

Уильям У. МАКЛЕНДОН-младший

Бостон, США

Директор Технической службы информационных технологий SIL International

Магистр наук по специальности «Системное проектирование» (Бостонский университет), бакалавр наук (Университет Дьюка)

Технологии SIL в области многоязычия в киберпространстве

Вступление

SIL International, конфессиональная некоммерческая организация, деятельность которой направлена на выявление и документирование всех языков мира, а также содействие признанию их ценности, проявляет особое внимание к языкам этнокультурных меньшинств. Со дня своего основания в 1934 г. SIL работал с более чем 1800 языковыми общинами, представляющими более 1,2 млрд человек из 70 стран. Занимаясь вопросами лингвистики, грамотности, перевода и другими видами деятельности в области развития, связанной с языками, сотрудники SIL воспринимают свою работу как следствие приверженности христианским ценностям, которую они воплощают в служение, достижение научного совершенства, обмен знаниями и создание партнерских отношений. SIL стремится служить всем, независимо от верований, идеологии, пола, расы и этнической принадлежности.

Помимо работы, направленной на развитие отдельных языков, SIL также активно поддерживает языки меньшинств на местном, национальном и международном уровнях. Организация обладает статусом консультанта ЭКОСОС (Экономического и Социального Совета ООН) и ЮНЕСКО. С помощью этих и других своих партнеров SIL трудится во имя поддержки многоязычного образования, грамотности и профессиональной подготовки, ведет социально-ориентированную деятельность в сфере развития, а также преследует другие гуманитарные цели. Более подробная информация предлагается на сайте <http://www.sil.org/sil/>.

Одной из основных сфер работы SIL является развитие и координация технологий и технических ресурсов, обеспечивающих представленность и использование всех языков в быстро меняющемся мире информационных технологий и киберпространства.

Целью данного доклада является освещение многих из этих технологий и связанной с ними деятельности. Автор отмечает, что его доклад представляет деятельность многочисленных коллег SIL во всем мире¹⁶.

Основная деятельность

Для обеспечения возможности использования отдельного языка при работе на компьютере и представления его в киберпространстве следует решить три основные проблемы – определить методы ввода, кодировки и визуализации.

Методы ввода определяют процедуру введения в компьютер данных на каком-либо языке. Как правило, для набора текста используется клавиатура, где каждому знаку данного языка соответствует определенная клавиша или комбинация клавиш. Стандартные компьютеры разрабатывались для английского языка и содержат 70–90 клавиш, тогда как другие языки могут иметь меньшее или большее количество знаков.

Методы кодировки управляют преобразованием отдельного символа в конкретное численное значение, которое используется компьютером при обработке внутренних данных. Для максимально широкого употребления необходима стандартизация методов кодирования. В противном случае числовые коды одного компьютера теряют всякий смысл для другого, использующего иную кодировку. Мировой стандарт языкового кодирования задает Unicode, и SIL продолжает активно участвовать в работе его технических комитетов.

Наконец, методы визуализации предоставляют пользователю языковые материалы в письменной форме на экране или в виде распечатки. Визуализация первоначально предназначалась для английского и других языков, использующих латинское письмо, однако многие языки используют более сложные системы письменности, и их использование на компьютере и в киберпространстве невозможно без специальных поддерживающих технологий. Технология изготовления, поддержки и сопровождения шрифтов является одним из основных элементов визуализации. В шрифте каждому символу соответствует глиф, особая форма его представления. В некоторых языках для одного и того же символа требуются различные глифы в зависимости от контекста. Для поддержки таких языков необходимо использовать технологию «smart font».

Еще одним важным элементом представления языка в киберпространстве являются метаданные, описывающие сам язык. Название языка и стандарт-

¹⁶ Преобладающая часть материала взята с сайтов SIL и из других опубликованных трудов. Сноскам на каждый источник автор предпочел ссылки на интернет-адреса оригиналов использованных сообщений, которые также содержат дополнительную тематическую информацию.

ный код его представления насуточно необходимы, чтобы гарантировать правильное использование методов ввода, кодировки и визуализации всеми пользователями языковых данных.

Как и прежде, SIL активно участвует во всех видах этой деятельности. Статус некоммерческой организации позволяет широко распространять наши достижения с помощью свободного программного обеспечения или других общедоступных ресурсов. Далее приводятся соответствующие примеры для каждой из этих областей.

