

УДК 81'322

DOI 10.47388/2072-3490/lunn2021-53-1-41-52

ПЕРСПЕКТИВЫ СОВРЕМЕННЫХ СИСТЕМ МАШИННОГО ПЕРЕВОДА

О. И. Кузьмин

Московский государственный лингвистический университет,
Москва, Россия

Современный мир движется в направлении глобальной цифровизации и, в частности, к софтверизации. Повсеместно происходят процессы замены человеческого ресурса на цифровые сервисы или программы, имитирующие выполнение схожих задач. Несомненно, что применение таких технологий в долгосрочной перспективе имеет экономические выгоды для предприятий и компаний. Однако, несмотря на это, качество конечного результата часто оказывается неудовлетворительным. Системы машинного перевода (МП) не являются исключением. До сих пор невозможно обойтись без редакторской правки текста, переведенного с помощью сервисов онлайн перевода. Полностью автоматизированный перевод высокого качества на сегодняшний день не является возможным ввиду многих факторов. Основная причина кроется в тайнах естественного языка: существовании подъязыков, абстрактных слов, полисемии.

Повышение качества работы систем машинного перевода является одной из приоритетных задач обработки естественного языка (*NLP* — *natural language processing*). За последние годы был сделан существенный прорыв в скорости анализа информации, и в будущем этот вопрос не будет первостепенным. Главными критериями оценки качества переведенных текстов будут являться смысловая связность текста и семантическая точность использованных лексических материалов. Для усовершенствования систем МП необходимо внести элементы дифференциации данных и персонализации информации под индивидуальные задачи пользователя. Для этих задач необходимо использование способа тематического моделирования для отнесения текста к той или иной предметной области. В настоящее время существуют алгоритмы, основанные на глубоком обучении, которые способны выполнять данные задачи. Однако процесс выявления уникальных лексических единиц требует более детального лингвистического описания семантических особенностей. Методы парсинга, которые будут задействованы при анализе текстов, должны также предусматривать возможность кластеризации по подъязыкам. Создание электронных автоматизированных словарей профессиональных областей знаний поможет качественно улучшить работу систем машинного перевода. Стоит отметить, что до сих пор не было успешных проектов по созданию словарей машинных систем перевода конкретных подъязыков. Таким образом, возникает потребность в разработке таких словарей и интеграции их в существующие системы онлайн-перевода.

Ключевые слова: машинный перевод; нейросетевой перевод; статистический перевод; гибридный перевод; обработка естественного языка (*NLP*).

Modern Machine Translation Systems: Trends and Prospects

Oleg I. Kuzmin

Moscow State Linguistic University, Moscow, Russia

The modern world is moving towards global digitalization and accelerated software development with a clear tendency to replace human resources by digital services or programs that imitate the doing of similar tasks. There is no doubt that, long term, the use of such technologies has economic benefits for enterprises and companies. Despite this, however, the quality of the final result is often less than satisfactory, and machine translation systems are no exception, as editing of texts translated by using online translation services is still a demanding task. At the moment, producing high-quality translations using only machine translation systems remains impossible for multiple reasons, the main of which lies in the mysteries of natural language: the existence of sublanguages, abstract words, polysemy, etc. Since improving the quality of machine translation systems is one of the priorities of natural language processing (NLP), this article describes current trends in developing modern machine translation systems as well as the latest advances in the field of natural language processing (NLP) and gives suggestions about software innovations that would minimize the number of errors. Even though recent years have seen a significant breakthrough in the speed of information analysis, in all probability, this will not be a priority issue in the future. The main criteria for evaluating the quality of translated texts will be the semantic coherence of these texts and the semantic accuracy of the lexical material used. To improve machine translation systems, we should introduce elements of data differentiation and personalization of information for individual users and their tasks, employing the method of thematic modeling for determining the subject area of a particular text. Currently, there are algorithms based on deep learning that are able to perform these tasks. However, the process of identifying unique lexical units requires a more detailed linguistic description of their semantic features. The parsing methods that will be used in analyzing texts should also provide for the possibility of clustering by sublanguages. Creating automated electronic dictionaries for specific fields of professional knowledge will help improve the quality of machine translation systems. Notably, to date there have been no successful projects of creating dictionaries for machine translation systems for specific sublanguages. Thus, there is a need to develop such dictionaries and to integrate them into existing online translation systems.

