

Миссия выполнима

Новые возможности PROMT 7.0

Машинный перевод (МП) текстов уже отметил свой полувекковой юбилей, однако по-прежнему количество критиков этой технологии никак не желает уменьшаться — всегда находятся желающие доказать, что компьютер все еще не может конкурировать с живыми переводчиками. Тем не менее разработчики переводчика PROMT сумели добиться значительных успехов, доказывая обратное.



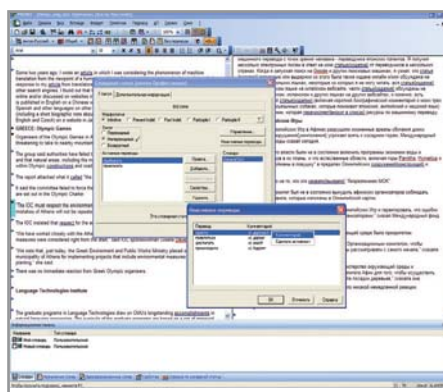
Стоит заметить, что в истории МП были даже такие моменты, когда эта технология признавалась экспертами неперспективной. Однако здор- ровые силы разработчиков смогли по- бороть скептицизм, и технология МП сейчас заслуженно признана вполне состоявшейся не только компьютерны- ми экспертами, но и миллионами поль- зователей программ-переводчиков во всем мире.

Сегодня в мире существует несколько десятков разработчиков систем МП. В нашей стране созданием машинных переводчиков занимается компания PROMT. Она хорошо известна пользова- телям благодаря своей системе перевода, которая первоначально носила название Stylus, а сейчас именуется PROMT. Эта система перевода существует на рынке уже 14 лет, и налицо очевидный про- гресс в плане повышения качества пе- ревода и функциональных возможнос- тей системы. В конце прошлого года на рынок была выпущена новая линейка систем перевода PROMT 7.0. Это со- провождалось громкими заявлениями компании-разработчика о создании но- вой технологии перевода и достижении очередных высот в сфере машинного перевода. Попробуем внимательно рас- смотреть, чем нас порадовали разработ- чики в новой версии системы перевода и какие сдвиги произошли в техноло- гии машинного перевода.

Архитектурные достижения

Системы перевода PROMT всегда имели довольно объемный главный словарь. Например, по данным разработчиков, генеральный (он же главный) словарь для англо-русской и русско-английской системы перевода PROMT 7.0 насчиты- вает 680 000 слов и словосочетаний. Это вполне закономерно, поскольку от объе- ма и качества словаря зависит качество самого перевода. Для получения точного перевода также важно иметь возможнос- ти настройки словаря, чтобы независи- мо от тематики текста система произвела правильный выбор варианта.

Система PROMT традиционно не да- ет пользователям возможности править генеральный словарь. Однако пользова- тель может создавать собственные сло-



▲ Рис. 1. Выберите активный пере- вод слова. Можно также добавить комментарий с пояснением

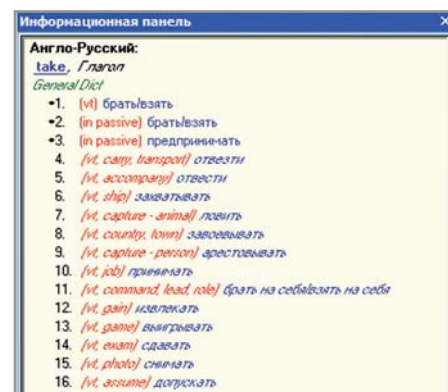
вари, в которые будут занесены нужные слова и требуемые в данном контексте варианты их перевода. Например, мож- но указать системе, что в переводимом тексте слово «switch» надо трактовать не как «выключатель», а как «хлыст».

В предыдущих версиях все варианты перевода, существующие в словаре, уча- ствовали в процессе перевода, это иногда усложняло выбор правильного варианта, и в результате получался некорректный результат. В версии PROMT 7.0 ситуация кардинально изменилась. Словарь те- перь содержит два типа переводов — активные и неактивные (рис. 1).

Активные переводы используется си- стемой при переводе, а пассивные хра- нятся в словаре и ждут своего часа. Пользователь может одним щелчком мыши перенести перевод слова из пас- сивного состояния в активное, и наобо- рот. Например, глагол «to take» имеет три наиболее распространенных вари- анта перевода: «брать», «взять» и «пред- принимать». Именно они и являются активными, и их система использует в первую очередь. Кроме того, для дан- ного слова существуют еще 16 вари- антов перевода (рис. 2), которые ис- пользуются значительно реже. И это, конечно, не предел!