Методы ввода

Партнерство с Tavultesoft Pty, Ltd. (<http://www.tavultesoft.com/>) помогло SIL разработать технологии и методы ввода данных с клавиатуры для всех языков, использующих программное обеспечение Keyman – приложение Microsoft Windows, позволяющее осуществлять ввод данных на основе формата Unicode при использовании других приложений Windows. Мы подготовили ряд учебных материалов по дизайну клавиатуры, опубликовали рабочую документацию и другую информацию по разработке методов ввода и выпустили программные клавиатуры для многих языков. Кроме того, нами разработаны редакторы раскладок клавиатуры: Ukelele системы Unicode для Mac OS X и Keyman для Linux (KMFL, <http://kmfl.sourceforge.net/>), поддерживающий раскладку Keyman для Windows. Дополнительную информацию о методах ввода и ссылках для загрузки этих ресурсов можно получить на сайте <http://scripts.sil.org/Input>.

Кодирование

Первоначально SIL были разработаны индивидуальные методы кодирования для языков тех стран, где работал наш институт. Эти методы не являлись частью какого-либо широкого стандарта и предназначались для использования нашими сотрудниками и определенными языковыми сообществами. После появления Unicode SIL принял его стандарт и стремился поддерживать все языковые сообщества и их системы письма в этом международном стандарте, «обеспечивающем любому знаку свой цифровой аналог независимо от платформы, программы и языка». Дополнительную информацию о системе Unicode можно найти на сайте: <http://www.unicode.org/standard/WhatIsUnicode.html>.

SIL участвовал в разработке многочисленных предложений по расширению стандарта Unicode, чтобы включить в него особые знаки, используемые в мигоритарных языках. Включение таких знаков в стандарт необходимо для адек-

ватной поддержки коммерческого программного обеспечения, программного обеспечения с открытым кодом и операционных систем. Без этой поддержки носители миноритарных языков остаются на дне цифровой пропасти. В последние годы SIL работал со следующими шрифтами:

- Arabic;
- Cham;
- Cyrillic;
- Dai Banna (New Tai Lue);
- Devanagari;
- Lanna (Tai Tham);
- Latin;
- Lisu;
- Myanmar;
- Tai Viet;
- Vai.

В проектах Unicode участвуют следующие партнеры:

Tai Viet – рабочая группа по созданию шрифта Tai Viet (Tai Viet Working Group), ЮНЕСКО, проект Script Encoding Initiative;

Vai – Лютеранская миссия переводчиков Библии (Lutheran Bible Translators), Либерийская организация по переводу Библии и распространению грамотности (Liberia Bible Translation and Literacy Organization);

Latin – Лютеранская миссия переводчиков Библии (Lutheran Bible Translators), организация Pioneer Bible Translators;

Cyrillic – Институт перевода Библии;

Dai Banna – Университет Пайап (Payap University), Юньнаньская комиссия по миноритарным языкам (Yunnan Minority Language Commission);

Lanna – Университет Пайап (Payap University), Майкл Эверсон (Michael Everson, Evertype), Университет Чианг Май (Chiang Mai University);

Lisu – Юньнаньская комиссия по миноритарным языкам (Yunnan Minority Language Commission) и др. (См. выражения признательности на сайте <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3424.pdf>);

Myanmar – Университет Пайап (Payap University), Майкл Эверсон (Michael Everson, Evertypе).

Дополнительную информацию о деятельности SIL, связанной с Unicode, можно найти на сайте <http://scripts.sil.org/PCUnicodeDocs>.

Визуализация

Важнейшим вкладом SIL в технологию визуализации является разработка шрифтов, совместимых с Unicode, и универсального модуля рендеринга с открытым кодом под названием Graphite.

SIL принимает участие в разработке шрифтов для наименее известных языков мира, а также высококачественных шрифтов для основных языков. За последние три года мы самостоятельно или в партнерстве разработали или усовершенствовали следующие варианты шрифтов на основе Unicode.

Разработанные шрифты:

- Arabic: Scheherazade and Lateef;
- Roman/Cyrillic/Greek: Doulos, Charis, Andika, Gentium;
- Tai Viet: Tai Heritage.

Обновленные шрифты:

- Dai Banna: Dai Banna SIL;
- Ethiopic: Abyssinica;
- Limbu: Namdthinggo SIL;
- Myanmar: Padauk.

