

12

АКАДЕМИЯ НАУК СССР  
ИНСТИТУТ НАУЧНОЙ ИНФОРМАЦИИ

ДОКЛАДЫ  
НА КОНФЕРЕНЦИИ ПО ОБРАБОТКЕ ИНФОРМАЦИИ,  
МАШИННОМУ ПЕРЕВОДУ И АВТОМАТИЧЕСКОМУ  
ЧТЕНИЮ ТЕКСТА

Выпуск II

Москва 1961

Сибирское Отделение АН СССР  
Институт математики

Э.В. Евреинов, Ю.Г. Косарев

ОБ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ УНИВЕРСАЛЬНЫХ  
ВЫЧИСЛИТЕЛЬНЫХ МАШИН ДЛЯ ЦЕЛЕЙ ПЕРЕВОДА

Новосибирск

- 1961 -

В настоящее время во многих организациях активно ведутся работы над алгоритмами перевода. По-видимому, недалекое то время, когда машинный перевод станет достоянием практики.

В связи с этим вполне своевременно поставить вопрос о типе машин, которые смогут обеспечить достаточную скорость перевода.

Представляется целесообразным прежде всего рассмотреть с данной точки зрения универсальные электронные вычислительные машины, серийно выпускаемые нашей промышленностью.

В этой связи нами была предпринята попытка оценить эффективность выполнения перевода на машине "М-20".

Расчеты времени велись на примере алгоритма перевода математических текстов с французского языка на русский, разработанный О.С. Кулагиной и И.А. Мельчуком.  
/Л.1/.

Важным достоинством данного алгоритма является то, что он позволяет разбить программу на ряд независимых участков, используемых однократно и работающих последовательно при переводе фразы.

На основании этого алгоритма группой сотрудников Института математики АН СССР, под руководством О.С. Кулагиной и Г.В. Вакуловской была составлена программа для машины "Стрела" и проведены опыты по переводу

математических текстов /Л.2/.

Приведем некоторые данные, характеризующие программу перевода:

Общий объем словарей - 10000 кодов.

Объем французского словаря - 5000 кодов.

Общий объем программы - 10000 ячеек.

Число частей программы - 17.

Объем каждой из частей программы не превышает 800 ячеек.

Число тактов, выполняемых при переводе фразы в 8-10 слов, составляет 45000 - 50000.

Из этого числа тактов 45,5% приходится на логические операции, около 25% - на операции переадресации и приблизительно столько же - на операции передачи управления. Таким образом, средняя длительность такта при выполнении перевода на машине "Стрела" оказалась равной примерно 250 микросекундам.

Словари и программа хранятся на магнитной ленте. Текст вводится с перфокарт.

На основании этих данных было подсчитано время перевода для машины "М-20". При этом было сделано допущение, что алгоритм, объем программы, распределение памяти оставались теми же, что и для машины "Стрела". Расчет велся для фразы длиной в 10 слов, записанной в 15-ти ячейках. Как видно из 2-й строки таблицы, средняя скорость перевода одиночной фразы оказалась равной 8-ми знакам в секунду, что явно недостаточно для практических целей.

Т и п машин.	Число одновре- менных фраз.	Примечания	Время вывода фразы	Время вывода слов- рей и прог- рамм.	Время работы схем пере- вода.	Общее время пере- вода фразы.	Скорость перевода.	Время перевод- ки объемом 300 стр.
			сек.	сек.	сек.	сек.	зн./сек	ч а с
1 "Стрела"	1	-	0,7	50	12	74,2	1	167
2 " " -20"	1	-	0,7	7	1,2	8,9	8	20,3
3 " "	10	-	0,7	0,7	1,2	2,6	27	61,2
4 " "	20	-	0,7	0,35	1,2	2,3	30	51,6
5 " "	40	-	0,7	0,17	1,2	2,1	33	51,1
6 " "	40	Замечание вре- мя перевода слов в словаре.	0,7	0,17	0,7	1,6	44	30
7 " "	30	До того, что в стро- ке 5 и 6 введенна скорость ввода в фразы.	0,3	0,17	0,7	1,3	58	2,2
8 " "	40	До того, что в строке 5 и 6 введенна ско- рость ввода и со- ответственно ре- сульт схемы перево- да.	0,7	0,17	0,5	1,1	100	1,7

1. Число в таблице " -20" совпадает с числом оставшихся устройств.  
2. При подсчете времени ввода и вывода: для "Стрела" - 15 фраз из 10 слов  
зачисляется 15 фразам.

Для увеличения скорости перевода необходимо прежде всего сократить время ввода словарей и программ с магнитной ленты в оперативную память. Это время может быть существенно уменьшено, если перейти к одновременному переводу многих фраз. Заигрыш во времени будет при этом пропорционален числу таких фраз. В таблице / строки 3, 4, 5 / даны подсчеты времени для случаев одновременного перевода 10, 20 и 40 фраз. Размещение такого числа фраз вполне реально, т.к. объем оперативной памяти "М-20" равен 4096 ячеек, а наибольшая из частей программы занимает не более 800 ячеек.

