

МОДЕЛЬ ДЛЯ ОЦЕНИВАНИЯ ФУНКЦИОНАЛЬНОСТИ СИСТЕМ МАШИННОГО ПЕРЕВОДА¹

© 2018 г. А. В. Костин, В. В. Смирнов

Москва, Центральный экономико-математический ин-т РАН
e-mail: kostin.alexander@gmail.com, vitaly_smirnov@mail.ru

Поступила в редакцию 31.05.16 г. После доработки 22.06.17 г.

Представлены основные понятия, модели и методы оценивания качества программного обеспечения. В связи с противоречивостью определений понятия “функциональность”, имеющей место в существующих стандартах в области оценивания качества программного обеспечения, в работе это понятие определено как частный случай понятия качество в квалиметрии применительно к оцениванию качества программного обеспечения. Предложена обобщенная модель оценивания функциональности программного обеспечения. Представлена разработанная на ее основе модель оценивания функциональности систем машинного перевода, которая может служить прототипом для построения более детализированных моделей. Рассмотрены некоторые способы получения численных оценок функциональности систем машинного перевода с помощью таких моделей.

DOI: 10.7868/S0002338818010158

Введение. Машинный перевод быстрее ручного и при обработке текстов большого объема высвобождает человеческие ресурсы, которые могли бы быть заняты ручным переводом. Освободившиеся ресурсы возможно использовать для развития рынка машинного перевода (Machine Translation (MT)). Однако доступность бесплатных средств MT, таких, как Google Translate, Microsoft Bing Translator, Яндекс.Переводчик, Translate.Ru (сервис онлайн-перевода ООО “ПРОМТ Сервис”), WorldLingo и др., и отсутствие точности перевода, вероятно, сдерживают рост данного рынка. Недостаточно удовлетворенные потребности рынка MT связаны с повышением качества перевода, востребованностью специальных навыков, необходимостью постредактирования, преодолением конкуренции со стороны поставщиков услуг бесплатного перевода, а также с трудностями в измерении и оценивании качества [1].

На современном уровне развития средств MT не удается полностью исключить человека из переводческого процесса. Человек может участвовать в процессе через исправление результатов MT, подготовку текста к компьютерной обработке, участие в самом процессе перевода. В последнем случае, в зависимости от степени участия человека, можно говорить либо об *интерредактировании* [2], либо об *автоматизированном переводе* [3]. Системы MT различных производителей могут отличаться соотношением количества функций, выполняющихся автоматически, и автоматизированных функций. В этой ситуации решения по использованию инноваций, принимаемые производителями средств MT, не приводят к кардинальным изменениям в продуктах, связанных с сокращением числа автоматизированных функций за счет улучшения качества функций, выполняющихся автоматически. Кроме того, стандартизованные подходы к оцениванию качества программного обеспечения (ПО) не в полной мере соответствуют рекомендациям квалиметрии – науки о количественном оценивании качества. Развитию рынка средств MT в этих условиях может способствовать разработка новых оценочных моделей, упрощающих производителям принятие обоснованных решений по совершенствованию возможностей предлагаемых ими продуктов на основе соотнесения их функциональности с потребностями рынка.

1. Модели и методы оценивания функциональности ПО. 1.1. Правомерность оценивания функциональности ПО с помощью методов оценивания качества. *Функциональностью (функциональными возможностями, взаимосвязанным набором функций, функциональными свойствами) ПО* в данной работе будем называть взаимосвязанную совокупность его потребительских свойств, за исключением характеристик использования ресурсов для его создания, хранения,

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 14-18-01999).

применения и удаления с запоминающих устройств. Заметим, что определенное таким образом понятие “функциональность” применительно к ПО является частным случаем понятия “качество” в квалиметрии.

Необходимость приведенного выше определения понятия “функциональность” вызвана тем, что в стандартах по оценке качества ПО, в частности ГОСТ Р ИСО/МЭК 9126-3 и заменяющем его стандарте ГОСТ Р ИСО/МЭК 25010-2015, понятие “качество” и связанные с ним понятия заметно отличаются от принятых в квалиметрии. Так, анализируя определение понятия “качество программного обеспечения” в ГОСТ Р ИСО/МЭК 9126-93, определяемое в нем как весь объем признаков и характеристик программной продукции, который относится к ее способности удовлетворять установленным или предполагаемым потребностям, нетрудно заметить, что оно является частным случаем квалиметрического понятия “интегральное качество” применительно к ПО. А функциональность ПО в этом же ГОСТе понимается почти в том же смысле, что и в данной работе, а именно что функциональность характеризует реализацию установленных или предполагаемых потребностей. Однако в ГОСТ Р ИСО/МЭК 25010-2015 функциональность ПО противопоставлена его качеству. Функциональность и качество ПО в нем рассматриваются как две части *неотъемлемых свойств* ПО. Но, в соответствии с этим же ГОСТом функциональные свойства определяют, что ПО в состоянии сделать, а свойства качества определяют, насколько хорошо ПО выполняет свои функции. Из чего логически вытекает, что свойства качества в ГОСТ Р ИСО/МЭК 25010-2015 являются лишь оценками функциональных свойств, что приводит к противоречию в этом ГОСТе между понятиями “функциональность” и “качество”.

В ГОСТ Р ИСО/МЭК 25010-2015 введено понятие “качество системы”, означающее “степень удовлетворения системой заявленных и подразумеваемых потребностей различных заинтересованных сторон, которая позволяет, таким образом, оценить достоинства”. Это понятие также можно рассматривать как частный случай квалиметрического понятия “интегральное качество” применительно к ПО. Предлагаемый в этом ГОСТе тип оценочных моделей, которые представляют собой иерархические декомпозиции сложного свойства “качество системы”, отличается от традиционного квалиметрического дерева свойства отсутствием правил их построения. Кроме того, несмотря на то, что из контекста понятно, что имеют место отношения А-PART-OF (часть-целое) между характеристиками (свойствами) различных уровней иерархии, для обозначения совокупности характеристик, принадлежащих к одному множеству, вместо термина “группа” используется термин “класс”, что вводит в заблуждение, заставляя думать, что речь идет об отношениях IS-A (таксономии).

Существуют также другие типы оценочных моделей, применимых для оценивания ПО. Перечислим некоторые из них: модели “факторы-критерии-метрики”; модели качества ПО, которые предложил J. McCall и расширил В. Boehm, в ней факторы характеризуют ПО с позиций пользователей и задаются требованиями, критерии характеризуют ПО с позиций разработчика и задаются как цели, метрики предназначены для количественного измерения и оценивания качества [4]; модели FURPS+ R. Grady [5], название которых образовано от первых букв слов *functionality, usability, reliability, performance, supportability* (функциональность, удобство работы, надежность, производительность, возможность поддержки), а “+” обозначает ограничения проектирования (*design*), разработки (*implementation*), интерфейса (*interface*), физические ограничения (*physical*); модели “цель-вопрос-метрика” [6]; модели “процесс/продукт” [7,8]; модели “цели-критерии-альтернативы”, процесс построения которых обычно основывается на методе анализа иерархий (МАИ) [9]; модели “функции-объекты-свойства”, являющиеся расширением квалиметрических деревьев свойств [10].

Из того, что мы определили понятие “функциональность” таким образом, что применительно к ПО оно оказалось частным случаем понятия “качество” в квалиметрии и в ГОСТ Р ИСО/МЭК 9126-93, а в ГОСТ Р ИСО/МЭК 25010-2015 свойства качества являются оценками функциональных свойств, следует правомерность оценивания функциональности ПО с помощью методов квалиметрии, которые используются для оценивания качества.

