к исследовательской работе по определению необходимых перечней вспомогательных стилей оформления для правильной передачи таких полиграфически сложных произведений как «Капитал».

Итак, прочитаем ещё раз как товарищи организовали контекстную текстологическую работу:

… *мы делаем два формата - оставляем ПДФ как образец оригинала для сверки и делаем гипертекст в формате html. <…> нам необходимо понимать, соответствует ли наша работа требованиям текстологии для подготовки текстов для хранилища?*

На поставленный вопрос можно ответить утвердительно. Потенциально PDF образец, совершенно разумно используемый для сверки, пригоден для описания в т. н. таблицах Радкевичюте (номер страницы оригинала и графические координаты в фотокопии для каждого абзаца). Важнейшей функцией PDF прототипа является также отнесение абзаца к странице, которое должно выполняться для всякого издания, признаваемого важным или авторитетным в результате обилия библиографических ссылок. Издания, не получившие широкого распространения в бумажном виде, однако, едва ли стоит дополнять постраничными сопоставлениями абзацев в FZR-схеме. В таком случае единственной системой публичных ссылок будет та, которая будет основана на UUID абзацев в конкретной базе данных. В общем случае подобные абзацные ссылки точнее страничных, хотя применение UUID делает их, по словам товарища Фроша, устрашающими.

В §7 меморандума товарищи рассказывают о типичных проблемах полного текстологического цикла, который на жаргоне называется циклом Радкевичюте[[19]](#footnote-108).

*§7. <…> о мелочах, которые усложняют нашу работу и которые мы не понимаем, как преодолеть, в том числе как минимизировать ручной труд в процессе сканирования, конвертации и подготовки текста книги для печати.*

Разберём указание на проблемы по этапам.

**Сканирование**, то есть создание нормализованной фотокопии. Этот процесс с точки зрения даталогического описания затронут в [реплике по поводу полемики Фроша и Загорского](http://propaganda-journal.net/10319.html). Каковы перспективы автоматизации этого процесса? Для небольших самообразовательных сообществ таких перспектив нет. Даже при условии промышленной организации всех иных стадий текстологической работы (что под силу любителям), автоматизировать сканирование практически невозможно. При разговоре об автоматизации сканирования невольно вспоминается арендованный немецкими товарищами полиграфический зал со сканером, где гильотиной уничтожали переплётную часть какого-то тома ленинских сочинений, а потом один товарищ раскладывал листы по порядку на широком столе, а другая товаришка по очереди выравнивала их в камере сканирования и складывала в архивную стопку. Подобный стиль работы, уничтожающий бумажное издание является предельным с вещественной стороны. Однако и с точки зрения базы данных фотокопия, например, рукописи, это предельная основа. На этом заканчивается техническая текстология произведений Нового времени. В общем случае текстологическое исследование заканчивается фотокопией достоверной рукописи на языке оригинала как первым объективным свидетельством. Это означает одновременно и то, что FZR-схема способна нести поабзацное описание до уровня авторской рукописи, и то, что в общем случае нет необходимости размещать в базе данных FZR-схемы все произведения с оригинальной графической основой из фотокопии рукописи.

Отдельные jpg или tiff файлы, накапливаемые в результате сканирования уже представляют из себя несомненный текстологический интерес и могут устанавливать связь с базой данных схемы Фроша-Загорского-Радкевичюте. Как это происходит практически?

1. Определяется некий единый каталог для хранения результатов сканирования. В этом каталоге создаётся каталог конкретного бумажного произведения письменности. Его целесообразно назвать по названию произведения с указанием автора. Так каталог хранения постраничных фотокопий работы http://hegel-marks.pl/downloads/teksty-siemek32.pdf можно назвать «Dlaczego Sokrates? Marek Jan Siemek».

2. В каталоге конкретного произведения размещаются пронумерованные любым очевидным способом файлы постраничных фотокопий.

3. Постраничные фотокопии подвергаются первичному распознанию программами tesseract и/или cuneiform, которые доступны как в RedHat ветви дистрибутивов (например, в Fedora) так и в Debian ветви дистрибутивов (например, в Ubuntu). Эти программы содержат шаблоны распознания для значительного числа европейских языков, в том числе, для украинской и болгарской кириллицы. Формат вызова tesseract довольно прост

tesseract файл-фотокопия файл-результат -l код\_языка

Детальную справку[[20]](#footnote-109) даёт вызов программы без параметров: tesseract.

