

ЛИНГВИСТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ЭЛЕКТРОННОГО СЛОВАРЯ TECHLEX

¹Швед М.О., ²Попова Ю.Б., ³Макарич М.В.

¹БНТУ, г. Минск, Беларусь, Mischa.schwed@gmail.com

²БНТУ, г. Минск, Беларусь, julia_popova@mail.ru

³БНТУ, г. Минск, Беларусь, 2348843@tut.by

Введение. Настоящая работа посвящена исследованию процессов становления и развития терминологии информатики и вычислительной техники в английском, белорусском и русском языках. Это довольно молодая отрасль знания. За короткий срок вычислительная техника проделала такой путь в своем техническом совершенствовании, масштабах применения и влиянии на человеческую жизнь, что оправданным стало определение компьютера как одного из важнейших изобретений человека. Проникновение вычислительной техники в повседневную жизнь миллионов людей приводит к довольно широкому распространению лексики по данной предметной области и вхождению ее в устную речь многих людей. В этом контексте изучение лексики информатики и вычислительной техники приобретает особую значимость.

Возникнув на стыке многих наук, новая отрасль и ее подъязык складывались на основе этих наук и их подъязыков. По мере развития формировались и особые черты, характерные только для подъязыка информатики и вычислительной техники. Лексика этой предметной области представляет собой особый интерес в связи с тем, что в ней ярко проявляется взаимодействие факторов экстралингвистического и лингвистического характера.

На современное состояние специальной лексики в области информатики и вычислительной техники влияет несколько факторов. Большинство новинок вычислительной техники приходит из Западной Европы и США, вместе с тем появляются также новые термины. Вычислительная техника развивается с поражающей быстротой, оборудование и программное обеспечение быстро устаревают вместе с обозначающими их словами. В свою очередь, появляется новое оборудование и программы, которые требуют наименований. Различные фирмы вводят свои термины для обозначения одного и того же понятия, что ведет к появлению синонимов, а переводчики технической литературы также подбирают иногда разные слова для обозначения одного и того же.

Вышеупомянутые причины свидетельствуют об отсутствии стабильности в терминологии информатики и вычислительной техники и дают основание утверждать, что она находится в процессе формирования при сильном влиянии терминологии этой предметной области в английском языке.

Разработка лингвистической базы данных для словаря TechLex. В процессе создания лингвистической базы (ЛБ) данных для предлагаемого электронного словаря была проведена предварительная работа по отбору и описанию терминологии подъязыка информатики и вычислительной техники на английском, белорусском и русском языках, осуществлена гармонизация этих описаний и согласование терминологических систем вышеупомянутых языков [1].

На первом этапе нами были отобраны и проанализированы 30 научно-технических англоязычных статей, взятых из периодических изданий «International Journal of Innovative Research in Computer and Communication Engineering», «Journal of Computer Engineering & Information Technology», «Journal of Applied & Computational Mathematics», «American Journal of Computer Science and Information Technology», «Journal of Informatics and Data Mining». В результате для включения в ЛБ словаря TechLex отобраны 7000 терминов. В общее количество терминов нами были включены как однокомпонентные: *frame* (англ.) – *фрэйм* (бел.) – *фрейм* (рус.); *storage* (англ.) – *захоўванне* (бел.) – *хранение* (рус.); *subset* (англ.) – *падмноства* (бел.) –

подмножество (рус.) так и многокомпонентные термины: *access control* (англ.) – кіраванне доступам (бел.) – управление доступом (рус.); *information query language* (англ.) – інфармацыйна-пошукавая мова (бел.) – информационно-поисковый язык (рус.).

Так как значительную часть подъязыка информатики и вычислительной техники составляют аббревиатуры, они также рассматривались для внесения в лингвистическую базу словаря: *AIS/Alarm Indication Signal* (англ.) – СІАС/сiгнал індыкацыі аварыйнага стану (бел.) – СИАС/сигнал индикации аварийного состояния (рус.); *BIOS/Basic Input-Output System* (англ.) – базавая сістэма ўводу-вываду (бел.) – базовая система ввода-вывода (рус.); *GUI/Graphical User Interface* (англ.) – графічны інтэрфейс карыстальніка (бел.) – графический интерфейс пользователя (рус.); *TB/terabyte* (англ.) – тэрабайт (бел.) – терабайт (рус.). Общее количество аббревиатур составило 8% от общего количества отобранной лексики. Это обусловлено наличием большого числа многосоставных технических терминов, отражающих инновационные разработки в информационной сфере.