Скорость перевода при этом составит 27-33 -х знаков в секунду, т.е. на перевод книги объемом 300 страниц потребуется 5-6 часов, что, по-видимому, уже может представлять практический интерес.

Возможно дальнейшее увеличение скорости перевода за счет совершенствования схемы алгоритма. Как указывалось в упомянутой выше работе /Л.2/, около 45% времени работы схемы перевода занимает поиск слов во французском словаре, т.е. на поиск одного слова тратится в среднем 2000 тактов. Указанное число тактов может быть существенно уменьшено, если, например, принять следующую структуру словаря:

1. Словарь разбивается на 10 частей по 500 кодов в каждой.

2. Каждая часть словаря, в свою очередь, разбивается на 10 участков приблизительно равной длины.

3. Для каждой части словаря вводится таблица, в которой содержатся первые слова всех участков и их адреса, и последнее слово последнего участка.

Тогда схема поиска слова в словаре будет иметь следующий вид:

1. Вводится очередная часть словаря и соответствующая ей таблица.

2. С помощью таблицы определяется:

а. Может ли содержаться данное слово фразы в этой части словаря.

б. При положительном результате определяется участок, в котором искомое слово может содержаться.

3. Производится сравнение данного слова со всеми словами участка.

Заметим, что при такой схеме нет необходимости выстраивать слова текста в алфавитном порядке.

Общее число тактов на поиск слова по этой схеме, по нашим подсчетам, не превышает 200.

Таким образом время поиска слова в словаре может быть уменьшено в 10 раз. Скорость перевода при этом превысит 40 знаков в секунду.

Как видно из таблицы, для дальнейшего увеличения скорости необходимо уменьшить время ввода текста. Повышение скорости ввода приблизительно в 2,5 раза может быть достигнуто путем небольших изменений в устройстве ввода, как это сделано в ряде образцов машины "М-20". В результате этого скорость перевода может быть поднята до ~ 50 знаков в секунду.

Следует отметить, что указанная скорость не является предельной для машины "М-20".

1. действительно, применение фоточитателей, устройства для ввода с перфоленты, либо совмещение ввода с работой машины позволяют затраты времени на ввод фраз практически свести к нулю.

2. Можно также ожидать, что в ходе дальнейшей работы над совершенствованием алгоритма и программы удастся несколько сократить время работы схем перевода.

В итоге скорость перевода будет определяться временем печати результатов, т.е. может быть доведена примерно до 100 знаков в секунду.

Все приведенные выше расчеты делались для сравнительно небольшого объема словарей. На практике может встретиться необходимость применять словари гораздо большего объема.

На основании таблицы можно видеть, что увеличение объемов словарей не сопровождается пропорциональным ростом общего времени перевода. Так например, при увеличении объемов словаря в 2 раза скорость перевода уменьшится всего на 5-20%. При увеличении объема словарей в 10 раз общее время перевода увеличится не более, чем в 2-3 раза, т.е. останется все еще в допустимых пределах.

Следует отметить, что при увеличении объема словарей будет иметь смысл перейти к одновременной обработке большего числа фраз путем использования магнитных барабанов и магнитных лент для хранения текста и выборочного словаря.

Таким образом, и в этом случае имеется возможность достичь сравнительно высоких скоростей перевода.



В ы в о д ы: 1. Универсальная вычислительная машина "М-20" позволяет осуществлять переводы специальной литературы за практически приемлемое время.

2. Внедрение машинного перевода задерживается в настоящее время из-за отсутствия алгоритмов и программ перевода.

3. Учитывая возможность достижения достаточно высоких скоростей перевода на существующих универсальных вычислительных машинах / и тем более на машинах ближайшего будущего/, нет необходимости ставить вопрос о разработке для этих целей специализированных машин.

4. В комплект устройства вычислительной машины "М-20" следует добавить буквонечатающее устройство и фототрансмиттер для ввода перфоленты / либо совместить работу вводных устройств с работой остальных устройств машины/. Указанные усовершенствования нужны не только для машинного перевода, но и для решения задач из других областей математической лингвистики, а также для решения экономических, информационных и других задач.

#### ЛИТЕРАТУРА:

1. О.С. Кулагина и И.А. Мельчук. Машинный перевод с французского языка на русский.- "Вопросы языкознания", № 5, 1956.
2. О.С. Кулагина и Г.В. Вакуловская. Опытные переводы с французского языка на русский на машине "Стрела". - Сб. "Проблемы кибернетики", вып.2. М., Физматгиз, 1959.