

Е.А. Киселева

Кодирование профессиональных данных в целях проведения сравнительно-исторических исследований

Ключевые слова: кодирование, историческое профессиональное исследование, классификация, профессия, занятие населения, базы данных.

С развитием статистики как науки (учреждением Международного статистического института и регулярным проведением статистических конгрессов) формировались критерии учета населения, в том числе по признаку основного вида деятельности. По рекомендациям петербургской сессии международного статистического конгресса (1872 г.) указание на занятие, положение или отношение к предприятию являлось обязательным пунктом для разработки формуляров переписи [1, с. 107], а также учитывалось при других статистических обследованиях. Вместе с тем специалистами отмечалась необходимость разработки единой для всех стран классификации профессий в целях проведения сравнительных исследований профессиональных и социальных структур [2, с. 111–112].

В 1893 г. Международным статистическим институтом была принята наиболее современная для XIX в., по мнению Г.Г. Швиттау [3, с. 89], классификационная схема. Для активного профессионального населения устанавливалось 8 главных групп:

1. Эксплуатация поверхности земли: земледелие, лесное хозяйство, охота и рыболовство;
2. Горное дело как добыча ископаемых богатств;
3. Промышленность, переработка сырья;
4. Торговля;
5. Транспорт;
6. Внутренняя и внешняя охрана (войско и полиция);
7. Государственная служба;
8. Лица свободных профессий.

Наряду с основными группами устанавливалась группировка отдельных видов профессий, которая была важна для целей международного сравнения. Далее, отдельные виды занятий могли иметь более или менее детальные подгруппы, выделение которых оставлялось на усмотрение каждой из стран.

Также международное статистическое сообщество рекомендовало прилагать к публикуемым данным о профессиях особый словарь и указатель с соответствующими ссылками. Все основные группы и каждый отдельный вид занятий или профессий подлежал тщательному описанию, которым руководствовалась официальная статистика. Кроме того,

существовала идея: выделить некоторые виды занятий и профессий и сделать обязательным для всех стран распределение по ним профессионального населения. Международный статистический институт предлагал минимум 499 видов [3, с. 90]. К сожалению, несмотря на теоретические разработки и стремление ученых и исследователей прийти к единой схеме профессионального учета населения, практическая реализация статистических программ значительно отличалась от задуманных. Номенклатура и классификация, разработанная и принятая статистическим институтом, не получила широкого распространения.

Только в 1958 г. появилась первая версия Международной стандартной классификации профессий (ISCO), разработанная Международной организацией труда для целей сравнительного сопоставления. В основе ISCO была первичная общая классификация по 10 основным группам:

1. Законодатели, старшие должностные лица и управляющие;
2. Специалисты-профессионалы;
3. Специалисты и вспомогательный персонал;
4. Конторские служащие;
5. Работники сферы обслуживания и торговых предприятий;
6. Квалифицированные работники сельского хозяйства, рыболовства и рыбоводства;
7. Квалифицированные рабочие промышленности и рабочие родственных профессий;
8. Операторы и сборщики промышленных установок;
9. Неквалифицированные работники;
10. Вооруженные силы.

В качестве дополнительных факторов в классификации учитывались: цель выполняемой работы, используемое сырье, применяемые орудия труда и инструменты, необходимые стандарты, условия работы, профессиональная подготовка работников.

ISCO использовался при анализе данных, получаемых в результате обследования рабочей силы и при решении вопросов занятости [4]. Всего существует три варианта ISCO – 58, 68, 88 (числовое обозначение варианта соответствует году разработки и утверждения) [5]. Необходимость периодического обновления подобных классификаций возникает в связи с развитием структуры секторов экономики, а следовательно, и связанных с ними профессий. В обозначенных вариантах современной классификации ISCO наборы

