

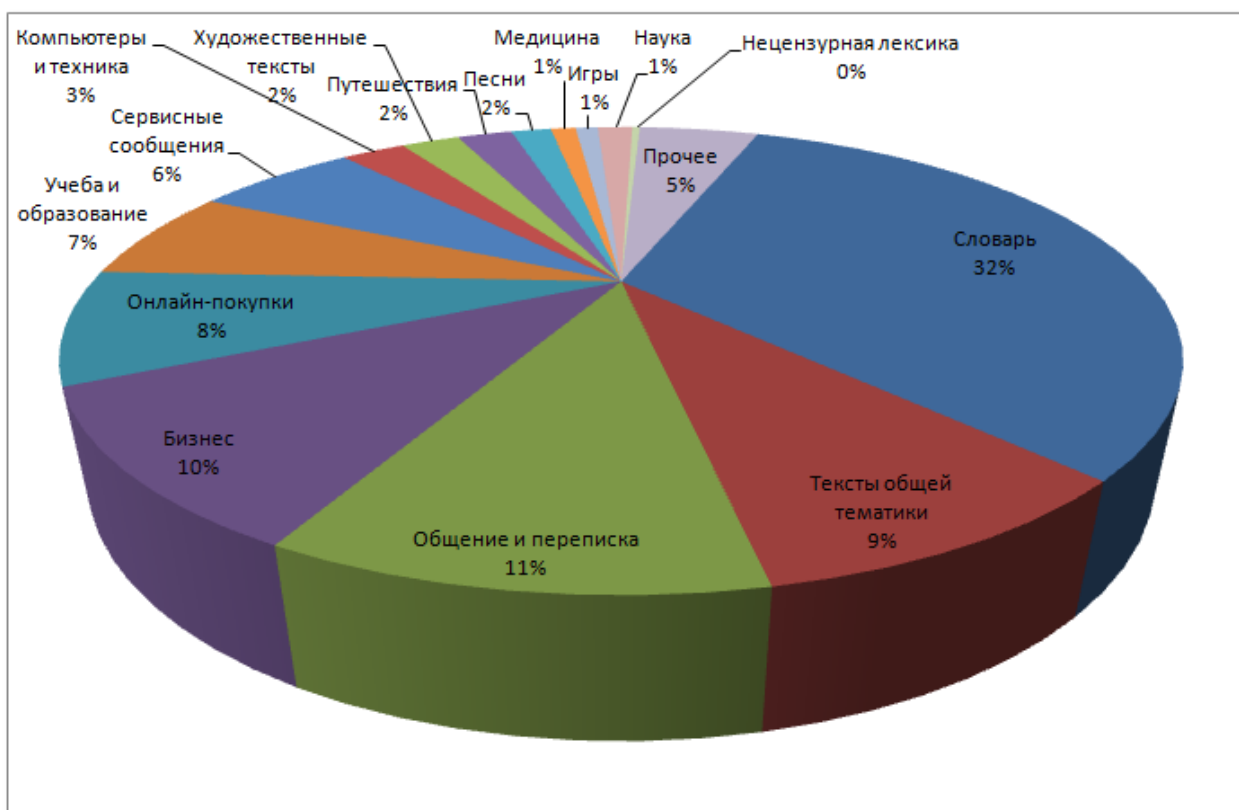
Что переводят в Интернете

*Мы выяснили, что за год пользователи Translate.Ru стали больше интересоваться наукой и меньше ругаться матом.*

Основой для данного исследования послужили результаты обработки запросов пользователей к сайту Translate.Ru в феврале 2011 года. Для построения сравнительного анализа мы использовали аналогичные данные, полученные ровно год назад (в феврале 2010 года), и дополнительно проанализировали данные за октябрь 2010 года.

Сам процесс классификации текстов построен на базе общих механизмов автоматического определения тематики перевода – тех же самых, которые применяются во всех программных продуктах PROMT и хорошо знакомы нашим корпоративным пользователям. За основу берется множество критериев, основным из которых является наличие определенных ключевых слов и словосочетаний в тексте. Также мы учитывали и тот выбор тематики, который пользователь сам делал при переводе своих текстов на сайте.