Суть применения программы tesseract довольно проста: на основе шаблонов начертания букв и словаря нужного языка она строит текст, находящийся на странице. При этом в большинстве случаев правильно соблюдено расположение абзацев. Аналогично работает программа cuneiform, также имеющая модуль украинской кириллицы. В ручном и/или наладочном режиме работа с cuneiform и tesseract обычно ведётся через оболочку yagf или аналогичные ей.

Весьма интересен у программы cuneiform формат вывода [hOCR](https://uk.wikipedia.org/wiki/HOCR), который позволяет сохранить в гипертексте ссылку на графическое расположение конкретного слова. После упрощения hOCR документа возможно получение нормализованного гипертекста и т. н. таблиц Радкевичюте[[21]](#footnote-111). В частности для абзаца, не пересекающего страницу нужно вычислить наименьшую и наибольшую координаты по вертикали и по горизонтали. В большинстве европейских текстов в пределах указанного полигона будет расположен весь текст, а, в случае разделения абзаца между страницами, таблицы Радкевичюте будут хранить два полигона для одного абзаца на разных страницах. По результатам майского (2019) консилиума задача преобразования [hOCR](https://uk.wikipedia.org/wiki/HOCR) во-первых в текст-кандидат для вычитки и во-вторых в описание полигонов абзаца была оценена как задача общего программирования высокой сложности (до нескольких сотен строк Java кода). Однако несомненно, что эта задача поддаётся автоматизации и вполне разрешима в обозримой перспективе.

Завершив изложение лучших практик, связанных со сканированием перейдём к следующему этапу. Трудности этапа **конвертации** изложены в §7 меморандума следующим образом:

*… Во-первых, для дальнейшей работы в Ворде, т. е. подготовки текста к переводу и заодно и гипертекста, мы бросаем его в Notepad++ после конвертации после сканирования или «передирания» с ПДФ. При этом сталкиваемся с тем, что предложения дробятся на части. Их восстановление и, заодно, проверка орфографии требуют много ручной работы.*

Итак, товарищами указано два принципиально разных источника:

1. Существующие PDF, имитирующие бумажные издания или не имитирующие их, но не содержащие связного текстового слоя.

2. Фотокопии

Общим дефектом является нечёткая абзацная структура. В результатах распознания фотокопии кроме того, нередки орфографические дефекты. Готовых средств автоматизации текстологической работы по преобразованию фотокопий или аномальных (PDF) текстов в нормальный гипертекст нет. Однако ноябрьский (2018) консилиум выявил, что в большинстве случаев специальные редакторы типа Notepad++ излишни. По сути задача может решаться простым окном редактирования (реализованном хотя бы на языке Java или Python), которое выполняет функции абзацного разбиения и проверки орфографии[[22]](#footnote-112). Так, элементарная проверка по словарю реализуется сразу двумя словарными группами программ hyspell и aspell, причём каждая группа включает словарь украинской, немецкой, российской и польской лексики[[23]](#footnote-113).

Задача абзацного разбиения реализуется на любом языке программирования довольно просто. Установив курсор в нужное место раздела абзацев, можно вставить туда любой необычный символ, не встречающийся в авторском тексте. Обычно вставляется т. н. [вертикальная табуляция](https://unicode-table.com/pl/000B/) или её аналог «␋», бывает что знак абзаца «¶». Довольно простая программа, получив сигнал разбора текста, автоматически может отделить текст указанного абзаца от фрагментов следующего абзаца. При этом также происходит уничтожение внутри текста абзаца переносов, табуляций и кратных пробелов. Как правило, ничто из названного не является элементом авторского текста и потому подобная автоматизации вполне допустима для 98-99% реально встречающихся абзацев на европейских языках и не создаёт лишних проблем за пределами встречающихся в книгах таблиц[[24]](#footnote-115).