Классификационная таблица дополнительной категории STATUS

OWNERSHIP	СОБСТВЕННИК
11 Owner, proprietor	11 Владелец, собственник, хозяин
12 Lease-holder, share-cropper	12 Арендатор, издольщик
13 Poor	13 Бедный
ARTISAN CAREER	КАРЬЕРА РЕМЕСЛЕННИКА
21 Master	21 Мастер
22 Journeyman	22 Подмастерье (кроме того, другое значение – наемный квалифицированный работник или ремесленник, работающий обычно поденно)
23 Apprentice, learner	23 Подмастерье, ученик
24 Artisan	24 Кустарь, ремесленник
PRINCIPALS AND SUBORDINATES	НАЧАЛЬНИКИ И ПОДЧИНЕННЫЕ
31 Principal	31 Начальник (когда контрольные функции являются второстепенными по отношению к основной деятельности)
32 'Worker'	32 Работник, сотрудник
33 Subordinate	33 Подчиненный (ассистенты и помощники)
34 Serfs and Slaves	34 Крепостные и рабы
TERTIARY EDUCATION	ВЫСШЕЕ ОБРАЗОВАНИЕ
41 Student	41 Студент (университета или колледжа, профессионального/ технического образования высшего уровня)
42 Graduate	42 Имеющие ученую степень (бакалавр, магистр, доктор наук)
'PURE' STATUS	«ЧИСТЫЙ» СТАТУС
51 Nobility	51 Дворянин
52 Prestige titles	52 Престижное звание

профессий существенно различаются. Такие «устаревшие» профессии в историческом контексте остаются носителями информации, которая имеет значение для историков и социологов [6, с. 89–90].

На рубеже XX–XXI вв. группа зарубежных исследователей, объединенных вокруг проекта HISMA (Международная группа анализа социальной мобильности), представила международную стандартную схему классификации исторических занятий XIX–XX вв. [7, с. 9–10].

Согласно HISCO профессии объединяются в частные, малые и большие группы (unit group, minor group, major group). В определениях частных групп обобщены функции работника, подробные описания которых даются в определениях отдельных профессий, входящих в эти группы. Первичные группы объединяются в малые, их определения составлены на основе сводных данных о первичных группах. Следующий уровень обобщения – большие группы, определения которых содержат указания на сферу экономики, в которой задействована профессия [6, с. 90]. Всего HISCO состоит из 284 частных групп, 83 малых и следующих 8 больших групп [7, с. 35–36]:

- 0/1. Специалисты и технические работники.
2. Административный и управленческий персонал.

3. Конторские служащие и родственные им профессии.

4. Работники сферы торговли.

5. Работники сферы обслуживания.

6. Сельскохозяйственные, животноводческие и лесного хозяйства работники, рыболовы и охотники.

7/8/9. Работники промышленности и родственные им профессии, транспорта, операторы оборудования и неквалифицированные работники.

Международный стандарт классификации исторических профессий основан на кодировании названий профессий и занятий населения, зафиксированных в исторических источниках. Кодификация – это ключевой момент HISCO, благодаря которому был решен вопрос проведения сравнительных исследований в области профессиональной структуры и занятости населения разных стран.

Код профессии состоит из 5 цифр, из которых первая определяет «большую группу» (например, группа 1 включает «Профессиональных и технических работников»), следующая цифра – указывает на «малую группу» (группа 1-9 «Профессиональные и технические работники, не входящие в другие малые группы»), «малые группы», в свою очередь, делятся на «частные группы» (например, группу 1–91 составляют

«Библиотекари, архивисты и хранители музеев»), последние 2 цифры относятся непосредственно к профессиональной деятельности (так, за кодом 1–91.30 скрывается профессия «архивариус») [8, с. 19].

Авторы и разработчики международного стандарта в руководстве HISCO предложили определенный алгоритм кодирования указанных в источниках занятий населения [7, с. 34].

Первым шагом является выделение «непрофессиональных» записей, т.е. таких, где занятие не обозначено вообще («пустая запись»), указанная информация косвенно относится к профессиональной деятельности, например, «в отставке», «бродяга», «нищий», или стоит пометка «без определенного занятия». Такая информация, которая отражает «непрофессиональную» занятость, может быть закодирована двумя кодами: либо -1, либо -2. Код -2 используется только в том случае, если запись о роде деятельности однозначно указывает на то, что человек не имеет работы, т.е. «без определенного занятия» – сюда же можно отнести и «временно неработающих». В остальных случаях применяется код -1.