Участие в работе:

- N’Ko: Conakry.

Текущая работа:

- Devanagari;
- Tifinagh.

Среди наших партнеров по разработке шрифтов: при разработке шрифта Conakry – Майкл Эверсон, проект Script Encoding Initiative; шрифта Padauk – Университет Пайап (Payap University).

Получить дополнительную информацию и скачать шрифты можно на сайте <http://scripts.sil.org/FontDownloads>.

SIL также разработал и предложил Лицензию на открытые шрифты (SIL Open Font License). Это проверенная система лицензирования, соответствующая принципам свободного и открытого программного обеспечения (FLOSS), которая позволяет осуществлять свободное сотрудничество в разработке, совместном использовании и распространении шрифтов. Лицензия доступна всем желающим участвовать в разработке шрифтов.

Хотя сложные системы письма требуют специальных знаний и практических навыков в области работы со шрифтами, эти ресурсы в большей части мира ограничены. Сотрудничество разнообразных партнеров, объединенных лицензией FLOSS и использующих общую методологию, позволяет многим языковым общинам получить качественные и удобные в обращении компоненты систем письма. В рамках лицензии на открытые шрифты в настоящее время предлагается более 50 шрифтов.

Дополнительная информация представлена по адресу: <http://scripts.sil.org/OFL>.

Graphite – программный пакет с открытым кодом для разработки шрифтов по технологии «smart font», способных обслуживать сложные системы письма, – появился специально для удовлетворения нужд миноритарных языковых общин, для письменностей которых не разработаны коммерческое программное обеспечение и, возможно, стандарты типа Unicode. В последние несколько лет SIL продолжал разработку пакета Graphite и добился значительных успехов в его поддержке платформами Windows и Linux.

Дополнительная информация представлена по адресу: <http://graphite.sil.org>.

С января по октябрь 2003 г. SIL вместе с ЮНЕСКО участвовал в коллективном проекте, являвшемся частью программы ЮНЕСКО B@bel. Основной целью было обеспечение разработки комплексной системы поддержки письменностей на основе ИКТ. Эта цель была достигнута за счет подготовки документации по управлению разработкой, технических средств комплексной поддержки письменности и двух примеров реализации письменности.

Дополнительная информация представлена по адресу: <http://scripts.sil.org/babel>.

Языковые коды стандарта ISO 639-3

Уже более 50 лет SIL принимает участие в идентификации и каталогизации языков мира. В первичном виде эта информация представлена в «Этнологе», который издается в печатном виде, а как интернет-ресурс доступен по адресу: <http://www.ethnologue.com/>.

Система стандартов ISO 639, основанная на материалах «Этнолога», расширена в феврале 2007 г. путем официального принятия стандарта ISO 639-3. В этой новой части стандарта предоставлен наиболее полный перечень языков, включая живые, мертвые, древние, искусственные – как распространенные, так и редкие, как письменные, так и бесписьменные.

Код ISO 639-3 предназначен для определения трехбуквенных идентификаторов всех известных языков человечества. Его основу составляют около 400 отдельных языков, ранее учтенных в стандарте ISO 639-2. Помимо этих языков в ISO 639-3 включено свыше 6500 живых языков, которые были взяты главным образом из 15-го издания «Этнолога», а также добавлены мертвые, древние, исторические и искусственные языки на основе ресурса Linguist List.

SIL International был назначен Центром регистрации (RA) ISO 639-3 для обработки заявок на обновление набора языковых кодов Alpha-3. Центр регистрации ISO 639-3 получает заявки на новые языковые коды и на изменение существующих по критериям, оговоренным данным стандартом, и ведет их учет.

Работа с новыми заявками в Центре регистрации начинается лишь после того, как они пройдут публичное рецензирование.

Дополнительная информация представлена на официальном сайте ISO 639-3: <http://www.sil.org/iso639-3/>.

Языковое программное обеспечение

SIL разработал более 60 различных программных приложений для поддержки деятельности своего выездного персонала. Большинство программ можно бесплатно загрузить с сайта SIL по адресу: <http://www.sil.org/computing/catalog/>. Это программное обеспечение включает приложения для акустического анализа звуков речи (Speech Analyzer), фонологического анализа (Phonology Assistant) и сбора и анализа лексических данных (FieldWorks Language Explorer). Последний является эффективным инструментом регистрации лексических единиц и создания подстрочников и словарей.