Key words: machine translation; neural translation; statistical translation; hybrid translation; natural language processing (NLP).

1. Введение

Современный мир — это мир *high tech*. Высокие технологии, построенные на алгоритмах и вычислениях, напрямую влияют на все сферы жизни человека. Сегодня почти каждый человек в мире обладает смартфоном и компьютером. Информационные технологии (ИТ) и банковские корпорации владеют огромными серверами, которые обрабатывают гигантские потоки информации ежесекундно. Области применения вычислительных технологий разнообразны — от портативных устройств и мобильных телефонов до роботов и космических ракет. Однако необходимо отметить,

что мощность процессоров, как ни странно, играет в вычислениях не ключевую роль. Без программной составляющей ни один процесс обработки информации невозможен. Еще в 1935 г. Алан Тьюринг предложил первую теорию программного обеспечения (ПО) или *software* (Turing 1937). В последующие годы и до сегодняшнего дня происходит постепенный процесс усовершенствования ПО, алгоритмов и способов обработки информации. Софтверизация и разработка новых программных продуктов с недавнего времени получили существенный потенциал развития и стали играть едва ли не ключевую роль в сфере развития ИТ (Kuleshov, Yusupov 2016).

Современные подходы к добыче данных, интеллектуальному и глубинному анализу получили название *data mining*. Такие подходы позволяют изучать большие массивы информации и с помощью фильтров выделять из них значимые компоненты (Witten, Frank, Hall 2011). Основной задачей систем, работающих на основе этих технологий, является «обнаружение знаний в базе данных (БД)» (англ. *knowledge discovery in databases, KDD*). Алгоритм поиска основан на технологии нечеткой логики (*fuzzy logic*) (Zadeh 2002). В процессе поиска данных используются методы обнаружения «скрытых знаний», особой информации, для поиска которой необходимы дополнительные условия и настройки. Запрос к БД и работа с ними возможна на языке программирования *SQL*, который предлагает возможности создания, изменения и извлечения хранимых данных (Chamberlin 2012). Алгоритмы поиска используются для проверки найденных закономерностей и впоследствии сверяются с исходной базой.

Также для задач анализа данных используется парсер (*parser*), или синтаксический анализатор. Инструменты парсинга (*parsing*) предназначены для извлечения и сбора любой открытой информации. Они извлекают необходимую информацию и сохраняют в удобном структурированном виде для последующего использования (Aho, Ullman 1972). С лингвистической точки зрения полученный материал может быть использован для дифференциации лексем подязыков. В этой связи необходимо разграничить общую базу данных лексического материала языка и базы данных отдельных подязыков для формирования нелинейных запросов.

При рассмотрении вопроса извлечения данных и представления их в какой-либо форме необходимо также упомянуть «науку о данных» (*data science*) (Cleveland 2001). Этот раздел информатики напрямую связан как с «большими данными» (*big data*), так и с искусственным интеллектом (ИИ). Специалист в данной области (*data scientist*) занимается вопросами анализа массивов информации и проектированием, разработкой баз данных с применением как статистических методов, так и методов моделирования графовых вероятностных моделей (Airoldi 2007).

2. Виды современных систем машинного перевода

Машинный перевод (МП) является популярным направлением развития ИТ. Технологии МП востребованы многими специалистами из разных стран, которые не владеют иностранными языками, но нуждаются в понимании текстового материала или взаимодействии с носителями других языков с целью обмена опытом и получения новой информации. С каждым годом появляются все новые и новые сервисы перевода, улучшается скорость и качество их работы. Системы МП можно встроить в браузер, и при работе со страницами на иностранном языке больше не нужно запускать дополнительные программы онлайн-перевода отдельно, поскольку текст сразу же будет конвертироваться на родной для пользователя язык. Несомненно, такие нововведения облегчают пользовательский опыт (*user experience*) и делают Интернет более универсальным и общедоступным, стирая невидимые границы между языками, народами и культурами.