1.2. Методы определения ситуаций оценивания программного обеспечения. В основе любого метода оценивания качества лежит операция сравнения, которая устанавливает такие отношения между объектами или свойствами объектов, которые позволяют утверждать, имеют ли одни объекты преимущества (достоинства) или недостатки перед другими объектами. Например, это могут быть бинарные отношения эквивалентности и превосходства [11, 12], отношения строгого порядка или нестрогого квазипорядка [13]. Сравнение выполняется специалистом по оцениванию качества на основе имеющихся в его распоряжении данных о свойствах объектов. Этот процесс всегда выполняется с точки зрения субъектов (стейкхолдеров), которые заинтересованы либо в самом процессе, либо в его результатах. В квалиметрии, чтобы учесть их интересы, первым этапом оценивания является этап *определения ситуации оценивания (требований к оцениванию)*, который включает в себя: определение

стейкхолдеров; определение зависящих от интересов стейкхолдеров требований к источникам и методам получения информации о свойствах и значениях свойств оцениваемых объектов.

Специалист по оцениванию качества ПО вынужден определять ситуацию оценивания в соответствии с действующей в компании-производителе ПО *системой управления качеством*, чтобы учитывать интересы потребителя ПО, самой компании-производителя, других стейкхолдеров, включая отдельных общественных субъектов, государство и общество в целом. Под *управлением качеством* будем понимать целенаправленный процесс скоординированных воздействий на объекты управления для установления, обеспечения и поддержания необходимого его *уровня качества*, удовлетворяющего требованиям стейкхолдеров [14].

Уровень качества характеризует результат сравнения *полезности*, определенной по совокупности достигнутых значений показателей качества рассматриваемого объекта (на одной из стадий его жизненного цикла или на всей совокупности этих стадий), применительно к условиям его использования, с соответствующими значениями показателей образцов конкурентов, эталонов, стандартов, аналогов и т.п. Методы оценивания уровня качества можно разделить на дифференциальные, комплексные и смешанные методы [14].

Если в управлении качеством приоритетным является *ориентация на потребителя*, то такая ситуация оценивания будет соответствовать первому принципу менеджмента качества, основанного на популярной в настоящее время концепции *всеобщего управления качеством* (Total Quality Management (TQM) [14]. Популярность данной концепции подтверждается тем, что многие фирмы-производители проходят сертификацию своих систем управления качеством на соответствие международным стандартам ИСО серии 9000, в третье издание которых, выпущенное в 2007 г., вошли базовые принципы TQM.

Выполняя определение ситуации оценивания, специалист по оцениванию качества ПО может принимать решения об используемых методах построения оценочных моделей и способах их применения на основе как своего личного опыта и интуиции, так и используя *принятие решений на основе фактов*, имеющихся у компании. Второй подход к определению ситуации оценивания в рамках системы управления качеством соответствует еще одному принципу менеджмента качества в TQM. Для оценивания ПО также имеет большое значение *процессный подход*, рекомендуемый в TQM и предусмотренный ГОСТ Р ИСО 9000-2001, поскольку этот подход акцентирует внимание специалиста по оцениванию ПО на связях ситуации оценивания с процессами, выполняемыми на различных стадиях жизненного цикла ПО.

Кроме TQM существуют также другие концепции управления качеством [15,16]: система обеспечения качества (Quality Assurance System (QAS)); всеобщий контроль качества (Total Quality Control (TQC)); статистический контроль качества (Statistical Quality Control (SQC)); гарантии продукции (Product Assurance (PA)); улучшение качества (Quality Improvement (QI)); передовой производственный опыт; интегрированный менеджмент процессов (Integrated Process Management (IPM)); интегрированный менеджмент качества (Integrated Management (IM)); всеобщий менеджмент качества в сфере охраны окружающей среды (Environmental Management System (EMS)); менеджмент качества в сфере разработки ПО (Software Quality Management (SQM)) и др.

Рассмотрим подробнее SQM, который включает в себя следующие три уровня [16,17]: гарантия качества (Software Quality Assurance (SQA)), которая выполняется путем мониторинга и корректировки процесса разработки с целью получения на выходе качественного продукта; планирование качества (Software Quality Plan (SQP)), в результате которого формируется план действий по достижению целей в отношении качества программного обеспечения и снижения рисков; контроль качества (Software Quality Control (SQC)), осуществляемый в соответствии с действующей моделью жизненного цикла (ЖЦ) программного обеспечения, который для каждого процесса, предусмотренного моделью ЖЦ, обеспечивает проверку правильности входных данных, результатов и самого процесса.

Несмотря на различие моделей ЖЦ ПО, в соответствии со стандартом ISO/IEC 12207 (ГОСТ Р ИСО/МЭК 12207) можно выделить группы процессов ЖЦ, которые в той или иной степени имеют место в каждой модели ЖЦ: *основные процессы*: приобретение, поставка, разработка, функционирование и сопровождение; *поддерживающие процессы*: документирование, конфигурирование, менеджмент конфигурации, оценивание, верификация, валидация (аттестация), анализ связей, аудит и разрешение проблем; *организационные процессы*: менеджмент, инфраструктурные процессы, усовершенствование и тренинг.

Среди перечисленных выше процессов на основании определений ГОСТ Р ИСО/МЭК 12207-2010 к оцениванию функциональности ПО имеют отношение оценивание (evaluation), верификация (verification), валидация (validation) и аудит (audit), поскольку все эти процессы связываются в данном стандарте с различными вариантами установления соответствия ПО требованиям к его

функциональности (качеству). С нашей точки зрения формулировок этого стандарта не достаточно, чтобы выделить существенные особенности каждого из четырех перечисленных выше процессов оценивания функциональности ПО. Поэтому приведем несколько вариантов формулировок, основываясь не только на стандарте, но и на публикациях, связанных с методами оценивания ПО.

Оценивание (оценка) – систематическое определение степени, с которой некоторый объект удовлетворяет установленным критериям (ГОСТ Р ИСО/МЭК 12207-2010). Если критерии задают ожидаемые стейкхолдерами отношения между входом и выходом ПО, то такое оценивание называется *тестированием ПО*. Оно выполняется для информирования стейкхолдеров о качестве ПО [18]. Как отмечено выше, в основе *оценивания функциональности (качества)* лежит операция сравнения, которая устанавливает такие отношения между объектами или свойствами объектов, которые позволяют утверждать, имеют ли одни объекты преимущества (достоинства) или недостатки перед другими объектами [11-13].

Для понятия *валидация* ПО существует большое количество определений. Некоторые из них представлены ниже.

1. *Валидация* – подтверждение (на основе представления объективных свидетельств) того, что требования, предназначенные для конкретного использования или применения, выполнены (ГОСТ Р ИСО/МЭК 12207-2010).

2. *Валидация* – процесс, показывающий, что построен правильный программный продукт [19].

3. *Валидация* определяет, будет ли окончательная программа отвечать потребностям и требованиям пользователя [19].

4. *Валидация* представляет собой процесс оценивания системы или компонент во время или в конце процесса разработки, чтобы определить, действительно ли она удовлетворяет предъявляемым требованиям или нет, и валидация должна достоверно установить, что программное обеспечение подходит для своей цели и делает то, что пользователь действительно требует [20].

Термин *верификация* используется в двух смыслах [21-23]. В более широком смысле он означает комплекс строго регламентированных процедур, применяемых на каждом этапе процесса разработки программного обеспечения для проверки соответствия продукта всем требованиям предыдущего этапа. В более узком смысле верификация означает доказательство правильности программ или доказательство соответствия программы своей спецификации.