Таким образом, словарь теперь может содержать неограниченное количество вариантов перевода каждого слова. Со- ответственно, настройка словаря за- ключается в выборе вариантов, кото- рые будут использоваться в процессе перевода определенного текста.

Такое изменение словарной архитек- туры является безусловным шагом вперед, поскольку максимально приближает сло-



▲ Рис. 2. Все переводы глагола «to take», из которых только три варианта являются активными

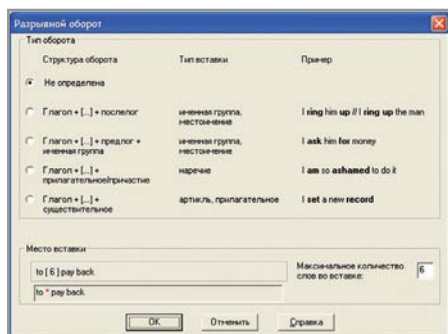
варь системы перевода к традиционному электронному словарю. С одной сторо- ны, теперь есть возможность поддержки неограниченного числа перевода кон- кретного слова, а с другой — множис- ленные варианты не будут мешаться при переводе (ранее система выдавала их в скобках рядом со словом). Новую словар- ную архитектуру разработчики назвали многомерной, поскольку в новой вер- сии системы есть переводы двух размер- ностей — активные и пассивные.

Дополнительным удобством является возможность добавления комментариев для каждого варианта перевода. В ком- ментарии можно указать, в каком кон- тексте используется данный вариант пе- ревода, время добавления или пожелания. Это также полезно, когда с системой ра- ботают несколько пользователей. Благо- даря комментариям можно узнать, кто до- бавил данный вариант перевода.

Возвращение к канонам

Пользователи старых версий словаря PROMT наверняка помнят, что про- граммма часто грешила очень своеобраз- ным словообразованием. Это было свя- зано с тем, что не были определены четкие правила распознавания, в связи с чем зачастую система могла выбрать вариант слова по собственному усмот- рению. Например, английское слово «radio beacon» (радиомаяк) система могла распознать следующим образом:

- ▶ radio beacon
- ▶ radio beacons
- ▶ radio beaconses
- ▶ radios beacon
- ▶ radioes beacon
- ▶ radios beacons



▲ Рис. 3. Выберите в этом окне структуру данного разрывного оборота, чтобы занести в словарь правильный перевод

- » ► radioes beacondes
- radioes beacons
- radios beacondes

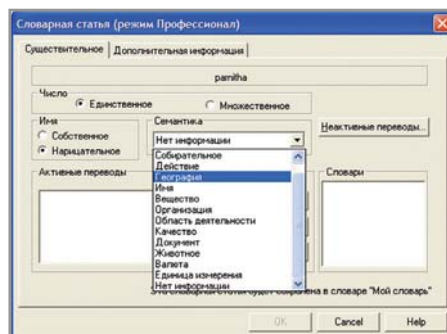
Из этого списка только первые два варианта являются правильными, остальные — фантазии самой системы.

В новой версии словарь хранит только канонические формы для каждого слова плюс четкие правила распознавания всех форм. Это означает, что, например, для английского глагола «делать» в словаре хранится только форма «to make», которую система умеет правильно склонять. Иначе говоря, интеллект системы существенно вырос.

«Разрывная», — подумал Штирлиц

Одна из главных проблем, с которой сталкиваются системы машинного перевода — это многозначность слов в любом естественном языке, а также существование устойчивых словосочетаний и разрывных оборотов. Причем это характерно как для языка, с которого делается перевод, так и для языка, на который осуществляется этот перевод. Такие обороты называются разрывными, потому что слова, входящие в их состав, не всегда следуют друг за другом, следовательно простого добавления в словарь недостаточно. Например, в фразе «They took everything off» оборот «to take off» разделен словом «everything». Ранее система перевода сразу спотыкалась на таких выражениях и выдавала чистый подстрочник, весьма далекий от правильного перевода («они взяли все прочь»). Правильный перевод — «они сняли все».