Наиболее популярными сервисами онлайн-перевода являются *Google Translate* и его прямой конкурент Яндекс.Перевод. Если сравнивать качество работы двух ключевых игроков на рынке подобных систем, то *Google Translate*, несомненно, имеет преимущество, так как его сервисами пользуются ежедневно миллиарды людей по всему миру. На данный момент компания *Google* является самой крупной и влиятельной зарубежной корпорацией в сфере ИТ. Благодаря тому, что поисковые запросы и тексты сохраняются в кэш-памяти поисковой системы *Google*, БД *Google* на данный момент является самой точной в мире. Конкурентом американским разработчикам выступает Яндекс.Перевод, который широко задействован только в России и странах СНГ, и, следовательно, имеет в своей базе меньше данных о других языках, кроме русского и английского. Таким образом, исходя из приведенного выше аргумента, можно сделать очевидный вывод о том, что *Google Translate* на данный момент времени переводит лучше остальных сервисов онлайн-перевода из-за наличия большой статистической базы текстов на разных языках. Российский сервис Яндекс.Перевод обучается на статистике Рунета и опирается на русскоязычные тексты, следовательно, качество перевода с / на русский язык у системы Яндекс.Перевод будет выше.

С точки зрения профессиональной терминологии перевод *Google Translate* будет более точным из-за большого количества содержащихся в его поисковых и веб-сервисах лексических единиц, которые относятся к самым различным областям (начиная от бытовых заканчивая научными). Данные лексические единицы хранятся в дата-центрах и на серверах компании *Google*, многократно задействуются в процессе работы системы или

обучения алгоритмов. Благодаря нейросетевым технологиям появилась возможность находить закономерности в информационных текстовых потоках, и в результате сложилась относительно точная система перевода, фактически не имеющая критических недостатков. Более крупные и влиятельные корпорации, такие как *Google*, обладают более мощными вычислительными и поисковыми ресурсами. Таким образом, логика современных машинных систем перевода достаточно простая: больше БД — лучше качество перевода.

На сегодняшний день существует несколько технологий, на основе которых работают современные системы машинного перевода. Среди них можно выделить **статистический метод (SMT)**. Это метод определения вероятности, с которой то или иное предложение может являться переводом исходному варианту. Системы статистического перевода являются наиболее распространенными среди других систем АП. Алгоритм работы основан на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов. Именно размер базы данных будет влиять в данном случае на качество перевода. Преимуществом статистического перевода является свойство «самообучения». Несмотря на все явные преимущества такой технологии, полученный результат не всегда соответствует требованиям к высокому качеству финального текста. Часто в нем можно найти большое количество ошибок, неточностей и смысловых искажений. Стоит отметить крайнюю неэффективность работы таких систем применительно к большим объемам профессиональных текстов, в которых встречаются уникальные случаи употребления того или иного слова и его контекстного значения. Подобные случаи невозможно встроить в алгоритмы статистики из-за простой валентности (возможность сочетания слова с другими лексическими единицами минимальна). Проблема выбора правильного значения слова является крайне сложной даже для профессионального лингвиста или переводчика, а для систем АП она и вовсе является невыполнимой.

Нейросетевой метод АП (NMT) появился относительно недавно, но уже успел стать культовым в кругах ИТ из-за тренда на так называемые технологии искусственного интеллекта (ИИ). Он имеет массу преимуществ перед статистическим подходом. Главное преимущество — это возможность обучения на конкретных примерах, а не на всей выборке в целом. Система может сама улучшать качество своей работы. Существует два вида обучения нейросетей: с учителем и без учителя. При обучении с учителем нейронная сеть обучается на размеченном наборе данных и предсказывает ответы, которые используются для оценки точности алгоритма на обучающих данных. При обучении без учителя модель использует нераз-

меченные данные, из которых алгоритм самостоятельно пытается извлечь признаки и зависимости (Goldberg 2017).