Аудит – независимая оценка программных продуктов и процессов, проводимая уполномоченным лицом с целью оценить их соответствие требованиям (ГОСТ Р ИСО/МЭК 12207-2010). В рамках SQA предметом аудита является качество ПО. Также аудит может быть направлен на проверку действия лицензий на ПО или на проверку правильности его конфигурации.

В ГОСТ Р ИСО/МЭК 25010-2015 отмечено, что при сравнении объектов их атрибуты (свойства, характеристики) могут иметь количественные или качественные различия. Это означает, что при разработке стандарта признавалось существование не только *количественных измерений*, но и *качественных измерений*. С особенностями каждого из двух этих видов измерений можно ознакомиться в работе [11]. Тем не менее, в ГОСТ Р ИСО/МЭК 25010-2015 рассматривается только использование оценочных моделей для количественных измерений.

1.3. Методы количественного измерения для оценивания качества программного обеспечения. Когда для оценивания качества ПО выполняются количественные измерения, возникает необходимость в использовании *метрик ПО*, предназначенных для получения численных значений некоторых свойств ПО или его спецификаций [24]. Метрики могут быть различным образом классифицированы. При этом они могут быть связаны с процессами оценивания качества ПО в рамках конкретной модели жизненного цикла ПО или модели оценивания качества ПО. Например, в рамках процесса тестирования ПО можно выделить “первичные или накопительные метрики”, имеющие числовые значения и являющиеся основой анализа тенденций и прогнозирования, и “вычисляемые метрики”, которые не могут быть получены непосредственно, а рассчитываться на основе “первичных метрик” [25]. В соответствии со стандартом ГОСТ Р ИСО/МЭК 25010-2015 различают показатели внутреннего и внешнего качества, соответственно можно различить внутренние и внешние метрики. В работе [26] рассматриваются такие внутренние *статические* метрики, которые связаны со структурными нарушениями ПО, и такие внешние *динамические* метрики, которые относятся к поведенческим аспектам ПО. “Первичные метрики” для оценивания качества ПО могут различаться единицами измерения, из которых наиболее популярными являются измерения, выраженные в количестве аномалий (ошибок, дефектов, багов, отклонений и др.), в количестве строк кода (Lines of Code (LOC)) и в количестве *функциональных точек* (Function Points (FP)) [27]. FP представляет собой комбинацию характеристик

ПО, относящихся к входным и выходным данным, интенсивности использования входных и выходных данных, пользовательскому интерфейсу, находящихся вне ПО интерфейсов, требуемым для работы ПО файлам. Вместе с термином “функциональная точка” используется термин “анализ функциональных точек” (Function Point Analysis (FPA)), отражающий название методик оценивания функциональности на основе FP, в которых обычно присутствуют этап определения функциональных элементов на основе анализа информации о ПО (прототипов, спецификаций, моделей и др.) и этап получения численных значений, характеризующих FP [28, 29].

Для оценивания качества средств МТ применяются специфические метрики. Одна из групп таких метрик основана на N-граммах, которые являются последовательностями из N однотипных элементов (символов, слов, предложений и др.). Подавляющее число метрик основано на N-граммах, к которым относятся BLEU, BLEU4 [30], posBleu, posP, posR, posF, wpF [31], morphF, wpBleu, morphBleu [32], NIST [33], LEPOR (Length Penalty, Precision, n-gram Position difference Penalty and Recall) [34]. Другие примеры метрик: F-measure [35] и ее усовершенствование METEOR (Metric for Evaluation of Translation with Explicit Ordering) [36]; WER (Word Error Rate), TER (Translation Edit Rate) [37]. Примеры комбинированных метрик [31]: mBLEU, mTER.

Метрики оценивания качества МТ применяются для улучшения качества средств МТ, как минимум, двумя путями. Во-первых, метрики используются для оценки качества перевода путем сравнения с эталоном, например, чтобы проверить, как меняется перевод при добавлении новых обучающих корпусов текстов для систем статистического МТ или при изменении программного кода средств МТ [32–36]. Во-вторых, метрики применяются для различных вариантов прогнозирования, направленных на улучшение качества перевода. Например, в [38] рассматриваются метрики BLEU и METEOR как средства прогнозирования человеческих оценок качества МТ. Другим примером является работа [39], в которой представлены вопросы использования метрик BLEU и HTER (Human-argeted Translation Edit Rate) для предсказания усилий, необходимых для выполнения постредактирования МТ-текстов. В [40] метрика BLEU включена в прогнозирование качества МТ с точки зрения пользователя средств МТ, выполняющееся при моделировании интерактивного МТ, в котором обратная связь с пользователем позволяет улучшить качество МТ.

1.4. **Обобщенная модель оценивания функциональности программного обеспечения.** В результате анализа существующих моделей и методов оценивания функциональности (качества) ПО, включая методы оценивания функциональности средств МТ, выделим множества элементов, характерных для данной области, которые можно представить в виде следующей модели оценивания функциональности ПО: $E = \langle F, H, A, M, S \rangle$, где F – множество простых (не декомпозируемых) и сложных (декомпозируемых) оцениваемых функций ПО; $H = \{h \mid \exists \langle f, h \rangle: f \in F, \forall \langle f, h \rangle, \langle f', h' \rangle: h = h' \Rightarrow f = f'\}$ – множество простых (не декомпозируемых) и сложных (декомпозируемых) свойств (характеристик) качества функций ПО; $A = A_F \cup A_H$ – множество атрибутов, значения которых извлекаются из доступных источников информации о функциях ПО или вычисляются на основе значений других атрибутов, $A_F = \{a_f \mid \forall f: f \in F \vee \exists \langle f, a_f \rangle\}$ – множество атрибутов F , $A_H = \{a_h \mid \forall h: h \in H \vee \exists \langle h, a_h \rangle\}$ – множество атрибутов H ; $M = MG \cup MC$ – множество методов оценивания качества ПО, где MG – множество методов извлечения значений атрибутов в A ; MC – множество методов расчета значений атрибутов в A ; $S = \langle I, R \rangle$ – ситуация оценивания, где I – множество стейкхолдеров, заинтересованных в оценивании качества ПО, $R = \{\langle i, r \rangle \mid i \in I, r: M \rightarrow A\}$ – множество ограничений, устанавливаемых стейкхолдерами на применение методов оценивания качества ПО.

Большинство работ, посвященных оцениванию функциональности МТ, связано с исследованием простых свойств, характеризующих качество средств МТ. Оценивание функциональности МТ, основанное на анализе сложных свойств, является мало изученным. Представленная выше обобщенная модель E может служить базой для оценивания функциональности МТ как по простым, так и по сложным свойствам. Ниже рассмотрена одна из возможных частных моделей, построенная с применением элементов модели E и предназначенная для оценивания качества систем МТ.