Пока изложенная несложная логика не реализована в готовом программном средстве, стоит указать, что существует примерная реализация обсуждаемых функций средствами операционной системы Unix-типа. Так в Германии для автоматизации зачистки текста от дефектов распознания (лишних табуляций, пробелов, аномальных тире, неуместных знаков переноса и пр.) обычно используются однострочные команды замены, передаваемые старому текстовому редактору под названием sed. Эти команды простой замены осуществляют склеивание строк из PDF, когда они уже вынесены в отдельный текстовый файл Уникода. Производимая замена обычно обеспечивает устранение переносов, кратных пробелов и иных дефектов исходя из потребностей автора. Замены в sed весьма просты и легко настраиваются на устранение типовых дефектов а также довольно легко понимаются. Например, простые замены могут использоваться даже для полной замены системы письма в тексте[[25]](#footnote-116).

Добавлением к командам замены в sed обычно служит графический текстовый редактор (gedit/pluma/xed или аналогичный) с проверкой слов по словарям aspell или hyspell. Вероятно мало известный в европейских текстологических кругах Notepad++ также может быть чем-то полезен на данном этапе.

Вслед за формальной подготовкой исходного текста обычно следует содержательная подготовка или восстановление корпуса авторских примечаний и академического аппарата. Трудности этапа **подготовки произведения к публикации** («подготовки книги к печати») изложены в §7 меморандума наших товарищей следующим образом:

…*сталкиваемся с тем, что предложения дробятся на части. Их восстановление и, заодно, проверка орфографии требуют много ручной работы. Тоже касается обычно и ссылок, когда цифры в Нотепаде идут строкой выше, а предложение ниже.*

В том случае, если цифры сносок или степеней, ошибочно помещаются в отдельную строку, то этот дефект считается для текста PDF происхождения органически неустранимым, а для текста, распознанного из фотокопий это считается результатом неверной настройки программы распознания. Нередко для типовой серии фотокопий страниц перед отправкой на tesseract или cineiform производится весьма затейливая серийная обработка программой [imagemagick](https://imagemagick.org/index.php), которая позволяет массово изменить резкость, контрастность, формальную детализацию и многое другое. Обычно в промышленных проектах обработка imagemagick и cuneiform или tesseract соседствуют в одном sh файле, где указываются подобранные на основе опыта параметры для всей серии фотокопий страниц.

§7 меморандума указывает и на другие технологические проблемы текстологической работы. В частности попытки продвижения в область специальной литературы привели к столкновению с проблемами, которые описаны весьма эмоционально:

… *Третье. В книгах Ведуты или Глушкова (а это только цветочки, есть еще «Энциклопедия кибернетики», первая в мире и на украинском языке, она вышла в 70-е годы, и т д.) есть очень много формул. Снова ручной труд?*

Короткий ответ: Да. Системы распознавания формул находятся в настоящий момент в зачаточном состоянии. Отсутствует текстологическая стандартизация представлений формул за исключением того, что они безусловно считаются авторским текстом, а следовательно, должны выражаться в гипертекстовом абзаце. В целях поиска не рекомендуется размещать формулы в виде изображений, однако конкретные требования к отражению взаимного расположения греческих и математических символов в гипертексте на настоящий момент не выработаны. Ясно только одно, что кодировка Уникод позволяет покрыть большую часть даже специальных математических трактатов. Например [один из буквенных диапазонов](https://unicode-table.com/pl/blocks/mathematical-alphanumeric-symbols/) вместе с [дополнительным диапазоном операторов](https://unicode-table.com/pl/blocks/supplemental-mathematical-operators/) позволяет писать весьма разнообразные формулы. Кроме того есть ещё 3 или 4 исключительно математических диапазона разного назначения. Формат хранения непосредственно формул в текстологической базе данных полежит всестороннему обсуждению. Текстологическая экспертиза возможности разбора в odt формат письменного наследия Конрада Цузе (основателя немецкой социалистической кибернетики) указала на уместность формата редактора формул LibreOffice Math, однако способ встраивания таких формул в гипертекст пока что не ясен, ибо до его обсуждения дело не дошло. Как видим, эта проблема находится на переднем крае текстологической технологии и каких-либо готовых средств ещё не выработано.

