
Э. А. АБГАРЯН

ПОДГОТОВКА СОЦИОЛОГИЧЕСКОЙ ИНФОРМАЦИИ К ОБРАБОТКЕ НА ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНОЙ МАШИНЕ ТИПА М-20, БЭСМ-3М, М-220

На данной стадии развития машинной техники пока что не решена проблема полного исключения объема механической работы. Современные машины лишены возможности приема информации в любом виде. Ввод информации в машину осуществляется либо с перфокарт, либо с перфолент, либо с магнитной ленты. Следовательно, для обработки социологического материала на электронно-вычислительной машине необходимо представить эту информацию в таком виде, который был бы удобным для восприятия машины. Этот процесс и таит в себе тот объем механической работы, на которую затрачивается почти девяносто пять процентов всего времени, используемого на обработку.

Период подготовки условно можно было бы разбить на три этапа:

1. Кодирование.
2. Шифровка.
3. Перфорация.

Качественное выполнение необходимой творческой и механической работы на каждом этапе обеспечивает точное воспроизведение полученной информации в электронно-вычислительной машине, что бесспорно содействует результатам с высокой степенью точности.

1. Кодирование.

Кодирование информации — наиболее ответственный этап при подготовке материала к обработке. Стремясь получить наиболее полную информацию в процессе одного обследования, социологи в ряде случаев используют анкеты открытого типа. Известно, что анкеты открытого типа являются носителями более точной информации. Но в этом случае мы получаем такое количество всевозможных ответов на поставленный вопрос, что возникает определенная трудность в воспроизведении всей информации в оперативной памяти машины. Заметим, что слишком большая дифференциация ответов не способствует также применению математических методов.

Поэтому, прежде чем приступить к кодированию, следует тщательным образом проанализировав материал определить тот необходимый объем информации, которая является достаточной для решения сформулированных задач.

Например, если требуется определить заработную плату опрашиваемого, то социолог, как правило, прибегает к одной из двух форм опроса.

1. Вопрос открытого типа, т. е. задается вопрос и оставляется место для ответа.

2. Вопрос закрытого типа, т. е. задается вопрос с определением наиболее вероятных интервальных групп.

В первом случае, задавая вопрос: «Какова Ваша месячная заработная плата?», мы можем получить множество разнообразных ответов (106 руб. 45 коп., 108 руб. 30 коп., 120 руб., 156 руб. 20 коп. и т. д.), которые, хотя и носят полную и точную информацию, по существу не представляют интереса для социологического анализа. Поэтому, прежде чем кодировать такой материал, следует представить его в удобном виде для планирования качественного анализа, количественного анализа и использования машинной техники. Наиболее удобным видом представления такого материала является интервальная группа. Кодирование осуществляется тогда, когда материал представлен в окончательном виде для ввода в машину.

Во втором случае все гораздо проще, так как данные группы уже определены.

Пример. Какова ваша месячная заработная плата?

От 60 до 80 рублей.

От 81 до 100 рублей.

От 101 до 120 рублей

и т. д.

В случае, когда анкета или опросный лист имеет вопросы и открытого и закрытого типа, то целесообразно закрытые вопросы кодировать до проведения опроса, а открытые вопросы в зависимости от полученной информации сгруппировать и лишь потом закодировать. Способы кодирования могут быть самыми различными. В каждом конкретном случае, в зависимости от постановки задачи, объема информации и принятой формы поставленных вопросов может быть принята вполне определенная форма кодирования. Ниже мы рассмотрим несколько примеров кодирования анкет закрытого типа.

Допустим, что исследователем разработана анкета закрытого типа, в которой каждый вопрос имеет по четыре взаимоисключающих ответа.

Например:

×₁ Какова ваша месячная заработная плата?

До 105 рублей.

105—135 рублей.

136—200 рублей.

Свыше 200 рублей.

X₂ Получаете ли вы премиальные?

Не получаю.

Раз в два года.

Раз в три месяца или раз в год.

Чаше, чем раз в три месяца

и т. д.

Кодировать такого типа анкеты можно следующим образом. Не принимая во внимание номер вопроса, ответам присваиваются цифры 1, 2, 4, 8, то есть выделенные вопросы принимают следующий вид.