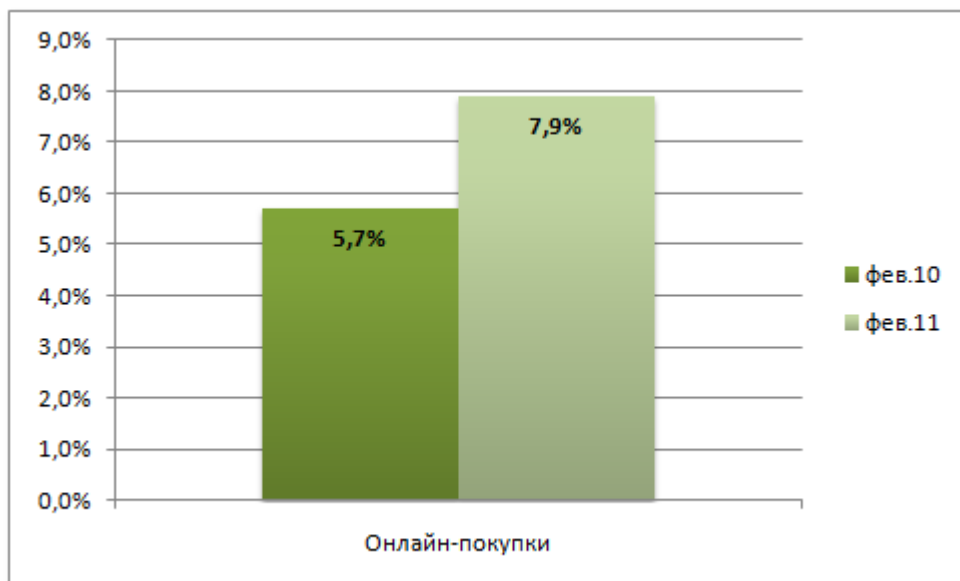
В результате мы получили следующее распределение запросов по тематикам:



Более 40 % всех переводов на Translate.Ru представляют собой либо отдельные слова (32 % всех запросов), либо тексты общей тематики (9 %), которые по разным причинам не могут быть классифицированы. Эти категории имеют относительно стабильные показатели на протяжении всего года, если не считать постепенного прироста переводов отдельных слов. Мы связываем его с тем, что словарный сервис постоянно развивается и предлагает новые возможности (например, недавно появилась грамматика и транскрипции).

Далее по популярности следуют общение и переписка на разные темы (11 %), а также перевод статей о бизнесе и бизнес-переписка – они имеют в совокупности 10 %. Эти 2 категории в среднем тоже стабильны.

Онлайн-покупки, к которым относятся переводы описаний товаров в интернет-магазинах, а также справочных страниц о способах оплаты и условиях доставки, занимают следующую позицию с 8 %, и эта категория заметно выросла за последний год (+39 %).



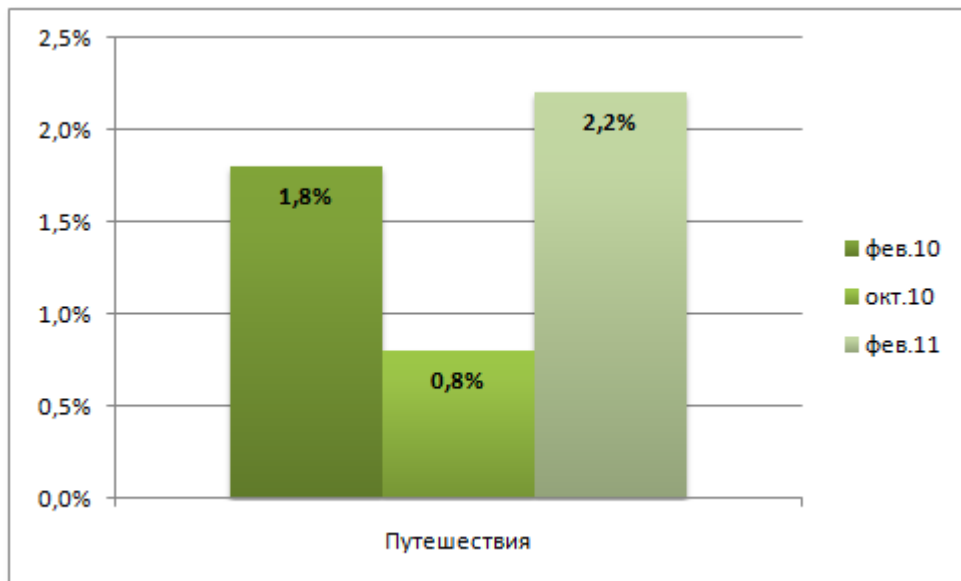
Тематика «Учеба и образование» также достаточно стабильна, занимает 7 % всех запросов и имеет некоторую тенденцию к увеличению. Это тексты из учебников, различные упражнения, топики, а также уставы и правила поступления в различные учебные заведения.