Помимо технологических вопросов, наши товарищи столкнулись и с даталогической проблемой, то есть с проблемой сохранения связи между академическим аппаратом и текстом. В §7 меморандума эта проблема выражена так:

… *При подготовке текстов античной литературы - Платона и Аристотеля мы столкнулись с тем, что в оригиналах на полях сделаны цифровые и буквенные ссылки. Но мы не понимаем, как их технически в формате html сохранить их для истории, тем более, что они связаны в оригинале с примечаниями!!!!*

Античная текстология в целом сильнейшим образом отличается от простой текстологии Нового Времени, имеющей обычно дело с авторитетным или оригинальным авторским прототипом. Разумеется, текстологическая база данных по FZR-схеме не предназначена для полноценного отражения связей античной текстологии. Однако в смысле публикационной витрины для античных произведений FZR-схема вполне способна выполнить функции текстологического фонда.

Конкретно проблема ссылок на полях может решаться двояким способом.

1. Все ссылки размещаются в виде невидимых ссылок, содержащих ближайший к месту их постановки пробел. В предыдущей части настоящих очерков подобным образом в ленинских сочинениях проставлялся невидимый, но обнаружимый разрыв страниц. Гипертекстовый формат невидимой ссылки таков: « ». В данном случае ссылка имеет название «s1». Невидимую ссылку можно использовать в качестве места, куда ведёт ссылка из примечания или комментария. Однако само примечание, как правило, лучше не отправлять из невидимой ссылки.

2. Помимо невидимой ссылки есть промежуточное решение, не ухудшающее поиск и текстологическое качество текста. Это невидимый комментарий «**<– текст –>**», который можно вставить в любое место (даже в середину слова). Потенциально этот комментарий должен быть преобразован в какой-либо явный вид либо программой автоматического разбора перед занесением в базу данных либо программой JavaScript, чтобы читатель увидел издательские пометки к том или ином виде. Упоминаемая программа JavaScript обычно создаёт узкий столбец сбоку от основного текста и приписывает в этом столбце все пометки напротив тех строк, где содержатся невидимые комментарии.

Несмотря на то, что невидимые ссылки и невидимые комментарии не позволяют показать читателю весь текстологический слой, они, даже в неполноценном виде, однако сберегают работу текстолога. Действительно, последующую работу с академическими и библиографическими пометками на полях можно будет автоматизировать через JavaScript для демонстрации читателю или через добавление соответствующих таблиц в FZ-схему, чтобы формирователь гипертекстов из базы данных автоматически правильно расставлял пометки. В любом случае работающие в контакте с нами авторы FZR-схемы открыты для диалога и готовы изменить спецификацию набора таблиц, чтобы она полнее соответствовала популяризаторским и просветительским нуждам.

§8 меморандума наших товарищей содержит довольно много специфических технических терминов, однако именно здесь ставится вопрос о связи оформления сайта и вообще механизма публикации гипертекстов с базой данных.

…  *удобный рубрикатор или каталог книг, тем более, что он должен быть на нескольких языках, или что ещё принципиально, это тематические подборки, для пропедевтики и дидактики <…> которые по идее, должны быть взаимосвязаны с каталогом или фондом, но мы не понимаем, как это сделать технически. Логически, если будет создан фонд теоретической текстологической литературы на различных языках мира, то это облегчит работу. Но эту работу надо делать параллельно и клонировать её.*

Приведённые строки содержат довольно много интересных и важных для использования текстологического фонда идей. Вот, скажем, каталог книг легко может быть автоматически сформирован в многоязычную таблицу вроде https://www.marxists.org/xlang/lenin.htm на основе таблицы «Laboroj» с добавлением перевода названия на запрошенный пользователем язык.

Как в таком случае организовать рубрикатор? Это тоже не составит большой трудности, если внедрить в базу данных таблицу-рубрикатор. Первая и недопустимо узкая реализация рубрикатора может быть представлена как таблица, где каждому UUID относящемуся к конкретному произведению будет сопоставлено название рубрики. В этой таблице из двух полей вполне естественно можно повторять один и тот же UUID произведения несколько раз вместе с разными названиями рубрик и наоборот, название рубрики в связи с несколькими произведениями. Однако *многоязычный рубрикатор не может быть реализован единственной таблицей*. В этом случае должен быть сформирован 1) каталог рубрик из единственного свойства типа UUID, 2) таблица переводов названия рубрики, где этот UUID упоминается вместе с текстом названия и кодом языка, а также 3) таблица рубрикации, где UUID рубрики упоминается рядом с UUID оригинальной работы (UUID переводных работ можно не упоминать).