Второй шаг – выделение записей, в которых указываются сведения о прошлой, настоящей или будущей профессиональной деятельности. Такая информация кодируется в соответствии с правилами международного стандарта. Это можно сделать как с помощью руководства HISCO, которое содержит описание и кодировку 1000 наиболее часто встречающихся профессий и занятий населения 8 европейских стран, так и на сайте «History of Work Information System» (<http://hisco.antenna.nl>).

Так как классификация создавалась с перспективой расширения и включения профессиональных данных других стран, то были разработаны 9 принципов кодирования для национальных, не обозначенных ранее или трудных названий профессий. Например, оригинальными российскими занятиями можно назвать «пимокат», «звонарь», «столовщик», «молотобоец», «квасник». К трудным для кодирования записям можно отнести слишком узкие или слишком широкие названия профессий. Подобного рода вопросы решаются также с помощью определенным образом подобранных кодов [13, с. 46–51]:

– для так называемых профессий общего характера, т.е. без указания специализации, код заканчивается на .10 или .05, например, записи «юрист», если нет ссылки на область права, в которой он специализируется, присваивается код 1-21.10;

– для записей, обозначающих только направление профессиональной деятельности даже без указания рода такой деятельности, код оканчивается на 0.00, например, «конторщик» (3-00.00) указывает только на то, что человек работает в какой-то конторе, для сравнения – запись «чиновник» определяет специалиста, занятого на государственной службе, поскольку неизвестна

ни должность, ни специализация, то можно определить большую и малую группы, а частная группа получает код .00, таким образом, код «чиновника» – 3-10.00.

Для названий профессий, частично относящихся к двум частным группам в пределах одной малой, создается промежуточный код .25 или .37. Так, между кодами 4-51.20 «Продавец в оптовой торговле» и 4-51.30 «Продавец в розничной торговле» появился код 4-51.25 «Продавец в рознично-оптовой торговле», с помощью которого в российской классификации был закодирован «торговец»;

– если название настолько неопределенно, что существует множество вариантов возможных кодов, то ставится код 9-99.99. В нашем варианте классификации исторических профессий такой код был присвоен записи «Кооператор».

Иногда название включает два отдельных вида профессиональной деятельности, например, «акушерка-фельдшерница». В таких случаях авторы руководства HISCO оставляют исследователю свободу выбора: или кодировать только первое занятие, или наиболее важное («главное занятие»). Кроме того, определенные трудности вызывают профессиональные названия, которые описывают и производство, и розничную продажу, в таких случаях разработчики международного стандарта рекомендуют кодировать только производственную деятельность [7, с. 34]. Следует отметить, что для России подобные положения не совсем адекватны. Они, безусловно, подходят для «мануфактурщиков», «промышленников», профессиональных ремесленников и т.д., для которых производство является преимущественным занятием, а продажей их товара могут заниматься «агенты», «купцы», «торговцы» и т.д. Но в то же время существовали «ремесленники», для которых основной целью было не производство, а продажа. В России число «чистых» земледельцев или фермеров было очень невелико, кроме земледелия крестьяне занимались кустарным промыслом. В переписи и других статистических обследованиях ремесло указывалось как «побочное» занятие, несмотря на то, что довольно часто являлось основным источником дохода. Конечно, для определенного процента крестьян ремесло было частью патриархального домашнего хозяйства, но к концу XIX в. этот процент значительно сократился. Таким образом, если кодировать кустарное ремесло только как производственную деятельность, то теряется целый слой сельских торговцев, представлявший в России значительный элемент рыночной системы.

Присвоение двух и более кодов может быть оправдано в исключительных случаях. Например, в России «кондитер» не только выпекал пироги и пирожные, как 7-76.30, но также делал конфеты и «сласти», как 7-76.60 [9].

Кроме основного кода международным стандартом предусмотрено кодирование дополнительной инфор-

мации с помощью категорий STATUS, PRODUCT и RELATION. Идея дополнительной классификации заключается в присвоении числового двузначного кода профессиональной или «непрофессиональной» информации, отражающей существенные моменты трудовой деятельности человека [7, с. 315].

