

1.2 Иностранные языки

УДК 81'322.4

МАШИННЫЙ ПЕРЕВОД: ПРЕИМУЩЕСТВА И НЕДОСТАТКИ

*Ст. преп. Степанов Д.А., студ. Федорцов П.А.
Витебский государственный технологический университет
г. Витебск, Республика Беларусь*

Машинный перевод – процесс перевода текстов (письменных, а в идеале и устных) с одного естественного языка на другой с помощью специальной компьютерной программы. Так же называется направление научных исследований, связанных с построением подобных систем.

Цель работы: исследовать принципы машинного перевода, его виды, проблемы, причины этих проблем и предложить способы их решений. В работе рассматривается эволюция машинного перевода от первых попыток создания до самых современных и совершенных на сегодняшний день систем перевода, которыми ежедневно пользуются люди по всему миру.

Существуют два принципиально разных подхода к построению алгоритмов машинного перевода: основанный на правилах (rule-based) и статистический, или основанный на статистике (statistical-based). Первый подход является традиционным и используется большинством разработчиков систем машинного перевода (PROMT, SYSTRAN, LINGUATEC). Ко второму типу до недавнего времени относились популярные сервисы Яндекс.Переводчик, Переводчик Google, а также ABBYY Lingvo. Сейчас же последние используют оба подхода, комбинируя их с помощью искусственного интеллекта (нейросетей).

Машинный перевод на основе правил (Rule-Based Machine Translation, RBMT) – общий термин, который обозначает системы машинного перевода на основе лингвистической информации об исходном и переводном языках. Они состоят из двуязычных словарей и грамматик, охватывающих основные семантические, морфологические, синтаксические закономерности каждого языка. Такой подход к машинному переводу еще называют классическим. На основе этих данных исходный текст последовательно, по предложениям, преобразуется в текст перевода. Эти системы противопоставляют системам машинного перевода, которые основаны на примерах (статистике). Принцип работы таких систем – связь структуры входного и выходного предложения.

Статистический машинный перевод (англ. Statistical machine translation, SMT) – это разновидность машинного перевода текста, основанная на сравнении больших объемов языковых пар. Языковые пары – тексты, содержащие предложения на одном языке и соответствующие им предложения на втором, могут быть как вариантами написания двух предложений человеком – носителем двух языков, так и набором предложений и их переводов, выполненных человеком. Таким образом, статистический машинный перевод обладает свойством «самообучения». Каждая система автоматического перевода имеет модель перевода – «словарь», в котором для всех слов и фраз на одном языке перечислены возможные переводы на другой язык с указанием вероятности этих переводов. Модели перевода для каждой пары языков содержат миллионы пар слов и словосочетаний.

Непосредственно переводом занимается декодер – алгоритм, который проводит анализ текста и для каждого слова подбирает все варианты перевода, упорядочивая в порядке убывания вероятности. Затем все полученные варианты декодер оценивает с помощью модели языка на частотность употребления и выбирает предложение с лучшим сочетанием вероятности и частотности.

Методы статистического перевода всегда работали через английский язык. При переводе, например, с русского на немецкий, машина сначала переводила текст в английский, а толь-

ко потом переводила на немецкий, что приводило к двойным потерям. Нейронный перевод впервые позволил напрямую переводить между языками, у которых не было ни одного общего словаря.

УДК 81'373.23

НICKNAME КАК ОСОБАЯ РАЗНОВИДНОСТЬ СОВРЕМЕННЫХ АНТРОПОНИМОВ

Преп. Бурдыко О.В., студ. Беляев В.А.

*Витебский государственный технологический университет
г. Витебск, Республика Беларусь*

В ходе нашего исследования мы определили, что антропоним называет, но не приписывает никаких свойств. Антропонимы обладают понятийным значением, в основе которого лежит представление о категории, классе объектов. Этому значению присущи, как правило, следующие признаки:

а) указание на то, что носитель антропонима – человек: Peter, Lewis в отличие от London, Thames;

б) указание на принадлежность к национально-языковой общности: Robin, Henry, William в отличие от René, Henri, Wilhelm;

в) указание на пол человека: John, Henry в отличие от Mary, Elizabeth.

При исследовании мы использовали сплошную выборку в сети Интернет: в блогах, чатах, форумах, в социальной сети ВКонтакте.

Материалом для исследования послужили 20 языковых единиц, представляющих собой никнеймы, которые используются коммуникантами в блогах.

Спорить о корнях и времени появления никнеймов в том виде, как мы их сегодня знаем, можно долго. Хотя скорее всего точкой отсчёта в Интернете можно назвать время появления первых unix-систем, где ник был именем для входа в систему. Никнейм является вашим лицом в сети, которое при желании может отражать вашу сущность, стремления, характер или интересы. Никнейм в Интернете – это не просто способ самоидентификации среди таких же, как вы, сетевых жителей, это также начало названия вашего почтового ящика вида: имя для входа или псевдоним на форуме, в чате, и любом другом сервисе, в котором будет необходимо представиться.

Наше исследование показало, что в основном обучающиеся используют никнеймы при регистрации в социальных сетях и используют для этого различные символы, объясняя это тем, что чем необычнее выглядит никнейм, тем интереснее, а также сложнее для продвинутых пользователей взломать ту или иную страницу.

Из собственного изучения никнеймов можно сделать выводы, что чаще всего используются латинские буквы для написания никнеймов, это же и подтверждает один из выводов опроса. По смысловому принципу в блогах преобладают никнеймы по профессии 30 %, затем имена с заменой символов 20 % и герои 20 %, самый маленький % выбора по названиям животных и растений, психоэмоциональные.

Проанализировав никнеймы, мы пришли к следующим выводам: информационное общество делает реальность самопрезентации окончательной реальностью, перенося этот принцип в реальное взаимодействие. Культура виртуальной реальности подталкивает коммуникантов к структурированию собственного определения, в нашем случае – никнейма, и идентичности между самоидентификацией в сетевых системах и их личностью.