Тематическая подборка может быть реализована как рубрика с особым смысловым положением однако без какого-либо технологического отличия от обычной рубрики.

Завершается меморандум наших товарищей указанием на важнейшую роль международного текстологического проекта:

*Логически, если будет создан фонд теоретической текстологической литературы на различных языках мира, то это облегчит работу. Но эту работу надо делать параллельно и клонировать её.*

Положение о параллельном[[26]](#footnote-120) наполнении и *клонировании* (копировании? объединении? резервировании? вероятно, товарищи правильно подразумевают всё и сразу) действительно принципиально для базы данных, созданной по FZR-схеме. Широкое использование UUID вместо целочисленных кодов, которое является характерной особенностью схемы Фроша-Загорского-Радкевичюте призвано обеспечить именно то, что независимо составленные базы данных, содержащие разные произведения смогут максимально безболезненно объединиться в единую международную базу данных так, что единожды данные коды абзацев, примечаний и иных подобных элементов не будет уже меняться[[27]](#footnote-121). Поэтому пока не существует инфраструктуры единой международной базы данных по схеме Фроша-Загорского-Радкевичюте, пока нет международного хранилища, пока нет его текстологической и программистской команды, остаётся только создавать фрагменты будущего единого текстологического фонда для местного применения. Их совместимость и результативность вложенного труда гарантируется тем, что везде соблюдается единый порядок заполнения таблиц, перечень которых устанавливается в действующей спецификации схемы Фроша-Загорского-Радкевичюте. Эта спецификация достаточно гибкая для того, чтобы, даже в случае отсутствия серверных хранилищ, быть реализованной даже в SQLite файле[[28]](#footnote-122), который сможет быть загружен в единый общий текстологический фонд без остатка.

За работу!

*Продолжение следует*

1. Например, фотокопией рукописи или математически подписанным файлом с текстом. [↑](#footnote-ref-21)
2. Литва и Латвия - Пер. [↑](#footnote-ref-22)
3. Anatol Kitow, Nikołaj Krynicki, Elektroniczne maszyny cyfrowe oraz programowanie, 1963, Warszawa, M-wo obrony narodowej. (tł. z rosyjskiego Maciej Stolarski)

   Anatol Kitow, Elektroniczne maszyny cyfrowe, 1959, Warszawa, M-wo obrony narodowej. (tł. z rosyjskiego Maciej Stolarski) [↑](#footnote-ref-23)
4. Лишь весьма редкая разовая настройка некоторых важных параметров требует во многих программах управления базами данных редактирования текстовых файлов. К особенностям функций этих программ как автоматизации труда канцеляриста или делопроизводителя эти настройки обычно не имеют прямого отношения. [↑](#footnote-ref-27)
5. Специальные неповторяющиеся коды, которые можно создавать не имея общего списка таких кодов где-либо и будучи уверенным что такого кода нет больше ни у кого. [↑](#footnote-ref-28)
6. Для испытаний нужно уничтожить таблицу “Lingva familioj”, пересоздать её и перезаполнить. Такова цена простой структуры SQLite.

   DROP TABLE "Lingva familioj";  
     
   CREATE TABLE "Lingva familioj2" (  
   "Kodo" int PRIMARY KEY,  
   "Nomo" varchar(32) NOT NULL  
   );

   **После этого повторить все команды INSERT, указанные выше.**

   Нетрудно заметить, что методологически программа SQLite ориентирована на то, что таблица-картотека будет создана уже со всеми подробными деталями, которые в будущем не поменяются. Причины такого перфекционистского подхода состоят в том, что SQLite не задумывалась как средство коллективной работы, где трудно сразу предугадать все мелкие детали. SQLite как программа управления базой данных работает с единственным местным файлом и не предназначена для коллективной работы. То есть SQLite аналогичен личному секретарю-канцеляристу, тогда как другие программы управления базами данных (Oracle, DB2, MariaDB, Firebird, PostdreSQL, MySQL) работают с сетевыми механизмами и аналогичны организатору обмена справками, либо коллективному делопроизводителю. Аналогия с работой министерств Народной Польши может быть продуктивно развита намного глубже, если читателю знакома эта область деятельности. [↑](#footnote-ref-32)
7. Для испытаний нужно уничтожить таблицу “Lingvoj” и пересоздать её. Такова цена простой структуры SQLite.

