**Приложение**

**Информация о работе отдела прикладной семиотики**

**Академии наук Чеченской Республики**

**по проекту «Чеченская локализация компьютерных программ»**

**за период с 2014 по 2018 гг.**

Проект «Чеченская локализация компьютерных программ» подразумевает: перевод компьютерных программ на чеченский язык и адаптация их к культуре чеченского народа, создание для пользователя комфортных условий для набора и обработки текстов на чеченском языке на компьютере и мобильных устройствах, автоматическую проверку орфографии и грамматики, распознавание и чтение чеченских текстов, создание обучающих компьютерных программ и т.д. Аналогичный проект по внедрению татарского языка в компьютерные технологии разрабатывается в НИИ Прикладной семиотики Академии наук Татарстана с 1993 года.

Для внедрения чеченского языка в компьютерные технологии в Академии наук Чеченской Республики в 2014 году создан отдел прикладной семиотики, штат которого в настоящее время составляет 5,25 единицы. Над реализацией проекта работают следующие сотрудники отдела:

1. Умархаджиев С.М., к. ф.-м. н., зав. отделом;
2. Асхабов Х. И., к. т. н., ведущий научный сотрудник (0,25 ставки);
3. Бадаева А. С., научный сотрудник;
4. Бекаев М. Х., научный сотрудник;
5. Израилова Э. С., старший научный сотрудник;
6. Султанов З. А., младший научный сотрудник,

а также – Абдулкадыров А., научный сотрудник отдела языка и литературы АН ЧР, и Эльсаев Я. В., младший научный сотрудник (0,2 ставки) КНИИ РАН.

**Этапы реализации проекта**

**Этап 1.** В первую очередь для реализации проекта “Чеченская локализация компьютерных программ”  необходимо было иметь электронные словари чеченского языка. Была проведена кропотливая работа по оцифровыванию основных словарей:

1. Мациев А.Г. «Чеченско-русский словарь». М. 1961 (словарь Мациева);
2. Карасаев А.Т., Мациев А.Г. «Русско-чеченский словарь». М. 1978;
3. Абдурашидов Э.Д. Чеченско-русский русско-чеченский словарь юридических терминов. Грозный. 2010.
4. Умархаджиев С.М., Ахматукаев А.А. Чеченско-русский, русско-чеченский словарь математических терминов. Тбилиси. 2010.

Эти словари, а также программы, созданные в отделе прикладной семиотики, выложены для свободного доступа на сайтах Академии наук [www.anchr.ru](http://www.anchr.ru/) и отдела прикладной семиотики <http://ps95.ru/category/than-belhash/>.

Разработчиками проекта подготовлены и изданы следующие словари:

1. Умархаджиев С.М., Асхабов Х.И., Бадаева А.С., Вагапов Ӏ.Д., Израилова Э.С., Султанов З.А., Астемиров А.В. Оьрсийн-нохчийн, нохчийн-оьрсийн компьютерийн лексикин дошам. Сольжа-ГӀала. 2017.
2. Абдулкадыров А.Т. Лексико-семантический тезаурус чеченского языка. Грозный. 2018.
3. Абдулкадыров А.Т. Словарь синонимических парадигм чеченского языка. Грозный. 2018.

На основе всех оцифрованных словарей создано веб-приложение «Нохчийн-оьрсийн, оьрсийн-нохчийн онлайн дошам “ДикДошам”, которым можно пользоваться через Интернет на сайтах отдела и Академии наук <http://ps95.ru/dikdosham/>.

**Этап 2.** Основным этапом локализации компьютерных программ является создание лексической базы языка – электронной базы всех слов языка. При создании такой базы чеченского языка возникли сложности, связанные с многозначностью лексических единиц на письме, что усложняет многие задачи компьютерной лингвистики. Речь идет, в первую очередь, о существовании огромного количества омографов в чеченском языке, существующих вследствие несовершенств графики и орфографии чеченского языка. Было принято решение: сотрудникам отдела создать лексическую базу чеченского языка на основе словаря Мациева и одновременно ученым-филологам активизировать исследования по совершенствованию норм орфографии.