Для большей точности результатов перевода с недавнего времени применяется технология, сочетающая правила статистики и нейронных сетей. Такой подход называется **комбинированной или гибридной системой перевода (HMT)**. Современные переводческие онлайн-системы, такие как Яндекс и *Google*, с недавнего времени перешли на комбинированную систему перевода. Это означает, что *SMT* и *NMT* работают в связке и дополняют друг друга. Там, где не справляется с переводом *SMT*, ему на помощь приходит алгоритм *NMT*. В результате качество перевода существенно повысилось.

Несомненно, современные системы машинного перевода, благодаря новым внедряемым технологиям, получили прирост в эффективности и качестве обрабатываемых корпусов текстов. Однако, несмотря на это, до сих пор остаются пробелы в качестве их работы и насущные проблемы, которые до сих пор не были разрешены. К примеру, современным системам не хватает модулей классификаторов, которые будут не только извлекать информацию, но и решать вопрос лексической неоднозначности (полисемии). В этой связи целесообразно применять новые способы разметки текстовых массивов, маркировки частей речи для улучшения качества семантического поиска по тексту и базам знаний (Bast, Bjorn, Haussmann 2016). Необходимость внедрения данных технологий обусловлена следующим фактором: статистический метод, используемый в большинстве систем машинного перевода, часто допускает лексические неточности, связанные, прежде всего, с полисемией. Таким образом, если бы система могла анализировать допущенные ошибки, используя при этом уже размеченные массивы текстов, содержащие уведомления о вероятности полисемии и отмечать другие лексические особенности, то многих ошибок можно было бы избежать.

3. Современные технологии и машинный перевод

Активное использование цифровых и веб-технологий привело к тому, что появилось большое количество баз данных. Также увеличился и объем информации, содержащейся в них. Это привело к затруднениям, связанным со структуризацией, хранением и обработкой информации. Часто вычислительных мощностей обычных компьютеров не хватает или вычисления происходят крайне медленно, что не удовлетворяет требованиям, заложенным разработчиками программ. Таким образом, очевидно, что возникает необходимость использования более быстрых и современных методов анализа. Для таких задач оптимально подходят квантовые компьютеры, которые предлагают возможности ускоренных вычислений и

моделирования нейронной сети экспоненциального размера. Квантовые компьютеры способны выполнять задачи машинного обучения. Так, в 2013 году корпорация *Google* открыла лабораторию квантовых вычислений в области изучения вопросов разработки искусственного интеллекта (*Google AI Blog: Launching the Quantum Artificial Intelligence Lab (googleblog.com)*).

Говоря о технологиях и способах хранения информации, стоит отметить все набирающую популярность технологию блокчейн (*blockchain*). Принимая во внимание ошибки, возникающие в процессе использования систем автоматизированного перевода, предлагается использовать технологию *blockchain* для усовершенствования работы систем машинного перевода. Для переводческой индустрии использование концепции блокчейна будет иметь такие же преимущества, что для крупных сфер бизнеса, таких как банковское дело или управление персоналом. Главное — это создать эффективную среду, в которой зашифрованная информация циркулирует между базами подязыков без централизованной системы управления. Определенные метаданные добавляются в блокчейны, которые содержат зашифрованную информацию о конкретном случае употреблении слова подязыка. Также технология блокчейн предусматривает существование открытого ключа, который можно будет использовать для расшифровки того или иного значения. Преимуществом такого подхода является более быстрый и точный оборот информации.

В отличие от основных программных решений для баз данных метаданные и блокчейны можно изменять и корректировать в процессе работы. Главным преимуществом такой технологии именно для систем машинного перевода является возможность отследить каждый этап процесса перевода (*translation tracking*) и возможность заглянуть в «черный ящик» (то есть увидеть подробный журнал всей проделанной работы в мельчайших деталях), а также, в случае необходимости, внести собственные авторские коррективы в процесс перевода и задать ему правильное направление.