Хотя основной особенностью языка научно-технической литературы является большое количество специальных терминов, он включает и некоторое количество общеупотребительных слов и сочетаний, которые также были включены в ЛБ нашего словаря. Значительную часть общеупотребительных слов составляют многозначные слова. В ряде случаев для определения значения многозначного слова приходится учитывать не только лексические связи, но и грамматические признаки. Такие лексические единицы представляют собой определенную трудность для перевода. Например, *bit* (англ.), в общеупотребительном смысле имеющее значение *невялікая колькасць* (бел.) – небольшое количество (рус.), в подъязыке информатики и вычислительной техники переводится как *адзінка вымярэння інфармацыі ў дваіковай сістэме вылічэння* (бел.) – единица измерения информации в двоичной системе исчисления (рус.).

Второй этап включал работу по нормализации пар терминов относительно языка перевода. Так как термины в текстах точных наук имеют строго определенное значение и точно выражают понятия, процессы и названия вещей, нами были использованы все три способа прямого перевода терминов: заимствование, калькирование и дословный перевод.

Самым простым способом прямого перевода является заимствование, которое позволяет заполнить пробел, обычно металингвистического характера (новая техника, неизвестные понятия). Заимствования составляют особый пласт лексики с точки зрения процессов номинации и представляют собой определённую экономию языковых усилий, используя готовые единицы чужого языка, чаще всего английского. Это обусловлено тем фактом, что авторами большинства современных научно-технических разработок последнего столетия являются англоговорящие исследователи. Например: *interface* (англ.) – інтэрфейс (бел.) – интерфейс (рус.), *display* (англ.) – дысплэй (бел.) – дисплей (рус.), *diode* (англ.) – дыёд (бел.) – диод (рус.), *kilobyte* (англ.) – кілабайт (бел.) – килобайт (рус.), *calculator* (англ.) – калькулятар (бел.) – калькулятор (рус.), *browser* (англ.) – браўзар (бел.) – браузер (рус.), *domain* (англ.) – дамен (бел.) – домен (рус.).

Калькирование – второй способ прямого перевода. Он является заимствованием особого рода: мы заимствуем из иностранного языка ту или иную синтагму и буквально переводим элементы, которые ее составляют. Мы получаем, таким образом, либо калькирование выражения, причем используем синтаксические структуры языка перевода, привнося в него новые экспрессивные элементы, либо калькирование структуры, причем привносим в язык новые конструкции. Например, *superpower* (англ.) – звышмагутны (бел.) – сверхмощный (рус.), *floating-point* (англ.) – плавальная кропка (бел.) – плавающая точка (рус.), *hash table* (англ.) – хэш-табліца (бел.) – хеш-таблица (рус.), *graphic display* (англ.) – графічны дысплэй (бел.) – графический дисплей (рус.), *simulation* (англ.) – мадэляванне (бел.) – моделирование (рус.), *ribbon cable* (англ.) – стужачны кабель (бел.) – ленточный кабель (рус.).

Дословный перевод, также использованный нами в данном исследовании, обозначает переход от исходного языка к языку перевода, который приводит к созданию правильного и идиоматического переводного эквивалента с соблюдением обязательных норм языка: *infrared keyboard* (англ.) – *інфрачырвоная клавіятура* (бел.) – *инфракрасная клавиатура* (рус.), *indicative data* (англ.) – *характарыстычныя дадзеныя* (бел.) – *характеристические данные* (рус.), *inference chain* (англ.) – *ланцужок высновы* (бел.) – *цепочка вывода* (рус.), *graphic display* (англ.) – *графічны дысплэй* (бел.) – *графический дисплей* (рус.), *home record* (англ.) – *пачатковы запіс* (бел.) – *начальная запись* (рус.), *n-color printing* (англ.) – *n-каляровы друк* (бел.) – *n-цветная печать* (рус.).