2. Модель для оценивания функциональности систем МТ. Современные системы МТ используются в различных областях, на разных типах оборудования (серверах, персональных компьютерах, мобильных устройствах), ориентированы на различную степень участия человека в переводческом процессе и предусматривают различные способы ввода и вывода данных (речевой, текстовый, в виде изображений в графическом формате). Тем не менее, можно выделить следующие функции (элементы множества F), в той или иной степени характерные для всех систем МТ: ввод текста (речевой, текстовый, в виде изображений в графическом формате), предварительная подготовка текста, преобразование текста во внутреннее

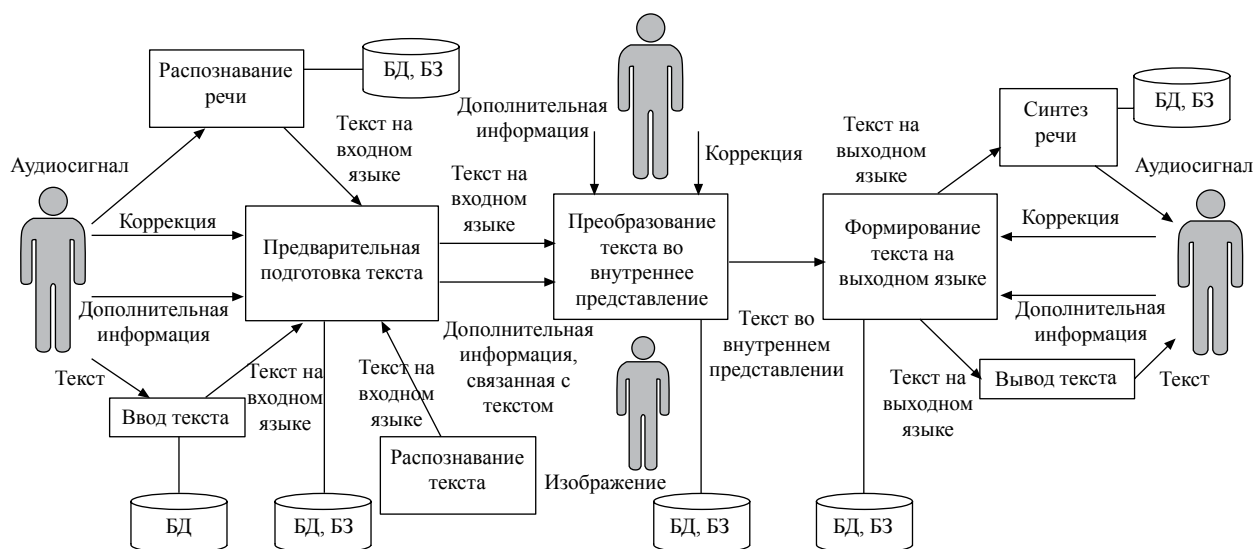


Рис. 1. Функции, характерные для всех систем машинного перевода

представление, формирование текста на выходном языке, вывод текста (речевой, текстовый). Кроме того, все системы МТ используют базы данных (БД) и/или базы знаний (БЗ). Функциональная модель гипотетической системы МТ, имеющей сразу все перечисленные выше функции, представлена на рис. 1.

Показанная на рис. 1 функциональная модель может быть использована для построения моделей оценивания функциональности системы МТ. Применение квалиметрии для построения таких моделей является одним из возможных подходов, представленных в работах [9, 10, 12, 13, 41]. Как отмечено выше, для предлагаемых в ГОСТ Р ИСО/МЭК 25010-2015 иерархических декомпозиций сложного свойства “качество системы” не предусматривается наличие правил их построения. Однако в квалиметрии важно наличие таких правил. Хотя модель Е имеет общие черты перечисленных в предыдущем разделе моделей различных типов, подходящих для оценивания качества ПО, она отличается от каждой из них. Поэтому необходимо вводить специальные правила для построения оценочных моделей функциональности средств МТ, использующих элементы модели Е. Для удобства и соответствия ГОСТ Р ИСО/МЭК 25010-2015 расширим модель Е, добавив корневой элемент h_0 для обозначения комплексного оцениваемого свойства. С учетом наличия элемента h_0 введем правила построения квалиметрической оценочной модели, которые перечислены ниже.

1. Корневому элементу h_0 могут быть подчинены только элементы, обозначающие функции ПО.
2. Элементу, обозначающему функцию ПО, могут быть подчинены либо только другие такие же элементы, либо простые или сложные свойства.
3. Элементу, соответствующему сложному свойству, могут быть подчинены только простые или сложные свойства.
4. Каждому элементу должны присваиваться коэффициенты важности (далее – важность), как это описано в [41].

В соответствии с квалиметрическим подходом в процессе оценивания простым свойствам из оценочной модели должны присваиваться значения на основе информации об оцениваемом объекте, т.е. о средстве МТ. Численные оценки комплексного свойства вычисляются для каждого оцениваемого объекта на основе значений простых свойств и коэффициентов важности. Такое использование значений простых свойств и коэффициентов важности обосновано исследованиями в области психологии оценивания, в соответствии с которыми эти два типа оценок можно соотносить с двумя типами внутренних конструктов: “шкалами ценности” и “шкалами субъективной значимости” [42].

У каждого функционального элемента из F и у каждого свойства из H , как минимум, должны быть по два атрибута (элемента множества A_F и элемента множества A_H соответственно): имя и важность, наличие других атрибутов зависят от конкретного метода оценивания функциональности системы МТ. С учетом участия человека в переводческом процессе у функций “предварительная подготовка текста”, “преобразование текста во внутреннее представление” и “формирование текста на выходном языке” должно

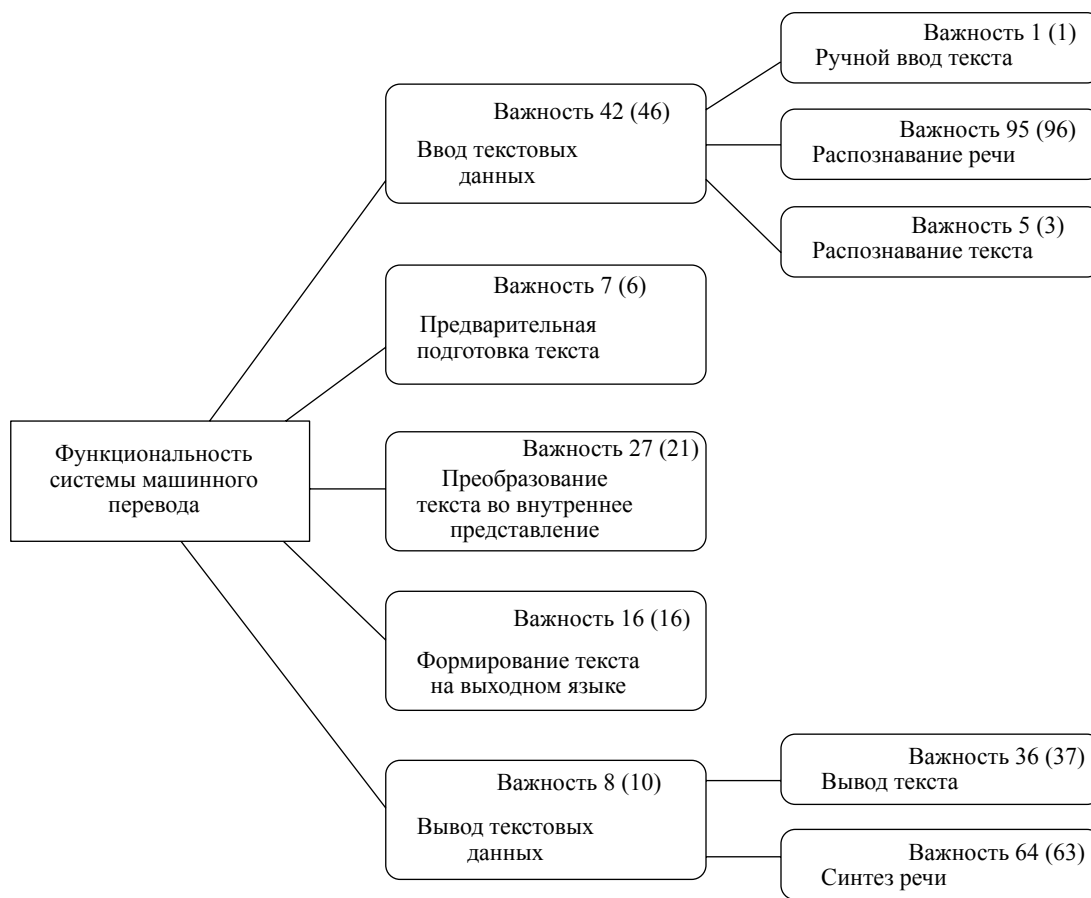


Рис. 2. Функциональные уровни оценочной модели системы МТ

быть свойство “корректируемость”. У всех функций должны быть свойства, характеризующие выходные данные.