X₁ Какова ваша месячная заработная плата?

1. До 105 рублей.

2. 105—135 рублей.

4. 136—200 рублей.

8. Свыше 200 рублей.

X₂ Получаете ли вы премиальные?

1. Не получаю.

2. Раз в два года.

4. Раз в три месяца или раз в год.

8. Чаше, чем раз в три месяца.

Не трудно заметить, что данная система кодирования основывается на десятичных цифрах, так как имеет в наличии цифру «8». Если же учесть, что программа в электронно-вычислительной машине записывается в восьмеричной системе счисления, то не трудно догадаться, что данный тип кодирования будет пробиваться и вводиться в машину как числа. Другими словами, в каждом адресе кода записывается не четыре цифры, а три, так как десятичные цифры занимают четыре разряда.

Таким образом, при данном кодировании в каждой ячейке размещаем девять вопросов, а на одной перфокарте мы можем разместить максимум сто восемь вопросов.

На бланке для перфорирования это будет выглядеть следующим образом.

Адрес	Команды и числа					Примечание
	=	00	X ₁ X ₂ X ₃	X ₄ X ₅ X ₆	X ₇ X ₈ X ₉	

Такая форма кодирования может быть оправдана в случае, если ответы на вопросы в анкете взаимоисключающие и если число вопросов не превышает 108.

Следует заметить, что в некоторых случаях такое кодирование не выдерживает критики. Во-первых, оно является не экономичным, это

видно из того, что используются не все разряды ячейки, а лишь с 1 по 36-й разряд, что, конечно, не говорит о совершенстве кодирования. Во-вторых, такое кодирование не позволяет увеличения числа ответов на вопрос (не более 4-х ответов).

Следующий тип кодирования, который предлагается, на наш взгляд является более совершенным и заслуживает большего внимания, несмотря на имеющиеся недостатки.

В некотором смысле, нижеприведенный вид кодирования является оригинальным. Оригинальность заключается в том, что это самая обычная нумерация ответов.

Рассмотрим данное кодирование на небольшой части анкеты.

I. Какая оценка вашего труда является наилучшей?

1. Личная удовлетворенность.
2. Доска почета, грамота, благодарность.
3. Оценка коллектива.
4. Материальная.
5. Оценка со стороны начальства.

II. Скажите, пожалуйста, рассчитываете ли вы в ближайшее время на продвижение по работе?

6. Да, в течение ближайшего года.
7. Да, но неизвестно когда.
8. Не знаю.
9. Нет, это исключено.
10. Меня это не интересует

и т. д.

Такое кодирование, на наш взгляд, является наиболее удобным при анкетном опросе. С точки зрения психологии заполняющего, здесь нет ничего непривычного. Преимущество такого кодирования заключается и в том, что оно использует все сорок пять разрядов ячейки, тем самым резко увеличивает объем информации, размещенной на одной перфокарте.

При обработке информации на электронно-вычислительной машине важное значение приобретает правильное размещение информации в оперативной памяти. Наилучшим видом кодирования в каждом конкретном случае является тот вид, который позволяет наиболее экономично использовать память машины.

Второй способ кодирования, хотя и полностью использует все сорок пять разрядов кода, в некоторых случаях не является оптимальным. Действительно, если анкета построена на вопросах с взаимоисключающими ответами и если число ответов в каждом вопросе не превышает семи, то целесообразно использовать следующий тип кодирования. Третий способ покажем на том же примере.

I. Какая оценка вашего труда является наилучшей?

1. Личная удовлетворенность.
2. Доска почета, грамота, благодарность.
4. Материальная.
5. Оценка со стороны

II. Скажите, пожалуйста, рассчитываете ли вы в ближайшее время на продвижение по работе?

1. Да, в течение ближайшего года.
2. Да, но неизвестно когда.
3. Не знаю.
4. Нет, это исключено.
5. Меня это не интересует

и т. д.

Сравнение двух последних типов кодирования наглядно показывает, что в каждом конкретном примере необходимо определить наилучший способ.

Критерием для определения наилучшего способа кодирования может быть занимаемый объем памяти машины. Другими словами, тот тип, с помощью которого данная конкретная информация занимает минимальный объем в памяти машины, является в конкретном примере наилучшим.