Анализ способов перевода терминов показал, что наиболее употребительным для английской терминологии подязыка информатики и вычислительной техники являются заимствование и прямой дословный перевод вследствие того, что у многих английских терминов имеются лексические эквиваленты в белорусском и русском языках, общепринятые и зафиксированные в словарях.

Результаты нашего терминологического исследования были размещены в Excel таблице для последующего заполнения базы данных технического англо-белорусско-русского электронного словаря TechLex. Фрагмент таблицы приводится на рисунке 1.

	A	B	C	D	E	F	G
5742	join	1.злучанне (операцыя рэляцыйнай алгебры); 2. (1.соединение (операция реляционной алгебры); 2. (операция включающее ИЛИ					
5743	journal file	часопіс, часопісны файл	журнал, журнальны файл				
5744	journalizing	журналяванне (запіс інфармацыі аб аперациях ў журналізацыя (запись информации об операциях в журнал)					
5745	joystick	каардынатная ручка, "джойстык"	координатная ручка, "джойстик"				
5746	jump I	пераход, перадача кіравання	переход, передача управления				
5747	jump II	пераходзіць, выконваць пераход, перадаваць	переходить, выполнять переход, передавать управление				
5748	jump instruction	каманда пераходу, каманда перадачы кіравання	команда перехода, команда передачи управления				
5749	jump table	табліца пераходаў, пераключальнік	таблица переходов, переключатель				
5750	justification	выраўноўванне [выключка] радкоў	выравнивание [выключка] строк				
5751	justify	выраўноўваць	выравнивать				
5752	Karnaugh map	карта Карно	карта Карно				
5753	Kbit	кілабіт, Кбіт (1024 біта)	килобит, Кбит (1024 бита)				
5754	Kbyte	кілабайт (1024 байты)	килобайт (1024 байта)				
5755	kernel	ядро (аперацыйнай сістэмы)	ядро (операционной системы)				
5756	kernel mode	прывілежаваны рэжым, рэжым ядра (аперацыйнай сістэмы)	привилегированный режим, режим ядра (операционной системы)				
5757	kernel operation	аперацыя ядра	операция ядра				
5758	key	ключ; клавіша (клавіятуры)	ключ; клавиша (клавиатуры)				
5759	key click	пстрычка пры націску клавішы	щелчок при нажатии клавиши				
5760	key field	поле ключа; ключавое поле	поле ключа, ключевое поле				
5761	key in	друкаваць	печатать				
5762	key sorting	сартаванне па ключы	сортировка по ключу				
5763	keyboard input	увод (дадзеных) з клавіятуры; дадзеныя, якія ўводзіцца	ввод (данных) с клавиатуры; данные, введенные с клавиатуры				
5764	keyboard	клавіятура	клавиатура				
5765	keyed access	доступ па ключы, ключавы доступ	доступ по ключу, ключевой доступ				
5766	keyed sequential-access method	паслядоўны метада доступу з ключамі	последовательный метод доступа с ключами				
5767	keypad	дапаможная [спецыялізаваная] клавіятура	вспомогательная [специализированная] клавиатура				
5768	keystroke	націск клавішы	нажатие клавиши				
5769	keyword parameter	ключавы параметр	ключевой параметр				
5770	keyword	ключавое слова	ключевое слово				
5771	kill	знішчаць (пра працэс); выдаляць (пра паведамленне)	уничтожать (о процессе); удалять (о сообщении или части текста)				
5772	kit	набор; камплект	набор; комплект				
5773	kludge	варыянты запіс (якая выкарыстоўваецца для апраўдання)	вариантная запись (используемая для обхода системы контроля типов)				
5774	knowledge acquisition	збор ведаў; пабудова базы ведаў	сбор знаний; построение базы знаний				
5775	knowledge base	база ведаў	база знаний				

Рисунок 1 – Фрагмент Excel таблицы с англо-белорусско-русским переводом

Описание программной реализации электронного словаря. В процессе реализации англо-белорусско-русского технического словаря было разработано клиент-серверное приложение на языке программирования Java с использованием архитектурного паттерна MVC (англ., Model-View-Controller), разделяющего данные приложения, пользовательского интерфейса и управляющей логики на три отдельных компонента: модель, представление и контроллер – таким образом, что модификация каждого компонента может осуществляться независимо.