Для того чтобы текст получился правильным и корректным, необходимо редактирование текста специалистом, который проверяет содержание на предмет конкретных терминов и точности. В нашем случае предлагается использовать исходные тексты подязыка как шаблоны для выявления неточностей. Для финального контроля качества необходимо использовать главный (основной) алгоритм для проверки переведенного материала по его завершении. Это означает, что после того, как файл расшифрован, существует полная запись изменений, внесенных каждой базой данных подязыка с указанием того, когда и что было сделано. Это позволяет видеть полную историю, возможные ошибки на каждом этапе, а также варианты правильного и неправильного перевода, выбирать из большего коли-

чества вариантов, даже если они с точки зрения системы содержат ошибки. Также предлагается добавить в систему возможность выдавать несколько разных вариантов перевода на основе того, какая лексика встречается в переводимом отрывке. Система должна уметь анализировать массив текста, делать вывод на основе представленного лексического материала и отображать несколько вариантов перевода, ранжированных в зависимости от вероятности того, к какому подязыку относится данный текст.

4. Разработка ПО для систем машинного перевода

Несмотря на большой прогресс, достигнутый в области работы машинного перевода за последние десятилетия, все же остается ряд моментов, которые необходимо улучшать. Основная проблема сконцентрирована на семантическом уровне языка из-за того, что некоторые слова не имеют точного значения ввиду своей абстрактности. До сих пор не существует программного способа представления знаний о мире, которые обладает человек. Если рассматривать тенденции развития систем машинного перевода, то в этой связи необходимо отметить все возрастающую популярность метода ранжирования слов и создания так называемых словарей конкретности / абстрактности (*dictionaries of concreteness / abstractness*) (Borghi, Binkofski, Castelfranchi, Cimatti, Scorolli, Tummolini 2017). Данный метод основан на вычислении сходства между векторами слов и расстоянием между словом и остальными словами из случайной выборки конкретных / абстрактных слов. Каждое слово сопоставляется с выборкой абстрактных или конкретных слов, и с помощью векторного представления вычисляется коэффициент подобия. Конкретные слова языка — это такие слова, которые описывают предметы или явления, существующие в окружающем мире. Абстрактные слова, наоборот, не имеют конкретных объектных представлений в реальном мире, значения таких слов возможно раскрыть с помощью описательного и объяснительно-иллюстративного метода.

Перевод абстрактных слов представляет наибольшую трудность как для статистических систем, так и для нейросетей, так как информация, содержащаяся в абстрактных словах, относится напрямую к чувствам, эмоциям и впечатлениям, их невозможно алгоритмизировать и кодифицировать. Понимание значения таких слов у человека происходит интуитивно (методом ассоциации), а моделирование представлений и понятий осуществляется с помощью умозрения. Способность мозга проецировать мысли для воссоздания образов в реальном мире, моделировать ситуации, прогнозировать и придавать всему этому лингвистическую оболочку послужила механизмом эволюционного развития человечества. Факт существования

абстрактных слов относится к основным парадоксам соотношения языка и мышления. Согласно исследованиям на понимание значения абстрактных слов, а также их генерации человек затрачивает больше времени и усилий по сравнению с конкретными (Kroll, Merves 1986). Попытки создания электронных словарей конкретности / абстрактности уже существуют. (Solovyev, Ivanov, Akhtyamov 2019). Ранжирование слов в зависимости от конкретности / абстрактности является необходимым условием для усовершенствования работы систем машинного перевода, также необходимо формировать не просто общие словари конкретности / абстрактности, а использовать их применительно к подязыкам. Для усовершенствования работы систем предлагается создать базы данных конкретности / абстрактности каждой подобласти языка и в дальнейшем внедрить их в систему вместе с исходными текстами подязыков для интеграции их с общей системой перевода. Такой подход позволит не просто представлять слова в виде цифр и алгоритмов, но и в какой-то степени приблизит автоматизированные системы к программному моделированию процесса восприятия текста человеком и внесет существенный вклад в развитие области обработки естественного языка (*NLP — natural language processing*).