   DROP TABLE "Lingvoj";  
   CREATE TABLE "Lingvoj" (  
   "Nomo originala" text NOT NULL,  
   "Nomo" text,  
   "Familio" int NOT NULL,  
   "ISO 639-1" char(2),  
   "ISO 639-3" char(3) NOT NULL  
   );

   **После этого повторить все команды INSERT, указанные выше.** [↑](#footnote-ref-33)
8. Замкнутой системой таблиц называется в данном случае группа таблиц, которая вместе полностью выражает некоторую сферу. В нашем случае исключение таблицы названий произведения делает систему таблиц бессмысленной, ибо сведения о произведении без названия бесполезны. С точки зрения учёта авторства и поиска по авторам или учёта издательств наша система таблиц не замкнута, однако поскольку эти возможности не рассматривалась, а система таблиц уже оказалась осмыслена то учёт авторства и издательств может рассматриваться как ближайшее усовершенствование, которое не отнимает смысла у имеющейся системы таблиц. [↑](#footnote-ref-35)
9. Эта команда не создаёт полноценные UUID, соответствующие стандарту в отношении неповторимости, однако для испытаний и уяснения сути работы в текстологической базе данных это не нужно. [↑](#footnote-ref-37)
10. Эта возможность реализуется в навыках общего и веб-программирования, которые обходятся вниманием поскольку нет возможности привлечь соответствующих специалистов к изложению проблем программирования для базы данных по схеме Фроша-Загорского. [↑](#footnote-ref-42)
11. Это вызвано тем, что диаграмма отражает более раннюю версию схемы базы данных, чем SQL директивы. Диаграмма создавалась не товарищем Фрошем, а по его просьбе другими людьми - TC.

    Фр. IDentifiant Universel Unique, анг. Universally Unique IDentifier [↑](#footnote-ref-50)
12. см. https://docs.postgresql.fr/10/datatype-uuid.html

    Украинский жаргонизм “всесвітня неповторна ознака” (всемирный опознаватель без повторений) - Ред. [↑](#footnote-ref-55)
13. Даталогия от нем. Daten - математическая наука о зависимостях данных, исследующая способы их выражения в формальных структурах, имеющих под собой самое разное техническое основание. Решающий вклад в обособление даталогии внёс Эдгар Кодд, выделивший понятие схемы данных - TC. [↑](#footnote-ref-95)
14. *Меморандум* — документ, выражающий достигнутое взаимное понимание или указывающий меру согласованности усилий - Ред. [↑](#footnote-ref-97)
15. В настоящий момент текстологическая работа фактически деградировала до нерегулярного составления фотокопий, хотя была начата с формирования pdf , стандартизованных специальной комиссией Студенческого Круга Философии Марксисткой. Подробнее см. по хронологии на http://marksizm.edu.pl/ - TC. [↑](#footnote-ref-100)
16. Это хорошо видно [по историческому примеру столетней давности](http://propaganda-journal.net/9402.html) (искать по словам «издательской программы») - Ред. [↑](#footnote-ref-101)
17. Marek Jan Siemek, Dlaczego Sokrates? [↑](#footnote-ref-103)
18. Сторонники шовинистических и устаревших кодировок вроде кодировки 1250 или кодировки 1252 до сих пор присылают различные возражения. Мы отказываемся рассматривать их впредь. Меньше двух недель назад текстологическая комиссия в ответ на вопрос, представимы ли такие фундаментальные ленинские произведения как «Материализм и эмпириокритицизм» или «Тетрадки по философии» в кодировках, отличающихся от Уникода, ответила - нет. Обоснованием являются символы дробей, стрелок, удвоенных восклицательных знаков, частые знаки французской и немецкой диакритики, несовместимой с чешским, польским, российским, украинским латышским или литовским текстом в любой из кодировок 1250-й серии - TC. [↑](#footnote-ref-107)
19. Подготовка к созданию нормализованных критических гипертекстов некоторых ленинских работ на литовском языке привела к необходимости осмысления **полного текстологического цикла**, то есть всей последовательности работ от создания постраничных фотокопий до размещения проверенных абзацев в базе данных. По причине широчайшего технического охвата указанных работ для произведений на литовском языке, эта последовательность условно названа «циклом Радкевичюте». Связанное понятие - т. н. **короткий текстологический цикл** или в жаргоне «цикл Стрэнбского», который основан на наличии распознанного и относительно связного текста, то есть предполагает относительно слабую связь с фотокопиями - TC. [↑](#footnote-ref-108)
20. Англоязычная подробная справка вызывается командой «man tesseract» - Ред. [↑](#footnote-ref-109)
21. Таблица «Абзац -полигон» отсутствует в диаграмме, поскольку ещё не была реализована. В общем случае она хранит в себе адрес исходного файла, четыре координаты полигона, принадлежащего абзацу в этом файле и UUID самого абзаца, то есть полученного текста. В такой формулировке один абзац может быть связан с несколькими полигонами. - TC. [↑](#footnote-ref-111)
22. Англоязычный намёк на существующие средства проверки по словарю https://stackoverflow.com/questions/20888326/using-aspell-library-in-java. [↑](#footnote-ref-112)
23. Сравнение качества предоставляемых словарей на указанных языках при подготовке этого очерка не производилось- TC. [↑](#footnote-ref-113)
24. Полиграфические таблицы являются наиболее проблемным элементом для выражения в гипертексте. В настоящий момент в Германии признано, что таблица должна рассматривать аналогично отдельному абзацу авторского текста, а её структура воспроизводится переводами **вручную**. В пользу этого способа текстологической работы говорит только лишь относительная редкость таблиц в классических прозведениях, однако этот аргумент перевешивает прочие- TC. [↑](#footnote-ref-115)
25. Пример. Реальная команда, используемая при переводах для полонизации украинских собственных имён. Вызов этой команды в терминале приводит к тому, что вводимые украинские имена выводятся близко к польскому написанию. Посредством перенаправления возможна автоматическая обработка файлов до нескольких миллионов знаков в секунду. В данных командах важен порядок замены букв - сначала идёт замена частных случаев, потом по более общим правилам.