С учетом вышесказанного, функциональные уровни оценочной модели системы МТ могут быть представлены так, как показано на рис. 2, а пример использования этих уровней в модели, связывающей функции и свойства системы МТ, – на рис. 3.

3. Определение коэффициентов важности функций систем МТ с помощью статистической обработки патентных документов. Важность элементов оценочной модели может быть определена различными методами. В квалиметрии для этого используются экспертные и статистические методы [13]. Для определения коэффициентов важности функциональных элементов модели, представленных на рис. 2, было разработано программное обеспечение, реализующее статистический метод обработки патентных документов (заявок на изобретение и патентов), основанный на соотношении кодов международной патентной классификации с функциональными элементами модели. Такой способ можно применять исходя из предположения, что, связывая патентный документ с некоторым классом, специалист выражает свое мнение о ценности патента в той области, которая обозначается кодом этого класса. Ценность патентного документа может проявляться, в частности, через получение коммерческой выгоды от использования в товарах или услугах тех новшеств, которые представлены в этих документах.

Разработанное нами программное обеспечение ориентировано на сбор и анализ патентных документов с применением ключевых слов, международной патентной классификации и ссылочной информации, содержащихся в патентных документах, включая ссылки на патентные документы. Данное узкоспециализированное программное обеспечение было создано специально для выполнения представленного в данной работе исследования совместно с программой Qualimetry Studio [43], предназначенной для автоматизированного построения и использования квалиметрических оценочных моделей.

Отбор патентных документов выполнялся по кодам международной патентной классификации G06F17/27 и G06F17/28, которые, с нашей точки зрения, характеризуют отношение патентных документов к реализации

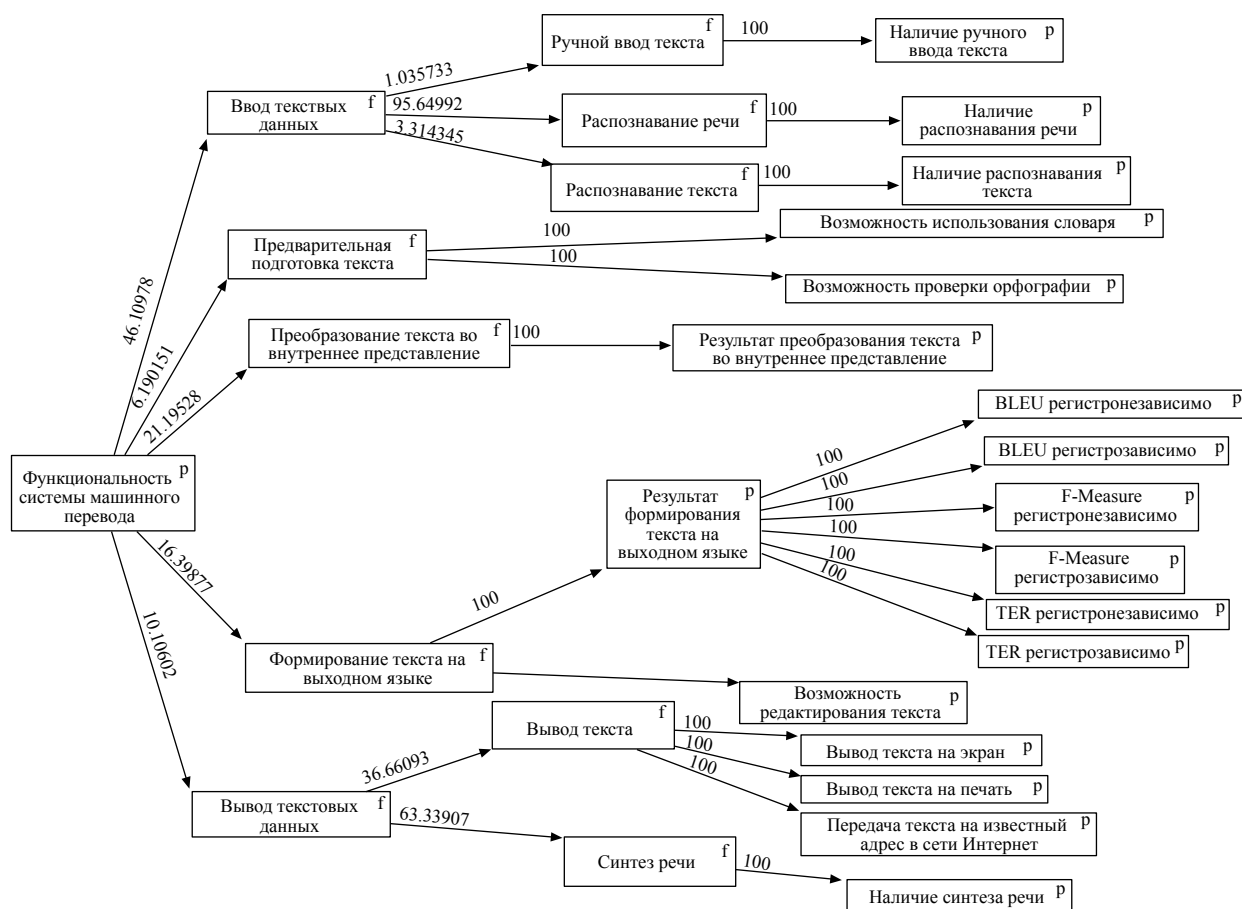


Рис. 3. Оценочная модель для сравнения онлайн-сервисов перевода

наиболее важных функций, необходимых для создания средств МТ. Эти коды на сайте www.eapatris.com расшифрованы в соответствии следующей иерархией:

G06F17/00 – устройства или методы цифровых вычислений или обработки данных, специально предназначенные для специфических функций:

G06F17/20 – манипулирование данными, представленными на естественном языке:

G06F17/27 – автоматический анализ, например, синтаксический разбор, коррекция орфографических ошибок;

G06F17/28 – обработка или перевод текстов на естественном языке.

Заметим, что в расшифровках этих кодов акцент делается на определенных “специфических функциях”, т.е. с точки зрения модели “функции-свойства” речь идет о функциональных элементах. Эти коды позволяют лишь отобрать патентные документы для исследований, однако анализ остальных кодов международной патентной классификации, содержащиеся в уже отобранных документах, позволяет получить более дельную информацию.

В качестве источников данных использована выборка из 12373 патентных документов (обозначим ее как АТ-документы), каждый из которых по международной патентной классификации имеет хотя бы один из кодов G06F17/27 или G06F17/28. На рис. 2 после слов “Важность” представлены значения коэффициентов важности, вычисленные по АТ-документам. В связи с тем, что патентообладатели на одно и то же решение получают патенты в нескольких странах, получается, что выборка включает в себя приблизительно 8.7 тыс. уникальных решений с кодом патентных документов G06F17/27 или G06F17/28 (более точное значение может быть получено, если учесть различные названия одних и тех же организаций и различное написание полных и сокращенных имен одних и тех же физических лиц). На рис. 2 в скобках представлены значения коэффициентов важности, вычисленные по этим уникальным решениям.

Набор остальных кодов из международной патентной классификации, которые указаны в АТ-документах, составляет 173 кода. В большинстве из них в рамках данного исследования не учитывались

подклассы (обозначения после косой черты), либо потому, что они встречаются в патентных документах именно в таком виде, либо потому, что с нашей точки зрения для данного исследования для них не требуется более подробная детализация, и поэтому некоторые коды были сгруппированы.