Препятствием в реализации задуманного может послужить существующая разница в естественных языках. Так, в русском языке больше абстрактных существительных, эквиваленты которым едва ли возможно найти в других языках. В английском и немецком языках, наоборот, больше глаголов. Среди них встречаются такие, которых нет и никогда не существовало в других языках. Во многих языках мира твердо укоренились так называемые англицизмы. Использовать английский вариант проще, чем придумывать эквивалент и сложное объяснение на своем языке и вводить его в обиход. В качестве иллюстрации можно привести английское слово *usability*, которое получило популярность в среде ИТ благодаря программистам и разработчикам веб-интерфейсов. В русском языке уже можно встретить слово «юзабилити» и его вариант — «юзабельный» (от англ. *usability* — «удобство и простота использования, степень удобства использования») (Cambridge Online Dictionary). Для подобных слов невозможно использование метода прямого перевода (трансфера) слово в слово, а будет применяться другая технология, основанная на подходе к описанию предмета, процесса или явления. Для этого предлагается создать и внедрить в существующие системы специальные словари, которые будут фиксировать уникальные слова, эквивалентов которым на других языках не существует. В случае с приведенным выше примером невозможно обойтись без использования метода транслитерации и введения в язык нового, ранее не существующего понятия.

Необходимым условием развития и улучшения качества систем машинного перевода является возможность дополнения или редакторской правки результатов машинного перевода. Для этой цели предлагается создать функцию правки текста, и предложенный пользователем вариант будет использован для улучшения качества машинного перевода, а лучший из предложенных вариантов должен быть отмечен как наиболее вероятный. Такой подход уже успешно применяется в системе машинного перевода *Google Translate* (*Help Us Improve the Google Translate Tool* — <https://translate.google.com/intl/en/about/contribute/>). Благодаря такому нововведению пользователи смогут активно участвовать в улучшении качества и вносить собственный вклад в развитие данных систем, редактируя варианты перевода и отправляя свои пожелания. Это позволит сделать систему актуальной, динамичной и гибкой.

5. Заключение

Невозможно отрицать технологический прорыв, благодаря которому появились технологии МП, которыми ежедневно пользуется большое количество людей. Сегодня в распоряжении пользователей огромное многообразие как онлайн-сервисов, так и автономных программ, не требующих подключения к Интернету.

Несмотря на то, что системы машинного перевода постоянно совершенствуются, остаются неразрешимые лексические трудности, связанные с языковыми особенностями. До сих пор допускается множество ошибок из-за невозможности формализации многих лингвистических явлений. Предложенные в статье программные методы улучшения качества перевода послужат теоретической рекомендацией в разработке систем машинного перевода.

Стоит отметить, что, осуществляя перевод с одного языка на другой, необходимо понимать, к какой предметной области, подязыку относится конкретное слово. Для этого необходимо ввести дополнительные классификаторы и настройки, на основе которых будут формироваться словари каждого подязыка. Также остро ощущается нехватка лингвистических знаний, которые необходимо интегрировать в системы обработки текстовой информации и системы машинного перевода. Предлагается создать дополнения для онлайн-переводчиков с целью усовершенствования качества перевода в подязыках. Кроме того, необходимо создать базы знаний языков для выявления индивидуальных языковых особенностей. Модули дополнений можно будет интегрировать в систему перевода. Так, если система встречает сочетания слов, которые определяются как конкретный подязык, то поиск результатов перевода будет происходить не по общей

базе, а по модулю подязыка. Кроме подязыков предлагается дополнить системы словарями конкретности / абстрактности, которые помогут избежать неточностей перевода. Несомненно, результаты, достигнутые в процессе изменений и дополнений, вносимых в системы машинного перевода, должны коррелировать с оценками экспертов-лингвистов, тогда данный подход будет релевантным для последующего внедрения в алгоритмы уже существующих систем.