Модель – это правила, которые используются для работы с данными, представляющими концепцию управления приложением. В любом приложении вся структура моделируется как данные, которые обрабатываются определённым образом в соответствии с правилами, например, адрес электронной почты должен быть задан в определённом формате. Модель даёт контроллеру представление данных, которые запросил пользователь (сообщение, страницу книги, фотоальбом и тому подобное). Модель данных будет одинаковой, вне зависимости от того, как мы хотим представлять их пользователю. Модель содержит наиболее важную часть логики нашего приложения, логики, которая решает задачу, с которой мы имеем дело (например, магазин, банк, словарь и т.д.).

Вид обеспечивает различные способы представления данных, которые получены из модели. Он может быть шаблоном, который заполняется данными.

Контроллер управляет запросами пользователя. Его основная функция – вызывать и координировать действия необходимых ресурсов и объектов, нужных для выполнения операций, задаваемых пользователем. Обычно контроллер вызывает соответствующую модель для задачи и выбирает вид, который подходит наилучшим образом для текущей ситуации. Веб-приложение обычно состоит из набора контроллеров, моделей и видов. Контроллер может быть устроен как основной, который получает все запросы и вызывает другие контроллеры для выполнения действий в зависимости от ситуации.

К основным преимуществам технологии MVC относят следующие:

- удобство вывода различного представления данных для всех типов устройств;
- облегченная поддержка и тестирование программного кода.

Клиент-серверная архитектура предлагаемого электронного словаря разделяет функции представления, обработки и хранения данных. Серверная часть разработанной системы содержит базу данных (БД) MySQL DataBase (DB) и сервер JAVA EE (англ., Enterprise Edition). Клиентская часть системы содержит базу данных SQLite DB и мобильное приложение для операционной системы Android. Для создания и оформления мобильного приложения использовались стандартные инструменты среды разработки Android Studio.

Для перспективного развития электронного словаря TechLex в базе данных, кроме слов по информационным технологиям, предусмотрены таблицы для следующих предметных областей: архитектура и строительство, водоснабжение, педагогика, транспортные коммуникации, экономика и энергетика. Каждая такая таблица имеет суррогатный первичный ключ в виде колонки id. Связи между таблицами отсутствуют, поскольку в этом нет необходимости. Такая структура позволяет легко обновлять БД в мобильном приложении, т.к. при добавлении новых слов на сервер будет обновляться лишь актуальная таблица без нерационального скачивания всех предметных областей словаря.

Веб-приложение электронного словаря TechLex развернуто в локальной сети Белорусского национального технического университета по адресу [<http://172.16.11.72:4325>], который необходимо набрать в адресной строке браузера с любого компьютера университета. После этого отобразится главное окно словаря, приведенное на рисунке 2.

Интерфейс приложения выполнен на белорусском языке, по умолчанию установлено направление перевода «Англійская мова» → «Беларуская мова». Предметная область «Інфармацыйныя тэхналогіі» заполнена лингвистическими терминами, описанными выше. Также предусмотрены вкладки для других предметных областей: «Водазабеспячэнне», «Будаўніцтва і

Архітэктэра», «Транспартныя камунікацыі», «Эканоміка», «Энергетычнае будаўніцтва», «Педагагіка».