    sed -e "s/ль/l/g;  
    s/зь/ź/g;  
    s/нь/ń/g;  
    s/ць/ć/g;  
    s/сь/ś/g;  
    s/ь//g;  
    s/ш/sz/g;  
    s/ч/cz/g;  
    y/абвгджзійклмнопрстуўфцыэґи/abwhdżzijkłmnoprstuŭfcyegy/;  
    s/щ/szcz/g;  
    s/х/ch/g;  
    s/ ю/ ju/g;  
    s/ я/ ja/g;  
    s/ е/ je/g;  
    s/ ё/ jo/g;  
    s/ є/ je/g;  
    s/ю/iu/g;  
    s/я/ia/g;  
    s/є/je/g;  
    s/е/e/g;  
    s/ї/ji/g;  
    s/łia/la/g;  
    s/łiu/lu/g;  
    s/łie/le/g;  
    s/łio/lo/g;  
    s/łi/li/g"; [↑](#footnote-ref-116)
26. Очевидно, первоочередные направления текстологической работы зависят от обстановки на национальном театре классовой борьбы. Так, от российских товарищей ожидается текстологическая база в виде оригиналов работ Ленина и Чернышевского. Украинские товарищи вполне могут считать первой значительной текстологической задачей публикацию двуязычного соответствия «Диалектики эстетического процесса» А. С Канарского, затем «Всеобщей теории развития» В. А. Босенко и «Диалектики» М. Л. Злотиной. К сожалению, как правильно отмечено в меморандуме, действительность такова, что сформировалось уже целое поколение украинских граждан, для которого оригиналы работ Чернышевского и Босенко представляют лексическую трудность. Впрочем, соответственно падению промышленности, на Украине упала стилистика *всех* употребляемых языков - TC. [↑](#footnote-ref-120)
27. Однако понадобится выбрать единственный UUID для каждого указанного автора, издательства и бумажного издания - TC. [↑](#footnote-ref-121)
28. В настоящее время ожидаются пробы размещения небольших работ в базе данных FZR-схемы, которую реализует SQLite файл. Официального описания FZR-схемы в виде команд SQL для SQLite в настоящее время не существует - TC. [↑](#footnote-ref-122)