На современном этапе системам машинного перевода не хватает узкой направленности и конкретики. Системы уже научились качественно переводить тексты по общей тематике. Однако в БД корпораций, таких как *Google* и Яндекс, на поисковых системах, на которые работают онлайн-переводчики, до сих пор не существует четкой классификации по областям знаний. Возможности каждого пользователя выставлять персонализированные настройки, самостоятельно контролировать степень достоверности перевода на основе шаблонов текстов подязыков и выбирать из предлагаемых Сетью подходящие варианты являются теми необходимыми нововведениями, которых не хватает современным системам машинного перевода. Фактически необходим индивидуальный подход, то есть фокусировка внимания на частных случаях употребления и возможность заглянуть в «черный ящик», достать из него лишь то, что нужно пользователю.

Список литературы / References

- Aho, Alfred V., & Ullman, Jeffrey D. (1972) *The Theory of Parsing, Translation, and Compiling*. Prentice Hall Professional Technical Reference.
- Airolidi, Edoardo M. (2007) Getting Started in Probabilistic Graphical Models. *PLoS Computational Biology*, vol. 3, 12, e252. DOI: 10.1371/journal.pcbi.0030252.
- Bast, Hannah, & Buchhold, Björn, & Haussmann, Elmar. (2016) Semantic Search on Text and Knowledge Bases. *Foundations and Trends in Information Retrieval*, vol. 10: 2–3, 119–271. DOI: 10.1561/15000000032.
- Borghi, Anna, & Binkofski, Ferdinand, Castelfranchi, Cristiano, Cimatti, Felice, Scorolli, Claudia, Tummolini, Luca. (2017) The Challenge of Abstract Concepts. *Psychological Bulletin*, 143(3), 263–292. DOI: 10.1037/bul0000089.
- Chamberlin, Donald D. (2012) Early History of SQL. *IEEE Annals of the History of Computing*, vol. 34, 4, 78–82.
- Goldberg Yoav, & Hirst, Graeme. (2017) *Neural Network Methods for Natural Language Processing (Synthesis Lectures on Human Language Technologies)*. Morgan & Claypool Publishers.
- Kroll, Judith F., & Merves, Jill S. (1986) Lexical Access for Concrete and Abstract Words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12 (1), 92–107. DOI: 10.1037/0278-7393.12.1.92.
- Kuleshov, Sergey V., & Yusupov, Rafael M. (2016) Is Softwarization the Way to Import Substitution? *SPIIRAS Proceedings*, 3 (46), 5–13. DOI: 10.15622/sp.46.1.

- Neven, Hartmut. (2013) *Launching the Quantum Artificial Intelligence Lab*. Retrieved from <https://ai.googleblog.com/2013/05/launching-quantum-artificial.html>.
- Solovyev, Valery, & Ivanov, Vladimir, & Akhtyamov, Rauf. (2019) Dictionary of Abstract and Concrete Words of the Russian Language: A Methodology for Creation and Application. *Journal of Research in Applied Linguistics*, 10 (SP), 218–230. <https://doi.org/10.22055/RALS.2019.14684>.
- Turing, Alan. (1937) On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, vol. 2-42, 230–265.
- Cleveland, William, S. (2001) Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review: Journal*. Willey & Sons, vol. 69, 1, 21–26.
- Witten Ian H., & Frank Eibe, & Hall Mark A. (2011) *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd Edition. Morgan Kaufmann.
- Zadeh, Lotfi. (2002) From Computing with Numbers to Computing with Words — from Manipulation of Measurements to Manipulation of Perceptions. *International Journal of Applied Math and Computer Science*, vol. 12, 3, 307–324.

Ссылки на электронные ресурсы / Electronic resources

- Яндекс.Переводчик [Электронный ресурс]. URL: <https://translate.yandex.ru> [Yandex.Perevodchik. Retrieved from <https://translate.yandex.ru>].
- Google Translate. Retrieved from <https://translate.google.com>.
- Usability. *Cambridge Dictionary*. Retrieved from <https://dictionary.cambridge.org/dictionary/english/usability>.