В PROMT 7.0 эта проблема окончательно устранена: во-первых, система теперь понимает большое количество



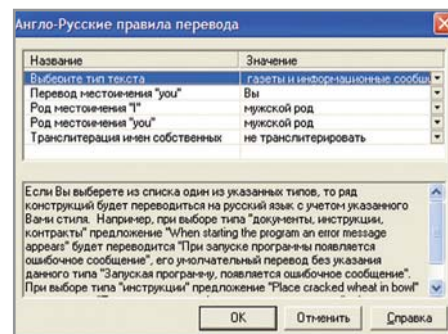
▲ Рис. 4. Для слова «Parnitha» выберите тип семантики «География» (это место в Греции), чтобы адекватно определить его значение

устойчивых разрывных оборотов (например, «keep promise», «pay bill», «keep away»). Во-вторых, пользователь может самостоятельно вносить в словарь новые разрывные обороты. Например, в предложении «We will pay them back for the trick they played on us» словосочетание «pay back» является разрывным оборотом. Для ввода выражения в словарь выделите «pay back» и с помощью правой кнопки мыши выберите команду «Словарная статья». Затем укажите часть речи — глагол (не забудьте выбрать уровень прав «Профессионал»). В диалоговом окне появится флажок «Разрывной оборот» (рис. 3), активизируйте его и нажмите кнопку «Укажите тип оборота». Далее в диалоговом окне «Разрывной оборот» выберите структуру данного оборота и затем введите его перевод в словарь.

Тонкая семантика

Еще одним новшеством словарей в PROMT 7.0 является возможность выбора семантического класса. Иначе говоря, новая процедура позволяет определить (или обозначить, так как семантика переводится с греческого как «обозначающий») тип слова (рис. 4).

Чтобы проиллюстрировать, как работает это новая возможность, переведем фразу «I bought some yogurt with live cultures». На выходе система выдаст: «Я купил некоторый йогурт с живыми культурами». Очевидно, что программа-переводчик не поняла, что такое «yogurt» и выбрала для слова «some» самый общий перевод — «некоторый». Поэтому для правильного перевода слова «йогурт» нужно указать, что это вещество. После этого перевод будет вы-



▲ Рис. 5. Выбрать правила перевода, соответствующие содержанию переводимого текста, совсем не сложно даже для новичка

глядеть следующим образом: «Я купил немного йогурта с живыми культурами».

Аналогично можно указать семантические типы для географических объектов, имен, животных, валют и других типов. Для этого зайдите в словарную статью по выбранному слову и в списке «Семантика» выберите его тип.

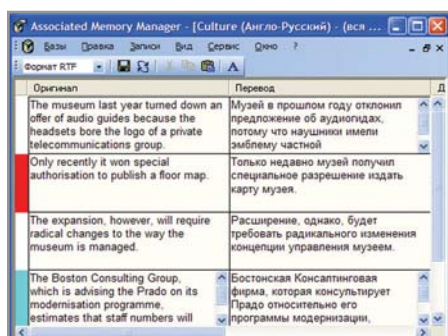
Конечно, даже при неправильном определении типа слова перевод будет понятен. Однако это улучшение приобретает существенное значение, когда приходится переводить объемные тексты и на выходе требуется качественный перевод. Чем меньше система будет выдавать «некоторых йогуртов», тем меньше требуется правки и существеннее экономия времени.

На вкус и цвет

Как вы, наверное, уже заметили, повышение управляемости работой системы перевода — это, так сказать, мейнстрим в новой системе PROMT 7.0. Разработчики уделили повышенное внимание различным настройкам, чтобы пользователи могли по максимуму адаптировать систему под свои тексты. Помимо выбора вариантов перевода и семантики у пользователей теперь есть также возможность выбирать правила перевода для конкретного текста.

Не секрет, что стиль написания кулинарного рецепта довольно сильно отличается от стиля делового контракта. Поэтому в системе PROMT 7.0 пользователь может указать индивидуальные правила перевода не только для всего документа, но и для каждого абзаца в отдельности.

Для выбора правил перевода документа выберите пункт «Тематика → Правила перевода документа», а для установки »



▲ Рис. 6. Ассоциированная память: сохраняем перевод для дальнейшего использования

» правил перевода абзаца щелкните правой кнопкой мыши по данному абзацу и выберите «Правила перевода абзаца».

Пользователю предлагается выбрать следующие настройки (рис. 5): тип текста (деловое письмо, газета, личная переписка); как переводить местоимение «you» — «ты» или «вы»; какой род задать для местоимений «I» и «you», а также установить транслитерацию собственных имен.