Анализируя вышеизложенные типы кодирования, предпочтение можно было бы дать третьему способу (впервые этот способ был применен для решения исторических задач на ЭВМ В. А. Устинозом, г. Новосибирск). Но он также не лишен недостатков. И одним из них является то, что он не предусматривает кодирование вопросов, где ответы не взаимоисключающие.

Способов кодирования социологической информации может быть множество, но вышеприведенные, на наш взгляд, являются наиболее приемлемыми.

Как мы уже отмечали, кодирование информации целесообразно произвести перед проведением опроса, особенно если опрос проводится с помощью анкет закрытого типа. В таком случае, после проведения опроса остается только зашифровать данный материал и перенести на специальные бланки для перфорации.

Шифровка и перенос информации на бланки

Шифровка — наиболее трудоемкий этап при подготовке информации к обработке. В зависимости от типа кодирования объем работы на этом этапе может быть сокращен. Вообще говоря, это один из факторов, влияющих на выбор того или иного типа кодирования. Некоторые типы

кодирования вообще не требуют шифровки. В таких случаях, на этом этапе работа ограничивается простым переносом кодированного материала на специальные бланки. Например, первый тип кодирования переносится на бланки без шифровки. Для наглядности мы увеличим число вопросов и условимся, что крестик перед ответом означает согласие опрашиваемого.

×₁ Какова ваша месячная заработная плата?

- 1. До 105 рублей.
- + 2. 105—135 рублей.
- 4. 136—200 рублей.
- 8. Свыше 200 рублей.

×₂ Получаете ли вы премиальные?

- 1. Не получаю.
- 2. Раз в два года.
- 4. Раз в три месяца или раз в год.
- + 8. Чаше, чем раз в три месяца.

×₃ С кем вы чаще проводите свой досуг?

- + 1. С друзьями по работе.
- 2. С друзьями по месту жительства.
- 4. С семьей.
- 8. Люблю одиночество.

×₄ Занимаетесь ли вы спортом?

- 1. Занимаюсь в секции.
- 2. Занимаюсь самостоятельно (регулярно).
- + 4. Занимаюсь от случая к случаю.
- 8. Не занимаюсь.

×₅ Принимаете ли вы участие в художественной самодеятельности?

- 1. Да.
- + 2. Нет.
- 4. Не регулярно.

×₆ Как часто вы ходите в кино?

- 1. Один раз в неделю и чаще.
- + 2. 2—3 раза в месяц.
- 4. Один раз в 2—3 месяца.
- 8. Несколько раз в год (очень редко).

×₇ Как часто вы посещаете театр?

- + 1. Ежемесячно.
- 2. 1 раз в 2—3 месяца.
- 4. 2 раза в год.
- 8. Очень редко.

- Х₃ Как часто вы читаете художественную литературу?
1. Не менее одной книги в месяц.

2. Одну книгу в месяц.

+ 4. Одну книгу в 3 месяца.

8. Читаю очень редко.

- Х₉ Ваш возраст (число исполнившихся лет)?
1. До 20 лет.

+ 2. С 21 до 25 лет.

4. С 26 до 30 лет.

8. С 31 года и выше

и т. д.

Таким образом, данный участок анкеты на бланке для перфорации будет представлен как:

Адрес	Команды и числа					Примечание
+ 1	0	00	281	422	142	
+ 2						

Заметим, что сорокапяти разрядная ячейка распределяется следующим образом: три разряда в разделе признак (45, 44, 43), шесть разрядов в разделе КОП (42, 41, 40, 39, 38, 37) и 36 разрядов в трех адресах, в каждом по 12 разрядов.

Второй тип кодирования в отличие от первого требует дополнительной работы по шифровке закодированного материала. Под шифровкой в данном случае мы понимаем перекодирование, то есть представление кодированного материала в таком виде, в каком необходимо записывать на бланки, для перфорации.