Дополнительная категория STATUS отражает:

– уровень образования («студент», «слушатель курсов»);

– служебный статус («сотрудник»);

– социальное положение («помещик»).

Классификационная таблица категории STATUS представлена в таблице 1.

Категория RELATION включает данные:

– о семейном положении («при муже», «солдатка»);

– времени занятости («пенсия», «отставной»);

– добровольной работе («общественная работа»);

– нетрудоспособности («недееспособный», «душевнобольной»).

Классификационная таблица категории RELATION представлена в таблице 2.

При этом за пределами классификации остались выборные должности, например, такие как «сельский староста» или «городской голова», которые кодируются как «административные и управленческие работники» без указания на выборность. Дополнительные кодификационные возможности охватывают следующие категории непрофессионального населения: лица, живущие доходом с капиталов и недвижимости, лица, живущие пенсией или на частном попечении и благотворительности. В то же время сведения о лицах, находящихся в разного рода учреждениях (воспитанники закрытых учебных заведений, призреваемые в богадельнях, приютах и т.д.),

заключенных в тюрьмах, оказываются утраченными, так как «не вписываются» ни в одну из классификационных групп.

Основные коды следующей категории PRODUCT построены в соответствии с Международной классификацией товаров объединенных наций (The United Nation Central Product Classification – CPC). Целью данной категории является сохранение информации о продукте, который производится или покупается. Основные группы производимых, продаваемых и покупаемых товаров, кодируемых в HISCO, имеют следующий вид:

01-04 Сельскохозяйственные продукты, рыболовства и собирательства;

11-18 Руда и минералы, электричество, газ и вода;

21-29 Продукты питания, напитки и табак, текстиль, одежда и кожа;

31-39 Другие перевозимые товары, кроме металлических изделий, машин и оборудования;

41-49 Металлические изделия, машины и оборудование;

51-53 Нематериальные фонды, земля, строительство и строительные услуги.

Некоторые группы из Международной классификации товаров не использовались в HISCO, например, торговые услуги; сдача жилья в наем; сервировка еды и напитков; транспортные услуги; коммунальные услуги; финансовые услуги; услуги по недвижимости; услуги по лизингу и ренте; услуги в сфере бизнеса и производства; общественные, социальные и персональные услуги.

Предложенная группировка носит несколько обобщенный промышленный характер, который не позволяет отражать национальный «колерит» занятия. Так, весьма разнообразный отечественный кустарный

Таблица 2

Классификационная таблица дополнительной категории RELATION

FAMILY RELATIONSHIP	СЕМЕЙНЫЕ СВЯЗИ
11 Wife or widow	11 Жена или вдова
12 Son	12 Сын
13 Daughter	13 Дочь
14 Other male relative	14 Другие родственники мужского пола
15 Other female relative	15 Другие родственники женского пола
TEMPORAL RELATIONSHIP	ВРЕМЕННЫЕ СВЯЗИ
21 Former or retired	21 Указание на занятие в прошлом, пенсионер (в отставке)
22 Future	22 Указание на занятие в будущем
VOLUNTARY OR HONORARY RELATIONSHIP	ДОБРОВОЛЬНАЯ НЕОПЛАЧИВАЕМАЯ РАБОТА ИЛИ НА ОБЩЕСТВЕННЫХ НАЧАЛАХ
31 Voluntary, honorary	31 Добровольная неоплачиваемая работа, на общественных началах
INCAPACITATED	НЕТРУДОСПОСОБНЫЙ
41 Physical or mental disability	41 Физическая или душевная нетрудоспособность
HOUSEWORK	ДОМАШНЯЯ РАБОТА
51 homemaker	51 Надомник (сотрудник, выполняющий работу на дому)

промысел гораздо шире и специфичнее указанных групп. Так, «пимокат» относится к группе «Другие производители обуви» (8-01.90), а записи «ситница», «жет подпруги», «позолотчик», «поправка старья» и т.п. попадают в «безликую» категорию «Ремесло».

В целом можно сделать вывод, что принципы кодирования Международного стандарта классификации исторических профессий универсальны для наиболее распространенных профессий и занятий, при этом предложенные механизмы позволяют сохранять национальные или региональные особенности профессиональной деятельности.