Поскольку речь зашла о выборе алгоритмов перевода, стоит отметить возможность выбора британского или американского варианта английского языка (эта возможность появилась в предыдущей версии PROMT).

Перевод графики

Исходно PROMT обрабатывал документы исключительно в текстовой форме. Затем к числу поддерживаемых форматов добавились документы Microsoft Office 2000-2003 и HTML-страницы. По сути, это было логично, поскольку данные форматы также являются текстовыми. Вот только перевод графических файлов оставался несбыточной мечтой. Потребность в переводе графики возникла в связи с ростом популярности формата PDF, в котором распространяется большое количество технической документации и деловых документов. Как можно перевести графический документ? Распечатать, отсканировать, распознать, сконвертировать текст в электронную форму и только после этого перевести. Процедура, прямо скажем, не из легких. Особенно если количество страниц в документе исчисляется десятками или даже сотнями.

PROMT 7.0 решает эту проблему! Система работает (рис. 7) с документами PDF, TIFF, BMP, JPG и рядом других форматов.

Это стало возможным благодаря встроенной системе OCR от компании I.R.I.S. (Readiris). Теперь в списке поддерживаемых форматов файлов появилась новая строчка — «Image file». После выбора этого типа файлов запускается встроенная OCR-система, которая на лету выделяет из графического файла текст и передает его системе перевода. Распознавание идет довольно быстро (в PDF-документах скорость зависит от количества содержащейся графики). Минута-другая, и перед нами документ в электронном виде. Далее его можно переводить, редактировать и сохранять в виде текстового документа.

Стоит отметить, что в данной версии OCR-система не поддерживает работу со сканером. Распознать можно только уже готовые графические файлы.

Перевод... «на три буквы»

Помимо графических файлов PROMT 7.0 переводит еще и XML-файлы. В настоящее время формат XML (Extensible Markup Language) становится все более популяр-

ным и активно используемым стандартом представления данных. XML-файл — это обычный текстовый файл, созданный по определенным правилам, который содержит данные и их структуру, описывающие какой-либо объект (например, товарную накладную). Наиболее часто XML используется при обработке и хранении больших объемов документов, поскольку в базе данных отдельно сохраняются текстовый контент и форматная информация документа.

В связи с большой популярностью стандарта XML появилась и потребность в переводе файлов этого формата. В новую версию программы PROMT включено приложение, которое предназначено для создания и редактирования файлов правил для перевода XML-документов. В отличие от файлов других подобных форматов или документов Word, XML-файлы не обладают заранее определенной структурой, поэтому для их корректного перевода требуется дополнительная информация, описывающая перевод конкретного файла (или ряда файлов с одинаковой структурой). Эта информа-

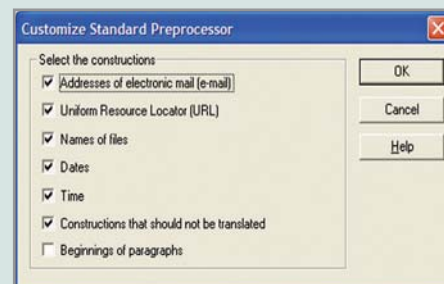
Как настроить PROMT 7.0

Тонкости взаимопонимания

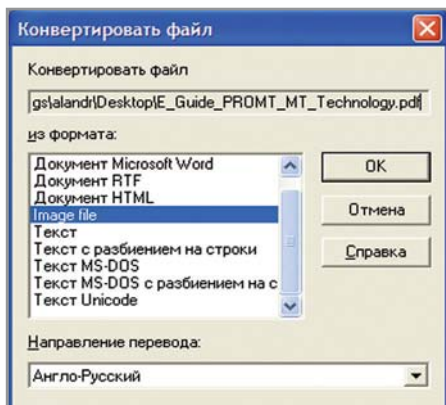
Для адаптации переводчика под собственные тексты и повышения качества перевода разработчики предлагают использовать следующие настройки:

- Подключение специализированных словарей, разработанных компанией PROMT. На данный момент насчитывается более 100 таких словарей по различным тематикам, которые разработчики рекомендуют использовать при переводе узкоспециализированных тематических текстов.
- Создание пользовательских словарей. Наиболее эффективно при переводе узкоспециализированных текстов.
- Создание списка зарезервированных слов. Зарезервированное слово — это слово (словосочетание), которое помечается во входном тексте как не требующее перевода. Чаще всего нужно резервировать имена собственные, аббревиатуры, специальные наименования, которые совпадают со значимыми словами (например, «Windows XP», «Miami Beach» и т. п.).