Следом идут так называемые «Сервисные сообщения» – целых 6 % запросов. Сюда попадают различные типовые сообщения и подписи к элементам интерфейса, сообщения об ошибках (вроде «Page not found» и «Access denied»), стандартные формулировки из рассылок и уведомлений.

С заметным отрывом – 2,5 % – за ними идут «Компьютеры и техника», включающие описания устройств и инструкции к различным гаджетам. За год они стали заметно менее популярны, что, видимо, связано с успехами в деле перевода инструкций самими производителями.

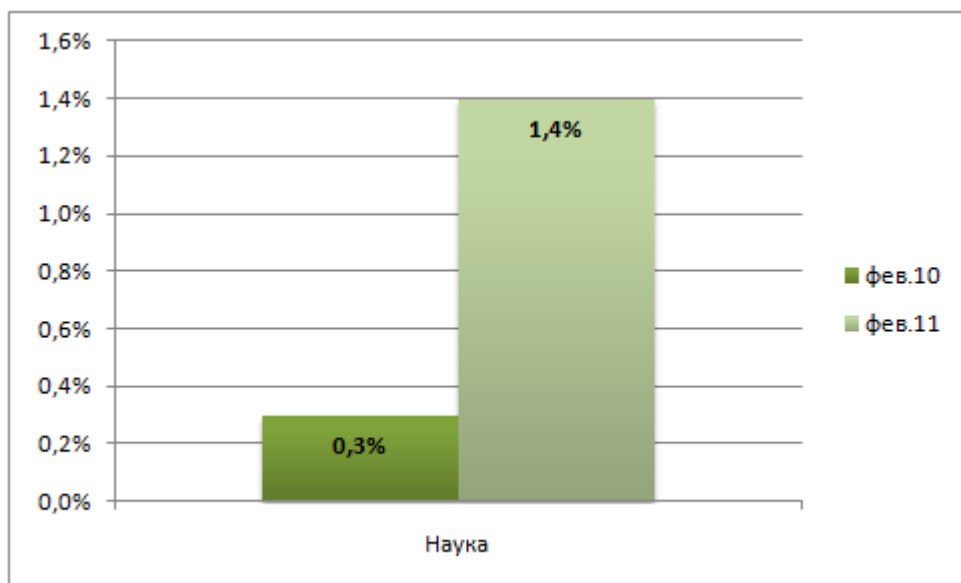


Затем идут художественные тексты, информация, связанная с путешествиями, песни и различные тексты медицинской тематики. Путешествия занимают в интересах пользователей более 2 %, и этот показатель сильно подвержен сезонным колебаниям, ведь интерес к туристическим услугам тоже зависит от сезона.



Путешественники переводят переписку с менеджерами отелей, условия тарифов авиакомпаний, информацию о достопримечательностях и городах.

Перевод материалов по научным дисциплинам (в основном по химии, математике, психологии, биологии), хоть и занимает в интересах всего 1,5 %, но зато именно в этой категории за год произошел самый внушительный рост. Чего не скажешь о нецензурной лексике, которой стало на сайте существенно меньше.





Из интересного также можно отметить, что среди записей, которые мы в данном исследовании определили как «Прочее», заметную часть занимают хобби: вязание, домоводство, рыбалка, охота, садоводство, кулинария.