Обработку материалов, отражающих занятия населения г. Тобольска, по данным Всероссийской переписи населения 1897 г. в целях создания пилотной версии русскоязычной классификации исторических профессий можно представить в виде четырех этапов, отличающихся от предложенных разработчиками модели HISCO.

I. Категоризация – сведение большого числа вариантов названий профессий в меньшее число категорий. При этом все одинаковые названия или записи, имеющие разночтения, но явно друг другу соответствующие, объединяются в одну категорию – название. Такая процедура необходима, чтобы объединить такие варианты как «изготовитель железных изделий для лошадей» – нетрудно догадаться, что речь идет о «кузнеце», а запись «государственная служба» соответствует «чиновнику» (база данных «Тобокск 1897»). Таким образом, из почти 12 тысяч записей было выделено порядка 160 профессий. Существует также определенное количество записей, которые сложно соотнести с какой-либо конкретной профессией, например, «делает веники» или «вяжет мережу». Если кодировать такие записи с помощью дополнительных категорий, то основным кодом является -1, а информация о занятии сохраняется в категории «STATUS» – «ремесло» (24) и соответствующей группы категории «PRODUCT».

II. Определение выделенных профессий по словарям преимущественно дореволюционных изданий. Смысл данного этапа заключается следующем:

- составить представление о профессии не в современном понимании, а по ее значению в конце XIX – начале XX в., т.е. на время создания источника («эксектор» – чиновник, ведавший хозяйственными делами и надзором за внешним порядком в канцелярии и присутственного места) [10, с. 80];

- выявить категории, имеющие два и более значения («объездчик» – тот, кто объезжает лошадей, и тот, кто объезжает определенную территорию) [11, с. 101];

- разграничить довольно схожие по названию, но отличающиеся по функциям занятия, о которых трудно составить представление без проведения соответствующего анализа (например, «акушер» – это обязательно дипломированных врач, специали-

зирующийся на акушерстве, а «акушерка» – это медицинский персонал, имеющий навыки в приеме родов) [10, с. 64].

III. Соотнесение русскоязычного определения и названия профессии с англоязычным вариантом по существующему руководству HISCO, в котором даются описания классифицируемых профессий в главе 4 «Группы HISCO, их определение и коды». Также для подбора точного английского эквивалента русскоязычным терминам профессии использовался энциклопедический словарь Уэбстера [12]. Наибольшие сложности вызывают профессии, аналог которым сложно найти в версии HISCO («акцизный», «водовоз», «кооператор» и т.д.), и записи, тяжело поддающиеся трактовке («крысоморщик»). В таких случаях следует переходить в четвертому этапу.

IV. Кодификация специфических названий осуществляется по правилам, указанным в руководстве HISCO и описанным выше.

При обработке и кодификации профессиональных данных Тамбова по спискам избирателей, имеющих право избирать в Учредительное собрание 1917 г., этапы выглядели немного по-другому, учитывая особенности самой базы данных и то, что пилотная версия русскоязычного стандарта классификации исторических профессий, основанная на базе данных по Тобольским переписным листам, уже существовала и успешно использовалась.

На первом этапе проводилась сортировка профессиональных данных в алфавитном порядке. Выделялись записи, в которых информация о занятиях отсутствует, а также профессии по схожести написания. Например, «чиновник», «чиновн.», «чинов.», «чин.» (база данных «Тамбов 1917»). Здесь же корректируются возможные ошибки заполнения и варианты написания, например, «портниха» и «партной», выделяются и уточняются непонятные сокращения и т.д.

На втором этапе проводится категоризация названий. Выделяются те категории профессий и занятий, которые будут составлять классификационную таблицу. Для каждой категории занятий выбирается один вариант написания. Так, если человек не имел определенного рода деятельности, то в базе данных могли быть следующие варианты записей: «не им.», «не им. зан.», «не работает», «неопр. зан.», «без опр. зан.», «без зан.», которые должны быть сведены к одному, например, «без занятий» с кодом -2. Таким образом, из 17 240 записей, содержащихся в базе данных, была выделена 291 категория занятий. Из них 13 профессии имеют код, указывающий только на общественный статус, семейные связи или производимый продукт, и 16 записей не закодированы.