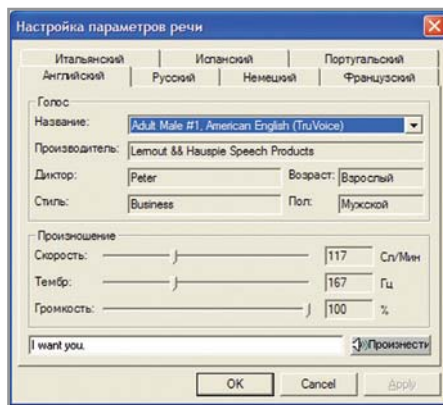
- Выбор правил перевода.
- Подключение баз ассоциированной памяти. В базе ассоциированной памяти можно сохранять словосочетания, предложения или целые абзацы исходного текста и их значения. Это упрощает перевод текстов по данной тематике.
- Использование препроцессоров. Препроцессор — это набор правил, которые используются при переводе таких конструкций как адреса электронной почты, адреса интернет-сайтов, имена файлов и т. д.



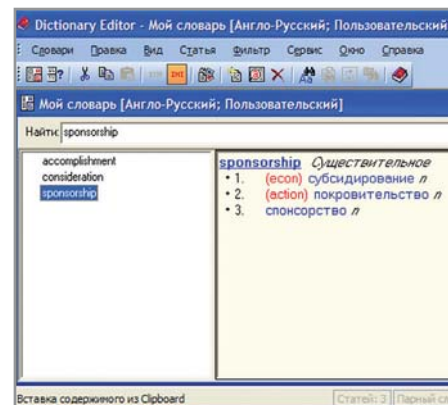
◀ Для настройки препроцессора нужно поставить галочки для соответствующих позиций



▲ Рис. 7. При открытии графического файла или документа в формате PDF укажите тип «Image file»



▲ Рис. 8. В окне настройки можно скорректировать скорость чтения текста, тембр и громкость голоса



▲ Рис. 9. Редактор словарей — мощный инструмент для улучшения качества перевода

» ция содержится в файле правил и описывает условия перевода различных тегов и их атрибутов.

Программа-говорун

PROMT 7.0 не только распознает и переводит документы. В программе также есть возможность чтения текстов вслух благодаря интегрированной технологии синтеза речи Microsoft Agent и модулям голосового синтеза Text-To-Speech. В комплекте PROMT поставляются речевые модули

для всех языков, которые входят в поставку (например, для PROMT в комплектации «Гигант» поставляются шесть речевых модулей: для английского, немецкого, французского, испанского, итальянского и русского языков).

Программа чтения предусматривает настройку скорости, тембра и громкости голоса (рис. 8). Также в программе существует режим, при котором озвучивание текста производится с использованием всплывающего персонажа Microsoft

Agent, благодаря чему процесс становится более живым и наглядным. При этом воспроизводимый текст дополнительно выводится во всплывающем окне. Возможность озвучивать переведенный текст помогает пользователю сберечь свое зрение, а заодно и время при прочтении различных текстов на компьютере. Естественно, для того чтобы компьютер смог заговорить, потребуются соответствующее оснащение: звуковая карта и колонки или наушники. »

Конкурс!

Уважаемые читатели!

Вы держите в руках юбилейный, 50-й выпуск нашего журнала, который мы постарались сделать для вас особенным, более праздничным и интересным.

Частью этой акции является и данный конкурс, для участия в котором вам надо всего лишь заполнить небольшую анкету на обороте, ответить на три несложных вопроса о вашем любимом журнале, наклеить все это на почтовую карточку и прислать к нам в редакцию по адресу: 127018, Москва, ул. Полковая, д. 17, с пометкой «Chip конкурс».

Срок приема анкет — до 20 июня. После этого путем жребия мы выберем счастливых, правильно ответивших на вопросы, которые получают прекрасные призы, указанные справа.