Сотрудниками отдела создана электронная лексическая база, содержащая более 2,5 миллиона слов и словоформ. Для автоматизации процесса подготовки этой базы созданы компьютерные программы на различных языках программирования. В Роспатенте получено 8 свидетельств об официальной регистрации программ для ЭВМ (см. Приложение 1).

В результате исследовательской работы сотрудников отдела в области компьютерной лингвистики и практической реализации новых программных разработок опубликовано в научных журналах 17 статей (Приложение 2), сделано 4 доклада на международных и общероссийских научных конференциях (Приложение 3) и поданы 2 заявки на конкурсы грантов Российского фонда фундаментальных исследований (Приложение 4).

**Этап 3.** Первоочередной задачей внедрения чеченского языка в информационные технологии является организация автоматической проверки правописания текстов на чеченском языке. Так как программные продукты фирмы Microsoft являются коммерческими программами с закрытым кодом, сотрудниками отдела организована проверка орфографии на основе созданной лексической базы в операционной системе Linux, аналогичной Windows, но имеющей открытый код.

Таким образом, в настоящее время компьютеры отдела прикладной семиотики при наборе на компьютере чеченского текста автоматически проверяют правильность написания слов, и предлагает варианты замены в случае неправильного написания.

Проводится работа по тестированию программы проверки орфографии. Выявлено большое количество неточностей. Это является следствием несоответствия словаря Мациева, изданного в 1961 году, современному чеченскому языку и несовершенства орфографии языка.

**Этап 4.** Чтобы любой пользователь мог настроить свой компьютер на проверку чеченских текстов, подготовленные в отделе программы должны быть встроены в редактор Word коммерческой компании Microsoft. Для доступа к коду редактора текстов Microsoft Word нужно получить разрешение компании Microsoft. На обращение в офис Microsoft было предложено выложить идею в чат сайта компании Microsoft для голосования. Идея получила более 12 тыс. голосов (<https://word.uservoice.com/forums/304924-word-for-windows-desktop-application/suggestions/32108578-chechen-spelling-dictionary>). Теперь остается ждать реакции компании Microsoft.

**Этап 5.** В рамках основного проекта сотрудниками отдела исследуются задачи по речевым технологиям, в частности, разрабатываются методы компьютерного синтеза чеченской речи и распознавания чеченских текстов. Изучены различные современные системы речевых технологий, налажены научные контакты с сотрудниками Санкт-Петербургского института информатики и автоматизации РАН, Эдинбургского университета (Великобритания), НИИ Прикладной семиотики Академии наук ТР. Научные результаты по этому направлению опубликованы в пяти статьях.

**Этап 6.** В отделе начата работа по переводу и адаптации к чеченской культуре операционной системы Windows фирмы Microsoft.Из-за необходимости перевода на чеченский язык большого объёма технического текста и нехватки научных сотрудников разработка этого проекта приостановлена.