Эту задачу можно решить двумя способами. Первый способ заключается в том, что все ответы независимо от вопросов разбиваются на триады, затем каждому признаку присваивается код соответственно 4, 2, 1. Далее, в зависимости от числа ответов в данной триаде берется сумма приведенных цифр. Не трудно заметить, что сумма цифр не может быть больше семи, то есть кодирование производится в восьмеричной системе. Анкета в этом случае может быть представлена в следующем виде:

I. Какая оценка вашего труда является наилучшей?

5	4	+	1. Личная удовлетворенность.
	2.		2. Доска почета, грамота, благодарность.
	1	×	3. Оценка коллектива.

0	4		4. Матернская.
	2		5. Оценка со стороны начальства.
	1.		II. Скажите, пожалуйста, рассчитываете ли вы в ближайшее время на продвижение по работе?
6	4	+	6. Да, в течение ближайшего года.
	2	+	7. Да, но неизвестно когда.
	1.		8. Не знаю.
0	4		9. Нет, это исключено.
	2		10. Меня это не интересует.
	1		III. Ваша квалификация (разряд)?
2	4		11. Первый разряд.
	2	+	12. Второй разряд.
	1		13. Третий разряд.
2	4		14. Четвертый разряд.
	2	+	15. Пятый разряд.
	1		16. Шестой разряд.
2	4		IV. Соответствует ли ваша квалификация выполняемой вами работе?
	2	+	17. Да.
	1		18. Нет.
			19. Трудно сказать

и т. д.

Такой способ шифровки, бесспорно, имеет право на существование, но в ряде случаев нагромождение анкеты всякими линиями и цифрами может отрицательно повлиять на качество заполнения. Поэтому целесообразно шифровку второго типа кодирования производить на специально подготовленных табличках. Данные кодировочные таблички могут быть отражены на последней страничке анкеты. При соответствующей инструкции, опрашиваемые сами с удовольствием производят шифровку. В этом случае, после того, как информация полностью зашифрована, остается лишь перенести данные на бланки для перфорации.

Шифровка с помощью кодировочной таблички осуществляется следующим образом. После заполнения анкеты, опрашиваемый, в зависимости от того, какие ответы он отметил, на таблице зачеркивает те числа, которые соответствуют номеру ответа. Например, вышеприведенный участок анкеты при переводе на кодировочную таблицу принимает вид:

1 2 3	4 5 6	7 8 9	10 11 12	13 14 15	и т. д.
46 47 48	49 50 51	52 53 54	55 56 57	58 59 60	

Далее, после того, как информация с анкет перенесена на кодировочные таблички, начинается процесс шифровки (перекодирование). Здесь уместно напомнить, что три двоичных разряда могут содержать максимально цифру семь, в восьмеричном счислении. Другими словами, если мы имеем три разряда (триада), то в зависимости от того, какой разряд имеет признак, мы можем получить одну из восьми возможных цифр. Рассмотрим пример, условимся обозначать разряды с наличием признака — крестиком. Отсчет разрядов ведется слева направо.

1

0		

— в данной триаде ни один из разрядов не имеет признака, в результате триаду шифруем как «0».

2

		+
1		

— в данной триаде имеет признак первый разряд, в результате триаду шифруем как «1».

3

	+	
2		

— в данной триаде имеет признак второй разряд, в результате триаду шифруем как «2».

4

+		
4		

— в данной триаде имеет признак второй разряд, в результате триаду шифруем как «4».

В случае, если в триаде два или три разряда имеют признак, то триаду шифруем цифрой, соответствующей сумме обозначенных признаков.

Так, например:

1

	+	+
3		

2

+		+
5		

3	+		
	6		
4	+		
	7		

Таким образом, при переводе информации с анкеты на кодировочную таблицу, зачеркивая соответствующие числа, мы тем самым указываем на наличие признака в данных разрядах. Отсюда, информацию на рис. 4 можно зашифровать как:

1 2 3	4 5 6	7 8 9	10 11 12	13 14 15
5	0	6	0	2
46 47 48	49 50 51	52 53 54	55 56 57	58 59 60

Зашифрованная информация переносится на бланк для перфорации в следующей последовательности. Первая цифра записывается в графе «признак», следующие две цифры записываются в графе «КОП», далее по четыре цифры в каждом адресе A_1 ; A_2 ; A_3 . Итак, в первой ячейке мы записали 15 цифр, то есть разместили 45 ответов. Но число ответов в анкете, как правило, бывает больше чем 45, поэтому при переходе на следующую ячейку необходимо выдержать тот же принцип, то есть одна цифра в графе «признак», две цифры в графе «КОП», и по четыре цифры в адресах A_1 , A_2 , A_3 .