Призы победителям



Оперативная память KVR533D2N4/512 (DDR2 533 МГц 512 Мбайт);
флеш-драйв KUSBDE 256 Мбайт

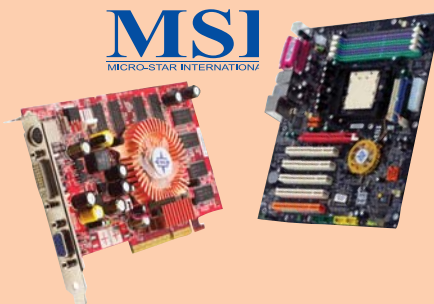


iRiver
Плеер IRIVER H-10

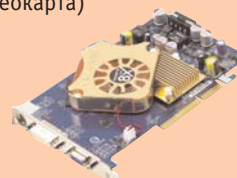


Интернет-переводчик
(10 экземпляров)

Выражаем
благодарность
компаниям,
предоставившим
призы



Комплект K8N Neo2 Platinum (системная плата) и NX6600-VTD128 (видеокарта)



Видеокарта Gigabyte GV-N66T128D



Корпус Diabolic Silver Blue
Материнская плата NF3 250K8AA-ERS



Что в итоге?

Итак, полноценный компьютерный перевод — уже не мечта, а реальность. Обзор новинок в PROMT 7.0 показывает, что многие из задач, казавшихся несколько лет назад невыполнимыми, успешно решены, а сама система стала гораздо удобнее и дружелюбнее. Несомненно, в данном случае разработчики PROMT учли многие пожелания и нужды современных пользователей. В итоге получился качественный и удобный в использовании продукт.

Машинный перевод: быстрее, выше, сильнее

С каждой новой версией совершенствуются алгоритмы и повышается качество перевода. Но даже самый лучший машинный переводчик никогда не сможет конкурировать с человеком. Призвание систем перевода в другом — служить надежным и удобным инструментом для преодоления языкового барьера. По сути, программа решает две задачи: дает быстрый и черновой перевод в тех случаях, где важно понять смысл текста на иностран-



PROMT 7.0 Family

Члены одной семьи

PROMT 7.0 — это общее название целого семейства ПО. В его состав входят следующие продукты:

- PROMT Express 7.0 — самый «легкий» продукт, предназначен для быстрого перевода текстов в домашних условиях. Подойдет пользователям, ограниченными в средствах. Цена: \$29.
- PROMT Internet Premium 7.0 — переводчик веб-сайтов. Цена: \$49.
- PROMT Standard 7.0 — «младший» член семейства. Программа умеет переводить тексты, электронную почту и графические файлы. Простой интерфейс, базовый набор настроек. Пози-

ционируется как продукт для дома или малого офиса. Цена: \$99.

► PROMT Professional 7.0 — продукт для бизнес-пользователей. Весь набор возможностей для офиса: интеграция в Microsoft Office, перевод веб-сайтов, графических файлов. Полный набор настроек для повышения качества перевода. Цена: \$300.

► PROMT Expert 7.0 — продукт для профессиональных переводчиков. Отличается от версии Professional наличием возможностей интеграции с TRADOS® и программы для извлечения терминологии. Цена: \$599.

ном языке, и служит мощным инструментом повышения производительности труда профессионалов. Первая задача благополучно решена уже давно, а для решения второй разработчики постоянно совершенствуют систему и предлагают все но-

вые настройки для адаптации переводчика в соответствии с тематикой текстов пользователя. Чем удобнее будет работать с переводчиком, тем больше будет благодарных пользователей.

■ ■ ■ Федор Дементьев

Конкурс!

Вопросы для участников конкурса

1. Как назывался и чему был посвящен первый спецвыпуск журнала Chip?
2. В каком году Chip выходил с колонкой редактора каждого раздела?
3. Обложка какого номера Chip была черной? (месяц, год)

Ответы на вопросы:

- 1.
- 2.
- 3.

Анкета:

- | | | |
|----------------------|---------------------|-------------------------|
| 1. Сколько вам лет? | с. Всеобщее среднее | 4. Покупали бы вы |
| а. Меньше 19 | d. Высшее | Chip с DVD? |
| b. 20-29 | 3. Сколько часов | а. Да |
| с. 30-39 | в неделю вы | b. Нет |
| d. 40-49 | пользуетесь | 5. Какие еще издания |
| e. более 50 | Интернетом? | Vogel Burda Commu- |
| 2. Ваше образование? | а. Меньше 1 часа | nications вы покупаете/ |
| а. Неполное среднее | b. 1-3 часа | читаете (обведите)? |
| b. Среднее | с. 5-10 часов | а. Foto-Video Digital |
| специальное | d. более 10 часов | b. Chip Special |