На третьем этапе (определение наименований профессий) проводится пополнение уже существующего словаря определений. Это актуально в основном для новых категорий, которые не встречались в Тобольске.

Последний, четвертый этап – кодирование – для профессий Тамбова осуществлялся по двум направлениям. Первое включало соотнесение категорий занятий Тамбова с уже существующей и опубликованной русскоязычной классификацией, основанной на Тобольских профессиях [10, с. 68–72]. Второе направление заключалось в работе с новыми для русскоязычной классификации профессиональными названиями. Их кодирование осуществлялось с помощью web-сервиса HISCO, который расположен на сайте «History of Work Information System».

Безусловно, возможности Internet значительно облегчают и ускоряют процесс кодирования, но, тем не менее, остаются такие русскоязычные категории, которые сложно соотносятся с европейской классификацией, например, «ассенизатор», «артельщик». Кодирование таких названий профессий происходит по уже рассмотренной выше схеме.

Два источника, используемые в целях создания классификации исторических профессий и проведения сравнительного анализа профессионального состава городов, существенно различаются по времени

(списки избирателей появились через 20 лет после проведения всеобщей переписи населения), территории (города Тобольск и Тамбов), целям и условиям возникновения, объему предоставляемой информации и характеру самих данных. Так, если в задачах переписи изначально значился учет населения по различным, в том числе и профессиональным признакам, то указание основного занятия в списке граждан, имеющих права избирать депутатов в Учредительное собрание, было скорее формальным требованием, чем основным критерием. Кроме того, база данных «Tobolsk1897» задумывалась и создавалась в целях работы именно с профессиональными данными, о чем свидетельствует создание таких справочников, как «Сферы занятий», «Профессии и занятия», «Типы учебных заведений», для сравнения – база данных «Tambov1917» носит более универсальный характер. Но применение основных принципов международной классификации исторических профессий (HISCO) позволило преодолеть столь значительные различия и провести сравнительные исследования профессионального состава двух городов [13].

Библиографический список

1. Ходский, Л.В. Основания теории и техники статистики / Л.В. Ходский. – СПб., 1896.
2. Сравнительная статистика населения / Ю.Э. Янсон. – СПб., 1892.
3. Швиттау, Г.Г. Профессии и занятия населения. Опыт критико-методологического исследования в области экономической статистики / Г.Г. Швиттау. – СПб., 1909.
4. Международная стандартная классификация занятий : пер. с англ. – М., 1998.
5. International Standard Classification of Occupations, ISCO-88. – Geneva, 1990.
6. Киселева, Е.А. От психологической классификации профессий к исторической / Е.А. Киселева // Историческое профессиональное исследование / под ред. В.Н. Владимирова. – Барнаул, 2004.
7. Leeuwen M.H.D. van, Maas I., Miles A. HISCO: Historical International Standard Classification of Occupations. Leuven University Press, 2002.
8. Левин, М. Создание HISCO / М. Левин, И. Маас, Э. Майлс // Историческое профессиональное исследование / под ред. В.Н. Владимирова. – Барнаул, 2004.
9. Толковый словарь Даля // Словари и энциклопедии [Электронный ресурс] // <http://dic.academic.ru>
10. Владимиров, В.Н. Историческая классификация профессий: первый русскоязычный вариант HISCO / В.Н. Владимиров, А.В. Дёмкин, Е.А. Киселева и др. // Историческое профессиональное исследование. – Барнаул, 2004.
11. Павленкова, Ф. Энциклопедический словарь / Ф. Павленкова. – СПб., 1913. – Ч. 1.
12. The New Webster's Encyclopedic Dictionary of the English Language. Gramercy Books. – N.-Y., 1997.
13. Владимиров, В.Н. К исследованию профессионального состава населения Тамбова в 1917 г. / В.Н. Владимиров, Е.А. Киселева, В.В. Канищев, Н.В. Стрекалова // Экономическая история Сибири XX века : материалы Всероссийской научной конференции. – Барнаул, 2006. – Ч. 1.