Изучение текстов, расшифровывающих эти коды, показало, что их подавляющее большинство носит функциональный характер и в сочетаниях с кодом G06F17/27 или G06F17/28 означает различные процессы извлечения, представления и обработки данных, представленных на естественном языке. Например, наибольшее количество АТ-документов содержит код G10L15, расшифровывающийся как “Распознавание речи” (3944 патентных документа); реже встречается код G06F17/30, расшифровывающийся как “Устройства или методы цифровых вычислений или обработки данных, специально предназначенные для специфических функций/информационный поиск; структуры баз данных для этой цели” (3233 патентных документа); еще реже встречается код G06F3, расшифровывающийся как “Вводные устройства для передачи данных, подлежащих преобразованию в форму, пригодную для обработки в вычислительной машине; выводные устройства для передачи данных из устройств обработки в устройства вывода, например интерфейсы” (2649 патентных документов); и т.д.

Для соотнесения кодов международной патентной классификации с функциональными элементами модели нами были разработаны онтология, содержащая понятия, относящиеся к функциональности средств МТ, и учитывающая такие типы отношений, как is-a, a-part-of, а также некоторые атрибутивные связи, и словарь ключевых слов, предназначенный для выявления этих понятий в расшифровках кодов и в текстах патентных документов в контексте, задаваемом кодами G06F17/27 или G06F17/28, означающими, что в патентных документах речь идет о процессах извлечения, представления и обработки данных, представленных на естественном языке. Для разработки онтологии были использованы соответствующие возможности программы Qualimetry Studio [43]. Например, на основе полученной онтологии в сочетании с набором ключевых слов, составленным по приведенным выше расшифровкам кодов международной патентной классификации, код G10L15 может быть напрямую отнесен к функции “Распознавание речи” в модели, показанной на рис. 2, а код G06F17/30 – к функции “Преобразование текста во внутреннее представление” в соответствии со связью: “Преобразование текста во внутреннее представление” a-part-of “Информационный поиск”.

Для разрешения конфликтов, возникающих вследствие сложности формулировок, расшифровывающих ряд кодов международной патентной классификации, выполнялось отнесение каждого такого кода и, как следствие, ссылающихся на него патентных документов либо к нескольким функциональным элементам модели, либо производился анализ встречаемости ключевых слов в формулах изобретений, чтобы выбрать соответствующие приоритетные функциональные элементы. Например, в результате такого анализа документы с кодом G06F15, расшифровывающимся как “Цифровые компьютеры вообще ...; оборудование для обработки данных вообще” (768 патентных документов), относились сразу к трем следующим функциональным элементам: “Предварительная подготовка текста”, “Преобразование текста во внутреннее представление” и “Формирование текста на выходном языке”.

Для расчета коэффициентов важности функциональных элементов, представленных на рис. 2, использована следующая формула: $G_i = 100 \text{Round}(D, 2)$, где

$$D = \sum_{j=1}^{N_i} G_{ij} / \sum_{i=1}^F \sum_{j=1}^{N_i} G_{ij};$$

G_i – коэффициент важности i -го функционального элемента; G_{ij} – количество патентных документов, относящихся к j -й группе документов, сопоставленной с i -м функциональным элементом, и имеющих один и тот же общий код M_{ij} по международной патентной классификации, $\forall M_{ij}: M_{ij} \neq \text{“G06F17/27”} \wedge M_{ij} \neq \text{“G06F17/28”}$; N_i – количество групп патентных документов, относящихся к i -му функциональному элементу; F – количество функциональных элементов, имеющих тот же родительский элемент в декомпозиции, что и i -й функциональный элемент; $\text{Round}(\langle \text{Значение} \rangle, 2)$ – функция округления с точностью до 2-го десятичного знака, использованная в данной работе для наглядности представления полученных результатов.

Примечание: для получения более точных результатов расчета коэффициента важности следует увеличить точность округления в функции Round или полностью исключить функцию Round, используя формулу (3.1); во втором случае сумма значений коэффициентов важности функциональных элементов, представленных на рис. 2, относящихся к одному общему родителю, окажется равной 100:

$$G_i = 100D. \quad (3.1)$$

Для конкретных ситуаций оценивания функциональные элементы, представленные на рис. 2, необходимо дополнить элементами, соответствующими свойствам и группам свойств оцениваемых объектов, чтобы получить модель “функции-свойства”. В качестве примера нами была построена модель для оценивания онлайн-сервисов перевода, представленная на рис. 3, которая может быть использована для сравнения таких сервисов, как Google (translate.google.ru) и Promt (translate.ru). В качестве набора метрик для оценивания результата формирования текста на выходном языке в модели использована совокупность из трех метрик: BLEU, F-Measure и TER с учетом и без учета регистра, применяемая для аналогичных целей в [44].

На связях, направленных на функциональные элементы (обозначенные буквой f), представлены значения коэффициентов важности, вычисленные с использованием формулы (3.1) на основе данных об уникальных решениях, полученных из информации об АТ-документах. На связях, направленных на элементы, обозначающие свойства (отмеченные буквой p), представлены одинаковые значения коэффициентов важности, равные 100.

4. Способы получения численных оценок функциональности средств машинного перевода. Способы получения значений комплексного свойства могут различаться в зависимости от возможностей использования различных типов шкал для представления значений простых свойств, зависящих от конкретных ситуаций оценивания. Ниже приведены примеры формул получения численных оценок (метрик) для комплексного свойства “Функциональность системы машинного перевода”, представленного в оценочной модели на рис. 3.

Приведенные ниже формулы (4.1) и (4.4), обоснованные в работах [13,45], применимы, если значения простых свойств представлены в шкалах отношений или в шкалах интервалов, а элементы модели (функции или свойства), относящиеся к одному родительскому элементу, можно считать независимыми по предпочтению:

$$P = \sum_{i=1}^m K_i G_i, \quad (4.1)$$

где P – значение комплексного оцениваемого свойства “Функциональность системы машинного перевода”; K_i – нормированное (приведенное к шкале от 0 до 1) значение простого свойства; G_i – вычисляемый коэффициент важности i -го простого свойства; m – количество простых свойств.

Для примера применения формул (4.1) и (4.4) в расчете значения комплексного свойства на основе оценочной модели, представленной на рис. 3, рассмотрим сравнение возможностей онлайн-сервисов Google (translate.google.ru) и Promt (translate.ru) при переводе с русского языка на английский. Мы использовали значения простых свойств для оценки результата формирования текста на выходном языке, полученные в [44] по метрикам BLEU, F-Measure и TER на примерах перевода текстов этими сервисами с русского языка на английский. Значения остальных простых свойств определили путем проверки фактического наличия и работоспособности соответствующих функций онлайн-сервисов и представили в шкале {0 (неработоспособно), 1 (работоспособно)}. Полученные значения представлены в табл. 1.

Следует отметить, что, несмотря на наличие у онлайн-сервиса Promt функции распознавания речи, нам не удалось добиться речевого ввода на русском языке. Хотя сервис был настроен на перевод с русского языка на английский, он пытался распознать русскую речь как английскую и подобрать английские слова, похожие на русские.

В табл. 2 отражены коэффициенты важности простых свойств G_i , вычисленные по формуле (4.4), нормированные значения K_i , рассчитанные на основе данных из табл. 1, и значения комплексного оцениваемого свойства “Функциональность системы машинного перевода” P , определенные по формуле (4.4). Вычисленные значения комплексного оцениваемого свойства показывают преимущество онлайн-сервисов Google ($P = 0.9414$) перед Promt ($P = 0.3853$) при переводе с русского языка на английский. Результаты получены с использованием программы Qualimetry Studio [43].