Заведующий отделом прикладной семиотики

Академии наук Чеченской Республики,

кандидат физ.-мат. наук С. М. Умархаджиев

**Приложение 1**

**Список свидетельств о регистрации программ для ЭВМ**

1. Свидетельство № 2018617362 об официальной регистрации программы DoshStat для ЭВМ / С.М. Умархаджиев, М.Х. Бекаев, А.С. Бадаева, Э.С. Израилова, З.А. Султанов, Х.И. Асхабов, Я.В. Эльсаев, А.А. Абдулкадыров; правообладатель Академия наук Чеченской Республики. (http://www1.fips.ru/ fips\_servl/fips\_servlet?DB=EVM&rn =5810&DocNumber=2018617362&TypeFile=html)
2. Свидетельство № 2017616337 об официальной регистрации программы Khiytsar для ЭВМ / С.М. Умархаджиев, А.В. Астемиров, А.С. Бадаева, С.А. Курбанова, З.А. Султанов; правообладатель Академия наук Чеченской Республики. (http://www1.fips.ru/ wps/PA\_FipsPub/res/Doc/PrEVM/RUNWPR/000/002/017/616/337/2017616337-00001/ DOCUMENT.PDF)
3. Свидетельство № 2017616041 об официальной регистрации программы DikDosham для ЭВМ / С.М. Умархаджиев, А.В. Астемиров, А.С. Бадаева, С.А. Курбанова, З.А. Султанов; правообладатель Академия наук Чеченской Республики. (http://www1.fips.ru/wps/PA\_FipsPub/res/Doc/PrEVM/RUNWPR/000/002/017/616/041/ 2017616041-00001/DOCUMENT.PDF)
4. Свидетельство № 2016662389 об официальной регистрации программы Nokhchiyn\_ tsIerash для ЭВМ / С.М. Умархаджиев, А.В. Астемиров, А.С. Бадаева, С.А. Курбанова, З.А. Султанов; правообладатель Академия наук Чеченской Республики. (http://www1.fips.ru/Archive/EVM/2016/ 2016.12.20/ Index.htm)
5. Свидетельство № 2016614148 об официальной регистрации программы Bilgaldosh для ЭВМ / С.М. Умархаджиев, З.А. Султанов, А.С. Бадаева, А.В. Астемиров, С.А. Курбанова; правообладатель Академия наук Чеченской Республики. (http://www1. fips.ru/Archive/EVM/2016/2016.05.20/Index.htm)
6. Свидетельство № 2016614149 об официальной регистрации программы Khandosh для ЭВМ / С.М. Умархаджиев, С.А. Курбанова, А.В. Астемиров, А.С. Бадаева, З.А. Султанов; правообладатель Академия наук Чеченской Республики. (http://www1.fips.ru/ Archive/EVM/2016/2016.05.20/Index.htm)
7. Свидетельство № 2016612664 об официальной регистрации программы TsIerdosh для ЭВМ / С.М. Умархаджиев, З.А. Султанов; правообладатель Академия наук Чеченской Республики. (http://www1.fips.ru/Archive/ EVM/2016/ 2016.04.20/Index.htm)
8. Свидетельство № 2014619220 об официальной регистрации программы Elp-I для ЭВМ / С.М. Умархаджиев, А.В. Астемиров; правообладатель Чеченский государственный университет. (http://www1.fips.ru/Archive/EVM/2014/2014.10.20/DOC/RUNW/ 000/002/014/619/220/DOCUMENT.PDF)