Lexique Pro (<http://www.lexiquepro.com/>) является еще одним свободным программным приложением SIL, важным для развития многоязычия в киберпространстве. Это приложение позволяет лингвистам публиковать онлайн-словари с гиперссылками между статьями, средствами определения категорий и обратными словарями. На сайте публикуются лексиконы, созданные с использованием Lexique Pro. Примеры представлены по адресу: <http://www.lexiquepro.com/library.htm>.

Языки жестов

Одной из новых областей деятельности SIL является рационализация технических решений, касающихся всемирной поддержки языков жестов. Технический прогресс в области анимации и видео сделал возможной компьютерную обработку таких языков. В отношении языков жестов ставятся те же технические задачи, что и в отношении письменных языков, а именно: методы ввода, кодирования и визуализации. Более подробная информация о нашей работе с мексиканскими языками жестов представлена на <http://www.sil.org/mexico/lenguajes-de-signos/00i-signed-languages.htm>

Документирование языков и документальная лингвистика

Многие всерьез обеспокоены тем, что в мире существует значительное число недокументированных языков, в большей или меньшей степени находящихся под угрозой исчезновения. Разделяя это беспокойство, SIL установил партнерские отношения с рядом организаций для определения путей документирования таких языков и распространения информации о них, чтобы предотвратить их исчезновение. Мы считаем, что любой язык обладает изначально присущей ценностью и что носители миноритарных языков и другие заинтересованные стороны нуждаются в технических средствах и методах защиты и поддержки своего культурного и языкового наследия.

Наш институт не жалеет усилий для распространения по всему миру через Интернет информации о тех языках, с которыми мы работали. Наглядным примером является сайт Papua New Guinea Language Resource (языковой ресурс Папуа-Новой Гвинеи) <http://www.sil.org/pacific/png/index.asp>. Другим примером является мексиканский сайт SIL <http://www.sil.org/mexico/00i-index.htm>. На сайты SIL, относящиеся к отдельным странам, можно перейти с нашего основного сайта <http://www.sil.org/> через меню «Who We Are».

SIL также поддерживает партнерские отношения с другими организациями для облегчения обмена информацией о языках и поддержки использования всех языков в киберпространстве. Одним из примеров такой деятельности является сайт Всемирной сети в поддержку языкового разнообразия МААҮА <http://www.maayajo.org/>.

Сообщество открытых языковых архивов

SIL играл ведущую роль в создании Сообщества открытых языковых архивов (Open Language Archives Community, OLAC) и продолжает работать над его совершенствованием. OLAC – это сообщество учреждений и людей, создающих всемирную виртуальную библиотеку языковых ресурсов.

Сейчас OLAC насчитывает 37 учреждений, которые представляют все свои языковые ресурсы в форме каталогизационных описаний. Речь может идти и о материалах, хранящихся в обычном архиве, и о традиционных печатных изданиях, и о страницах на веб-сайтах. Все эти каталоги связывает стандарт ISO 639-3 (языковые коды), которым участники пользуются для обозначения ресурсов на соответствующих языках или о них. Поскольку во всех архивах применяются одни и те же языковые коды, пользователь, интересующийся конкретным языком, найдет в этих архивах все, что имеет к нему отношение. На сегодняшний день в объединенном каталоге имеются записи по более чем 35 тыс. ресурсов, и это только начало. Членство в организации бесплатное. В ее работе может участвовать любое учреждение, располагающее ресурсами на тех или иных языках национальных меньшинств или о них.

Заключение

Долгий опыт работы SIL по развитию миноритарных языков во всем мире дает ему уникальную возможность для развития и обогащения технической компетентности как на местном, так и на международном уровне. Мы благодарны за эту возможность и сознаем свою ответственность за распространение своих знаний и помощь другим, занятым в этой области.

Мы признаем, что без посторонней помощи не можем надеяться надолго сохранить влияние на потребности языковых сообществ всего мира, желающих увеличить ценность своего языка и выиграть от продолжения его использования. SIL хочет быть достойным участником общего дела развития мировых языков. Мы рады случаю поделиться с вами нашими достижениями в надежде, что они принесут пользу другим, в особенности самим носителям миноритарных языков.