Если значения простых свойств могут быть представлены с помощью нечетких шкал, то для расчета значений комплексных оцениваемых свойств применимы приведенные ниже формулы (4.2)–(4.4), рассмотренные в работе [41], при условии, что элементы модели (функции или свойства), относящиеся к одному родительскому элементу, можно считать независимыми по предпочтению [45]:

$$U(x) = \int x\mu(x) / \int \mu(x) dx, \quad (4.2)$$

Таблица 1. Исходные данные для сравнения возможностей онлайн-сервисов Google и Promt по переводу с русского языка на английский

Возможности и их показатели	Значения	
	Google (translate.google.ru)	Promt (translate.ru)
Ручной ввод текста		
Наличие ручного ввода текста	1	1
Распознавание речи		
Наличие распознавания речи	1	0
Распознавание текста		
Наличие распознавания текста	0	0
Предварительная подготовка текста		
Возможность использования словаря	1	1
Возможность проверки орфографии	0	1
Преобразование текста во внутреннее представление		
Результат преобразования текста во внутреннее представление	1	1
Формирование текста на выходном языке		
Возможность редактирования текста	1	1
Результат формирования текста на выходном языке		
BLEU регистронезависимо	59.59	28.41
BLEU регистрозависимо	61.72	29.83
F-Measure регистронезависимо	83.09	67.62
F-Measure регистрозависимо	84.72	68.89
TER регистронезависимо	67.26	45.77
TER регистрозависимо	68.08	46.42
Вывод текста		
Вывод текста на экран	1	1
Вывод текста на печать	0	1
Передача текста на известный адрес в сети Интернет	1	0
Синтез речи		
Наличие синтеза речи	1	0

где $U(x)$ – значение комплексного оцениваемого свойства “Функциональность системы машинного перевода”;

$$\mu(x) = \max(\mu_{P_1}(x, G_1), \dots, \mu_{P_m}(x, G_m)),$$

где P_1, \dots, P_m – простые свойства; G_1, \dots, G_m – вычисляемые коэффициенты важности свойств; $\mu_{P_1}(x, G_1), \dots, \mu_{P_m}(x, G_m)$ – усеченные функции принадлежности простых свойств; m – количество простых свойств.

Таблица 2. Промежуточные и итоговые результаты расчета значений комплексного показателя

Показатели	Коэффициент важности, G_i	Системы машинного перевода	
		Google (translate. google.ru)	Prompt (translate.ru)
		Нормированное значение, K_i	
Функциональность системы машинного перевода			
Ввод текстовых данных			
Ручной ввод текста			
Наличие ручного ввода текста	0.0048	1	1
Распознавание речи			
Наличие распознавания речи	0.4410	1	0
Распознавание текста			
Наличие распознавания текста	0.0153	0	0
Предварительная подготовка текста			
Возможность использования словаря	0.0310	1	1
Возможность проверки орфографии	0.0310	0	1
Преобразование текста во внутреннее представление			
Результат преобразования текста во внутреннее представление	0.2120	1	1
Формирование текста на выходном языке			
Возможность редактирования текста	0.0820	1	1
Результат формирования текста на вы- ходном языке			
BLEU регистронезависимо	0.0137	1	0
BLEU регистрозависимо	0.0137	1	0
F-Measure регистронезависимо	0.0137	1	0
F-Measure регистрозависимо	0.0137	1	0
TER регистронезависимо	0.0137	1	0
TER регистрозависимо	0.0137	1	0
Вывод текстовых данных			
Вывод текста			
Вывод текста на экран	0.0124	1	1
Вывод текста на печать	0.0124	0	1
Передача текста на известный адрес в сети Интернет	0.0124	1	0
Синтез речи			
Наличие синтеза речи	0.0640	1	0
Результаты, P		0.9414	0.3853

$$\mu_{p_i}(x, G_i) = \begin{cases} G_i, \mu_{p_i}(x) \geq G_i, \\ \mu_{p_i}(x), \mu_{p_i}(x) < G_i, \end{cases} \quad (4.3)$$

где $\mu_{p_i}(x, G_i)$ – усеченная функция принадлежности i -го простого свойства; G_i – вычисляемый коэффициент важности i -го простого свойства; $\mu_{p_i}(x)$ – функция принадлежности i -го простого свойства;

$$G_i = G_d V_i / \sum_{z=1}^m V_{iz}, \quad (4.4)$$

где G_i – вычисляемый коэффициент важности i -го элемента в дереве; V_i – коэффициент важности i -го элемента в дереве; V_{iz} – важность z -го элемента, относящегося к тому же родителю, что и i -й элемент, $z \in \{1, \dots, m\}$; m – количество элементов, относящихся к тому же родителю, что и i -й элемент; G_d – вычисляемый вес родительского элемента для i -го элемента; i – идентификатор элемента, для которого рассчитывается вес; d – номер родительского элемента для i -го элемента; $G_d = 1$ для самого оцениваемого комплексного свойства “Функциональность системы машинного перевода”.

Заключение. Представлены основные понятия, модели и методы оценивания качества программного обеспечения. В связи с противоречивостью определений понятия “функциональность” в существующих стандартах в области оценивания качества программного обеспечения в работе это понятие определено как частный случай понятия качество в квалиметрии применительно к оцениванию качества программного обеспечения. Предложена обобщенная модель оценивания функциональности программного обеспечения. Представлена разработанная на ее основе модель оценивания функциональности систем машинного перевода, которая может служить прототипом для построения более детализированных моделей. Рассмотрены некоторые способы получения численных оценок функциональности систем машинного перевода с помощью таких моделей.

СПИСОК ЛИТЕРАТУРЫ

1. *Radiant Insights, Inc.* Machine Translation (MT) Market Analysis by Application (Automotive, Military & Defense, Electronics, IT, Healthcare), by Technology (Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT)) and Segment Forecasts To 2022. Research Report. Grand View Research, Inc., USA, CA San Francisco, 2015. URL: <http://www.grandviewresearch.com/industry-analysis/machine-translation-market>.
2. *Поддубицкий А.Ф.* Поддержка и управление процессами локализации // Компьютерные системы и сети. Материалы 49-й научной конф. аспирантов, магистрантов и студентов. Минск: БГУИР, 2013. С. 54.
3. *Голов А.Н.* Информационные технологии как инструмент перевода иноязычного технического текста // Тр. XI Междунар. научн.-практ. конф. “Наука и современность – 2011”. Новосибирск, 2011. С. 244–249.
4. *Кулямин В.В., Петренко О.Л.* Место тестирования среди методов оценки качества ПО // Тр. ИСП РАН, 2003, Т. 4. С. 163–176.
5. *Grady R.B.* Practical Software Metrics for Project Management and Process Improvement. USA, N.J., Englewood Cliffs: Prentice-Hall, 1992.
6. *Basili V.R.* Software Modeling and Measurement: The Goal/Question/Metrics Paradigm. Technical Report, CS-TR-2956, Department of Computer Science. University of Maryland, College Park, MD20742, September, 1992.
7. *Dromey R.G.* Cornering the Chimera // IEEE Software. 1996. V. 20. P. 33–43.
8. *Boegh J., Depanfilis S., Kitchenham B., Pasquini A.* A Method for Software Quality Planning, Control, and Evaluation // IEEE Software. 1999. V. 23. P. 69–77.
9. *Саати Т.* Принятие решений. Метод анализа иерархий. М.: Радио и связь, 1993.
10. *Смирнов В.В.* Квалиметрия в модели обучаемого в электронном обучении // Сб. научн. тр. XIX научн.-практ. конф. “Инжиниринг предприятий и управление знаниями (ИП&УЗ–2016)”. М.: ФГБОУ ВО “РЭУ им. Г.В. Плеханова”, 2016. С. 318–325.
11. *Ларичев О.И.* Теория и методы принятия решений, а также Хроника событий в Волшебных Странах: Учебник. М.: Логос, 2000.
12. *Субетто А.И.* Квалиметрия: малая энциклопедия. Вып. 1. СПб: ИПЦ СЗИУ – фил. РАНХиГС, 2015.
13. *Азгальдов Г.Г.* Теория и практика оценки качества товаров (основы квалиметрии). М.: Экономика, 1982.