**Приложение 2**

**Список публикаций**

1. [Izrailova](https://link.springer.com/search?facet-creator=%22E.+S.+Izrailova%22) E.S. [The Phonetic Alphabet of the Chechen Language as a Basis of a Speech-Synthesis System](https://link.springer.com/article/10.3103/S0005105518010077) / [E. S. Izrailova](https://link.springer.com/search?facet-creator=%22E.+S.+Izrailova%22) //[Automatic Documentation and Mathematical Linguistics](https://link.springer.com/journal/11952). — 2018 vol.52. — p. 51-55.
2. Алисултанова Э.Д., Израилова Э.С., Халиева Х.С. Синтез и распознавание речи с помощью нейронных сетей. // Сборник статей Всероссийской научно-практической конференции "Молодежь, наука, инновации". – Грозный. – 2016. – С. 219-223.
3. Асхабов Х.И. [Современные проблемы компьютеризации чеченского языка](https://elibrary.ru/item.asp?id=30716301) / Х.И. Асхабов // [Известия Чеченского государственного педагогического института](https://elibrary.ru/contents.asp?issueid=1908083). – 2017. – Т. 16. –  [№ 3 (19)](https://elibrary.ru/contents.asp?issueid=1908083&selid=30716301). – С. 15-21.
4. Асхабов Х.И. Файловые менеджеры: история и современность. Известия ЧГПУ, №4 (20), Серия №1. Гуманитарные и общественные науки. - 2017.
5. Асхабов Х.И., Батаева Я.Д., Джабраилова Р.Д. [Визуализация уравнений второго порядка средствами объектно-ориентированного программирования](https://elibrary.ru/item.asp?id=24329381) // [Информацион-ные технологии в преподавании и научно-технический перевод](https://elibrary.ru/item.asp?id=24327538), сборник научно-методических статей. – Москва. – 2014. – С. 144-152.
6. Бадаева А.С. О некоторых вопросах чеченской графики и орфографии в контексте локализации компьютерных программ. // Вестник Академии наук Чеченской Республики. – 2016. – № 3 (32). – С. 80-84.
7. Бадаева А.С., Израилова Э.С., Султанов З.А., Эльсаев Я.Г. Нохчийн меттан лексикин электронан база. // Таллам. 2017. №1.
8. Бадаева А.С., Курбанова С.А. Нохчийн меттан информацин технологишкахь кхиаран тӀегӀа а, башхаллаш а. // Таллам. 2016. – № 1. – С. 32-36.
9. Бадаева А.С., Султанов З.А. Нохчийн маттахь цӀердешнаш легаден «TCIERDOSH» программин белхан алгоритман коьрта принципаш // Таллам. 2017. – №2. – С. 70-78.
10. Израилова Э.С. «Фонетический алфавит» чеченского языка как основа системы синтеза речи / Э. С. Израилова // НТИ. сер. 2. Информ. процессы и системы/ ВИНИТИ РАН.— 2018 № 2.— С. 35-39
11. Израилова Э.С. О создании аудиословарей в редакторе Learnwords Editor /Э.С. Израилова // Компьютерные технологии и телекоммуникации КТиТК-2014: cб. ст. конф. / Грозненский гос. нефтяной тех. ун-т им. М.Д. Миллионщикова. – Грозный, 2014. – С. 112-115. (<https://elibrary.ru/item.asp?id=25660864>)
12. Израилова Э.С. О создании фонетико-акустической базы в рамках синтеза чеченской речи. /Э.С. Израилова // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. – 2017. – № 2. – С. 111-115.
13. Израилова Э.С. Создание электронных словарей – первый этап чеченской локализации компьютерных программ / Э.С. Израилова // Вестник Акад. наук Чеченской Республики. – 2013.– № 2. – С. 88–90. ([https://elibrary.ru/download/elibrary\_21134247\_ 42462899.pdf](https://elibrary.ru/download/elibrary_21134247_42462899.pdf))
14. Израилова Э.С., Шамсутдинов Х.Н. Автоматизированные системы общения с человеком. // Материалы международной научно-практической конференции «Фундаментальные и прикладные исследования: проблемы и результаты». ГГНТУ. г. Грозный. 2-4 ноября, 2017.
15. Умархаджиев С.М., Астемиров А.В., Асхабов Х.И., Бадаева А.С., Вагапов А.Д., Израилова Э.С., Султанов З.А. Оьрсийн-нохчийн, нохчийн-оьрсийн компьютерийн лексикин дошам. Грозный, АН ЧР, 2017. 238 с.
16. Умархаджиев С.М., Астемиров А.В., Бадаева А.С., Курбанова С.А., Султанов З.А. Компьютерная программа «KHANDOSH» для спряжения глаголов чеченского языка. Вестник Академии наук Чеченской Республики. 2015. № 4 (29). С. 88-92. - Грозный, 2015.
17. Якубов А.В., Асхабов Х.И. Толковый словарь пользователя информационно-коммуникационных технологий. // Махачкала: Алеф, 2014. – 134 с.

**Приложение 3**

**Доклады на международных и общероссийских научных конференциях**

1. Израилова Э.С. Доклад «Синтез и распознавание речи с помощью нейронных сетей» на Всероссийской научно-практической конференции «Молодежь, наука, инновации», Грозный, 20 октября 2016.
2. Асхабов Х.И. Доклад на республиканской научной конференции, посвященной Дню чеченского языка «Новый текстовый редактор с проверкой орфографии чеченского языка» 23 апреля 2018.
3. Израилова Э.С. Доклад на республиканской научной конференции, посвященной Дню чеченского языка «Создание системы автоматического синтеза чеченской речи» 23 апреля 2018.
4. Бадаева А.С. Доклад на республиканской научной конференции, посвященной Дню чеченского языка «Нохчийн маттахь нийсаяздар компьютер чохь таллар. Коьрта декхарш, уьш кхочушдаран некъаш а» 23 апреля 2018.

**Приложение 4**

**Заявки на конкурсы научных грантов отдела прикладной семиотики АН ЧР**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Год** | **Фонд** | **№ заявки** | **Тема** |
|  | 2017 | РФФИ | 18-312-00192 | Электронный грамматический словарь чеченского языка |
|  | 2017 | РФФИ | 18-07-00450 | Комплекс программных средств для грамматического анализа и обработки текстов на чеченском языке |