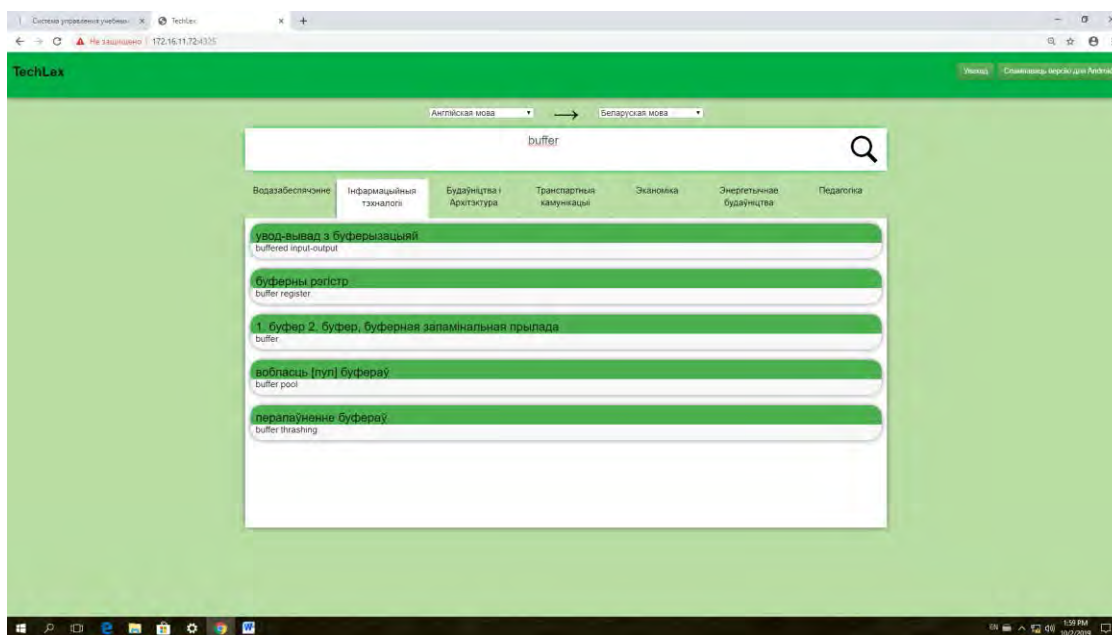


Рисунок 2 – Поиск перевода в словаре TechLex

Для перевода слов необходимо начать его ввод в поле поиска, при этом в строках для переведенных значений будут отображаться все слова и словосочетания, содержащие вводимую часть. Так на рисунке 2 приведена копия экрана электронного словаря TechLex для перевода в направлении для перевода в направлении «Английская мова» → «Беларуская мова» для слова «buffer». Поскольку в базе слов имеются словосочетания, содержащие в своей части слово «buffer», они будут найдены тоже, например, «buffered input-output».

Для использования мобильного приложения на смартфоне или планшете под управлением операционной системы Android необходимо подключиться к локальной сети Белорусского национального технического университета через Wi-Fi точки доступа, которые имеются во всех учебных корпусах университета, а также в библиотеке. Затем в строке браузера требуется ввести адрес электронного словаря [<http://172.16.11.72:4325>] и после появления главного окна словаря (рисунок 2) нажать на кнопку «Спампаваць версію для Android», расположенную в правом верхнем углу. После этих действий начнется процесс скачивания мобильной версии словаря, при котором возможен запрос операционной системы о подтверждении действий. В случае согласия файл с названием TechLex.apk скачается на мобильный телефон или планшет. После этого необходимо кликнуть по нему и подождать, пока пройдет процесс установки. После этого в меню телефона появится иконка приложения, на которую необходимо нажать для его запуска. При первом запуске необходимо быть подключенным к локальной сети БНТУ, чтобы приложение могло скачать базу слов. В дальнейшем словарь можно использовать в офлайн-режиме.

После окончания скачивания актуальной базы слов откроется основное окно приложения, изображенное на рисунке 3. По умолчанию установлено направление перевода «Английская мова» → «Беларуская мова» и предметная область «Інфармацыйныя тэхналогіі». Для перевода требуемого слова необходимо его ввести в поле для ввода и нажать на кнопку поиска. Переведенное слово отобразится в нижней части экрана мобильного приложения. Если вариантов перевода будет несколько, то отобразится весь список (рисунок 3 справа). Также как и в веб-

приложении, здесь возможны следующие направления переводов: «Английская мова» → «Беларуская мова», «Английская мова» → «Русская мова», «Беларуская мова» → «Английская мова», «Русская мова» → «Английская мова», «Беларуская мова» → «Русская мова», «Русская мова» → «Беларуская мова» [2-3].