В случае, если число ответов в анкете не превышает 90, то тогда на одной перфокарте мы можем разместить 6 анкет. Если число ответов больше 90, но меньше 135, то на перфокарте можем разместить 4 анкеты и т. д.

При шифровке восьмеричных цифр необходимо проявить максимальную осторожность, ибо любая ошибка на данном этапе влечет за собой вторую.

Например:

1 разряд — отведен под ответ «Личная удовлетворенность».

2 разряд — отведен под ответ «Доска почета, грамота, благодарность».

3 разряд — отведен под ответ «Оценка коллектива».

Как мы видели, опрашиваемый отметил первый и третий ответ. Данную триаду мы зашифровали так:

1	2	3
5		

Пять означает, что опрашиваемый ответил «Личная удовлетворенность и оценка коллектива». Но если в процессе шифровки по ошибке мы вместо пятерки записали бы тройку, то тем самым мы потеряли бы ответ «Личная удовлетворенность» и приписали бы «Доска почета, грамота, благодарность».

1	2	3
3		

— не верно.

Процесс шифровки и перенос информации на бланки для перфорации должен сопровождаться тщательным контролем.

Третий тип кодирования так же, как и первый, переносится на бланки для перфорации без шифровки. Но в отличие от первого типа, третий тип занимает все разряды ячейки за счет того, что кодирование не превышает цифру семь. Числа в этом случае записываются на бланке как команды, то есть так же, как и во втором случае. Напомним, что это стало возможным за счет того, что ответы взаимоисключающие. В приведенном выше примере ответы на первый вопрос не взаимоисключающие, но для примера мы допускаем, что они взаимоисключающие.

I. Какая оценка вашего труда является наилучшей?

1. Личная удовлетворенность.
2. Доска почета, грамота, благодарность.
- + 3. Оценка коллектива.
4. Материальная.
5. Оценка со стороны начальства.

II. Скажите, пожалуйста, рассчитываете ли в ближайшее время на продвижение по работе?

1. Да, в течение ближайшего года.
2. Да, но неизвестно когда.
3. Не знаю.
4. Нет, это исключено.
- + 5. Меня это не интересует.

III. Ваша квалификация (разряд)?

1. Первый разряд.
2. Второй разряд.
3. Третий разряд.
- + 4. Четвертый разряд.
5. Пятый разряд.
6. Шестой разряд.

IV. Соответствует ли ваша квалификация выполняемой вами работе?

- + 1. Да.
2. Нет.
3. Трудно сказать

и т. д.

Таким образом, отмеченные номера каждого вопроса переносятся на бланки без шифровки, записывая подряд, так же, как и во втором случае, то есть:

Адрес	Команды и числа					Примечание
+ 1	3	54	1000	0000	0000	
+ 2						
+ 3						

Итак, после того, как вся информация записана на бланках, переходим к последнему этапу подготовки информации к машинной обработке — перфорации.

Перфорация — процесс, осуществляемый на перфорационных машинах. Известно, что большинство из современных электронно-вычислительных машин работают в двоичной системе, естественно, и вводимая информация должна быть выражена в двоичной системе. Перевод восьмеричных или десятичных чисел в двоичную осуществляется на перфорационных машинах. Полученные перфокарты или перфоленты также нуждаются в контроле, так как работа оператора за перфорационной машиной механическая и не лишена возможности ошибки.

Контроль перфокарт осуществляется двумя способами. Первый способ состоит в том, что пробитые перфокарты с помощью электромеханического печатающего устройства распечатываются на рулоне бумаги, после чего сверяются с теми бланками, с которых производилась перфорация. Данный способ надежный, но утомительный. Гораздо проще поддается контролю второй способ, который заключается в том, что данный массив перфорируется дважды, причем разными людьми на разных перфорационных машинах. Полученные перфокарты сверяются обычным наложением соответствующих перфокарт.

С получением перфокарт завершается подготовка информации к обработке. Массив перфокарт, являясь долгой памятью машины, может быть использован неоднократно для решения новых задач.