14. *Мишин В.М.* Управление качеством: Учебник для студентов вузов, обучающихся по специальности “Менеджмент организации” (061100): 2-е изд. перераб. и доп. М.: ЮНИТИ-ДАНА, 2005.
15. *Ефимов В.В.* Средства и методы управления качеством: учебное пособие. М.: КНОРУС, 2007.
16. *Chinmaya Dash C., Ramesh D., Mr. Behera P.C.* Software Quality Management: A Process to Improve the Quality of Software Project Using SQA // Proc. 4th Annual Conf. on Engineering and Information Technology (ACEIT 2016). Japan, Kyoto, 2016. P. 69–73.
17. *Divya C.D., Bharath T.S.* A Glimpse of Software Quality Assurance and Factors Affecting It // Intern. J. Computer Science Trends and Technology (IJCST). 2016. V. 4 (1). P. 38–43.
18. *Sharma N.* Cycle of Software Quality // Intern. J. Innovative Research in Engineering & Science. 2014. V.3 (3). P. 26–38.
19. *Shaout A., Dusute C.* Verification and Validation of Embedded Systems – The Good, the Bad and the Ordinary // Intern. J. Advanced Research in Computer Science and Software Engineering. 2014. V. 4 (10). P. 8–15.
20. *Srinivasan K.P., Devi T.* Software Metrics Validation Methodologies in Software Engineering // Intern. J. Software Engineering & Applications (IJSEA). 2014. V.5. № 6. P. 87–102.
21. *Андерсон Р.* Доказательство правильности программ. М.: Мир, 1982.
22. *Apt K.R., Olderog E.R.* Introduction to Program Verification. Report CS-R9036. Center for Mathematics and Computer Science. Department of Software Technology. Amsterdam, 1990.
23. *Абрамова Н.А., Баталина Т.С., Гегемов Н.А., Коврига С.В.* Новый математический аппарат для анализа внешнего поведения и верификации программ. М.: Институт проблем управления РАН, 1998.
24. *Соколова Е.В.* Экспериментальная оценка эффективности применения облачных вычислений в бизнес-приложениях операционной системы Android // Радиоелектронні і комп’ютерні системи. 2015. № 4. С. 46–50.
25. *Рахманов В.Ю., Тимченко Б.Д.* Метрики в современном тестировании // Научно-технический вестник. СПб: СПбГИТМО (ТУ). Вып. 10. Информация и управление в технических системах. СПбГУ ИТМО, 2003. С. 108–111.
26. *Vesra A., Rahul H.* A Study of Various Static and Dynamic Metrics for Open Source Software // Intern. J. Computer Applications. 2015. V. 122. № 10. P. 17–21.
27. *Jones C.* Software Quality Metrics: Three Harmful Metrics and Two Helpful Metrics. Technical Report. 2012. URL: <http://www.ppi-int.com/systems-engineering/free%20resources/Software%20Quality%20Metrics%20Capers%20Jones%20120607.pdf>.
28. *Adri T.* Function Point Counting Practices Manual. Release 4.3.1. International Function Point Users Group (IFPUG). USA, N.J., 2010.
29. *Bagci V.H., Turgut U.O., Ciltik A., Cetin S., Ozcelik R.* Measuring a Software Production Line with IFPUG-based Function Points // Proc. Ninth International Conf. on Software Engineering Advances (ICSEA 2014). France, Nice, 2014. P. 202–208.
30. *Papineni K., Roukos S., Ward T., Zhu W.* BLEU: a Method for Automatic Evaluation of Machine Translation // Proc. 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadelphia, 2002. P. 311–318.
31. *Popović M., Ney H.* Syntax-Oriented Evaluation Measures for Machine Translation Output // Proc. 4th EACL Workshop on Statistical Machine Translation. Greece, Athens, 2009. P. 29–32.
32. *Popović M.* Morphemes and POS Tags for N-gram Based Evaluation Metrics // Proc. Sixth Workshop on Statistical Machine Translation (EMNLP 2011). UK, Edinburgh, 2011. P. 104–107.
33. *Zhang Y., Vogel S., Waibel A.* 2004. Interpreting BLEU/NIST Scores: How Much Improvement do We Need to Have a Better System? // Proc. Fourth Intern. Conf. on Language Resources and Evaluation (LREC-2004). Lisbon, Portugal, 2004. P. 2051–2054.
34. *Han A.L.F., Wong D.F., Chao L.S.* LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors // Proc. 24th Intern. Conf. on Computational Linguistics (COLING 2012). India, Mumbai, 2012. P. 441–450.
35. *Melamed I.D., Green R., Turian J.P.* Precision and Recall of Machine Translation // Proc. HLT-03, Edmonton, Canada, 2003. P. 61–63.
36. *Banerjee S., Lavie A.* METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments // Proc. ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. USA, Michigan Ann Arbor, 2005. P. 65–72.
37. *Cer D., Manning C.D., Jurafsky D.* The Best Lexical Metric for Phrase-Based Statistical MT System Optimization // Proc. 11th Annual Conf. of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies (NAACL HLT 2010). USA, CA, Los Angeles, 2010. P. 555–563.
38. *Paul M., Finch A., Sumita E.* Predicting Human Assessment of Machine Translation Quality by Combining Automatic Evaluation Metrics using Binary Classifiers // Intern. J. Computer Applications. 2012. V. 59. № 10. P. 1–7.

39. *Denkowski M., Lavie A.* Challenges in Predicting Machine Translation Utility for Human Post-Editors // Proc. Tenth Conf. of the Association for Machine Translation in the Americas. USA, CA, San Diego, 2012. P. 40–49.
40. *Sokolov A., Riezler S., Urvoy T.* Bandit Structured Prediction for Learning from Partial Feedback in Statistical Machine Translation // Proc. MT Summit XV, 2015. USA, FL, Miami, 2015.
41. *Смирнов В.В.* Квалиметрическая функция полезности // Тр. VIII-й Междунар. научн.-техн. конф. “Интегрированные модели и мягкие вычисления в искусственном интеллекте”. В 2-х т. Т. 2. М.: Физматлит, 2015. С. 692–701.
42. *Батурин Н.А.* Психология оценки: общие представления, дифференциация понятий и области изучения // Вестн. ЮУрГУ. Сер. Психология. 2008. Вып. 1. № 31. С. 17–30.
43. *Смирнов В.В.* Qualimetry Studio: программное средство построения квалиметрических моделей // Тр. 15-й нац. конф. по искусственному интеллекту с международным участием КИИ-2016. В 3-х т. Т. 3. Смоленск: Универсум, 2016. С. 279–286.
44. *Ulitkin I.* Human Translation vs. Machine Translation: Rise of the Machines // Translation J. 201. № 17(1). URL: <http://www.bokorlang.com/journal/63mtquality.htm>.
45. *Азгальдов Г.Г., Костин А.В., Садовов В.В.* Квалиметрия для всех: Учеб. пособие. М.: ИД ИнформЗнание, 2012.