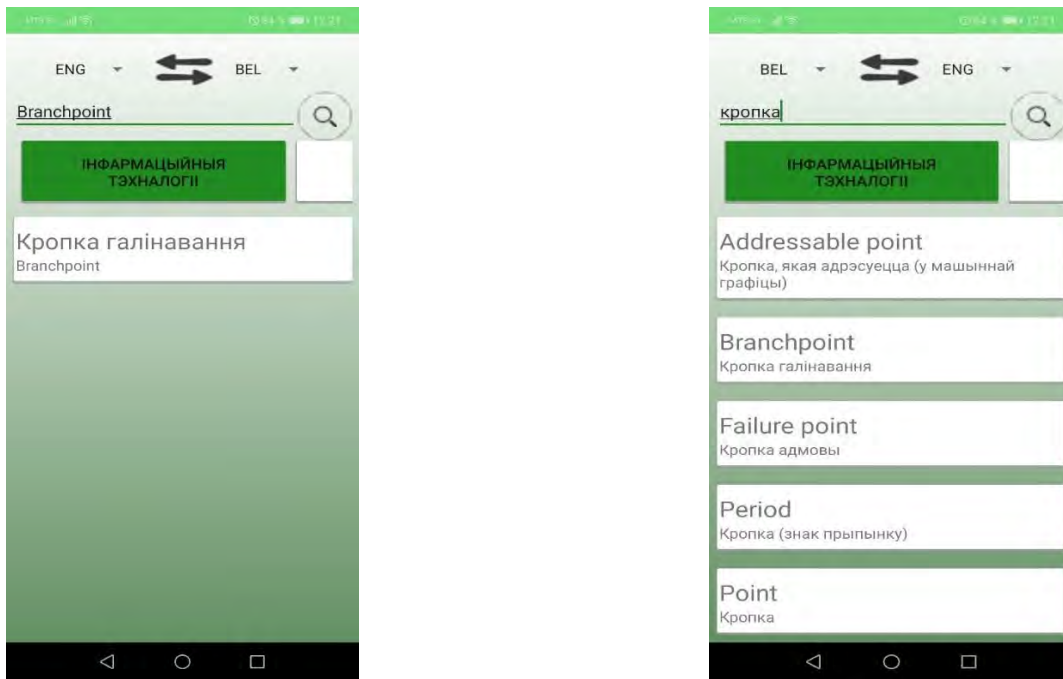


Рисунок 3 – Основное окно мобильного приложения TechLex

Заклучение. Лингвистическое и программное обеспечение электронного англо-белорусско-русского словаря TechLex спроектировано с учетом анализа современных электронных многоязычных переводных словарей и имеет ряд преимуществ:

- это первый технический многоязычный электронный словарь, имеющий англо-белорусско-русскую версию;
- лингвистическая база данных словаря составлена не традиционным способом переработки большого количества бумажных словарей и объединения полученных переводов, а путем последовательной обработки научно-технических англоязычных периодических изданий отдельных предметных областей;
- интерфейс электронного словаря разработан таким образом, что в соответствии с запросом активизируется лишь отдельно взятая зона, так как существует возможность добавления в базу данных дополнительных предметных областей и соответствующего им лексического наполнения.

Список использованных источников

1. Макарич, М.В. Лингвистическое и программное обеспечение англо-белорусско-русского словаря технических терминов / М.В. Макарич, Ю.Б. Попова, М.О. Швед // Системный анализ и прикладная информатика. – 2018. – №4. – С. 74–82.
2. Popova, Y.V. Mobile application for English-Belarusian-Russian technical dictionary / Y.V. Popova, M.V. Makarych, M.O. Shved // Тэарэтычныя і прыкладныя аспекты этналагічных даследаванняў: зборнік навуковых артыкулаў / пад навук. рэд.: Н.П. Мартысюк – Мінск : БНТУ, 2019. – С. 426-433.

3. Швед, М.О. Мобильное приложение для англо-белорусско-русского словаря технических терминов / М.О. Швед, Ю.Б. Попова // Информационные технологии и системы: проблемы, методы, решения (ИТС – 2018): сб. материалов Республиканской научно-технической конференции, Минск 22–23 ноября 2018 г. / редкол.: С.В. Харитончик [и др.]. – Минск: Четыре четверти, 2019. – С. 207–210.