**Подпроектное предложение**

|  |  |
| --- | --- |
| **Название подпроекта** | Оцифровка кавказские языки |
| **Название партнерской организации/инициативной группы** | 1. Благотворительный Фонд развития абхазского языка им. Баграта Шинкуба  2. Адыгейский республиканский институт гуманитарных исследований им. Т.М. Керашева  3. Издатели книг, авторы книг  4. Mozilla ([https://commonvoice.mozilla.org](https://commonvoice.mozilla.org/))  5. Wikipedia ([https://www.wikipedia.org](https://www.wikipedia.org/)) |
| **Общий профиль и мандат нашей группы** | Основные направления деятельности нашей группы, это создание и внедрение новых методик и технологий по развитию и изучению языка  Наша группа уникально подходит для достижения поставленных целей, в нашей команде уже задействованы специалисты всех необходимых направлений, есть техническая основа, на данном этапе необходим человеческий ресурс для создания объемной языковой базы данных. |
| **Общая сумма** | 4,440,000₽ |
| **Срок** | Один год |
| **География подпроекта** | Сухум, Майкоп |

**Контактная информация**

|  |  |
| --- | --- |
| **Подпроект-менеджер** | **Альтернативное лицо** |
| **Имя:** Леуан Микаа | **Имя:** Леуан Лагулаа |
| **Эл. почта:** leuanmik@gmail.com | **Эл. почта:** lagula@mail.ru |
| **Телефон:** +7(940) 700-37-77 | **Телефон:** +7(940) 722-48-08 |

**Цель подпроекта**

Цель подпроекта – расширение использования абхазского и адыгского языки в цифровом пространстве. Конкретная цель - предоставить набор данных с открытым доступом, который можно использовать для внедрения новых технологии на адыгском и абхазском языках. Проблема заключается в отсутствии открытого абхазского и адыгского наборы данных в нужном формате, чтобы использовать его в задачах НЛП (обработки естественного языка). Эти задачи позволят языку развиваться и идти в ногу с новейшими технологиями и гаджетами (т.е. Локализация приложений, перевод документов, голосовые помощники).

К таким задачам относятся, среди прочего:

• автоматический перевод текстов;

• преобразование речи в текст,

• преобразование текста в речь.

Наша абхазско-адыгская группа свяжется с авторами, чтобы опубликовать их текстовые материалы в открытом доступе, затем переформатирует текст в предложения, и мы привлечём людей, чтобы записывать их голоса в Комон Войс, а также для перевода статей Википедии, подходящие для обучения искусственному интеллекту на абхазский и адыгский языки.

**Обоснование выбора заинтересованных сторон/партнеров:**

1. Издатели книг, авторы книг (писатели, переводчики, поэты) — предоставляют произведения без авторских прав.

2. Mozilla (https://commonvoice.mozilla.org) Роль - предоставить площадку, необходимую для реализации набора голосовых данных с открытым доступом.

3. Wikipedia (https://www.wikipedia.org) Роль - предоставить площадку, необходимую для реализации набора текстовых данных с открытым доступом.

**Стратегия подпроекта**

*Если мы предоставим большие наборы данных с открытым доступом (Комон Войс, И Википедия), то* мы можем выполнять множество низкоуровневых задач НЛП (обработки естественного языка). (т.е. машинный перевод, преобразование речи в текст);

*Если* мы выполняем низкоуровневые задачи НЛП (обработки естественного языка), *то* мы можем быстро выполнить задачи высокого уровня (т.е. Локализация приложений, перевод документов, голосовые помощники);

*Если* мы можем быстро выполнить задачи высокого уровня, *то* мы позволим абхазскому и адыгскому языкам широко присутствовать в цифровом мире;

*Если* мы позволим абхазскому и адыгскому языкам широко присутствовать в цифровом мире, и обеспечить в глобальном масштабе легкий доступ для тех, кто не говорит на нем, *то* абхазский и адыгский языки станут функциональными языками.

Если абхазский и адыгский языки станут функциональными языками, то мы можем предоставить опыт и знания о том, как возродить уязвимые языки.

**Общая цель подпроекта**

Мы позволим абхазскому и адыгскому языкам широко присутствовать в цифровом мире и обеспечить в глобальном масштабе легкий доступ для тех, кто не говорит на нем. Они станут функциональными языками, так как мы выполняем низкоуровневые задачи НЛП (обработки естественного языка), и мы можем быстро выполнить задачи высокого уровня (т.е. локализация приложений, перевод документов, голосовые помощники).

**Конкретные цели подпроекта**

Мы предоставим большие наборы данных с открытым доступом, чтобы создать набор голосовых данных с помощью Комон Войс (https://commonvoice.mozilla.org), а также текстовых данных с помощью Wikipedia (https://www.wikipedia.org).

Также предоставим прототип абхазско-адыгско-русского машинного переводчика и голосового помощника.

**Предлагаемые мероприятия**

|  |  |  |
| --- | --- | --- |
| **Мероприятие** | **Методология** | **Результат** |
| **Комон Войс** | • IT специалист и специалист по языку для очистка и проверка предложений путем разработки и внедрения мер обеспечения качества.  • Помощники предстоит записывать голоса людей, в среднем 40 человек в месяц. | 400 тысяч предложений на адыгском  566 аудиочасов от 1600 человек (адыгский)  475 аудиочасов от 1345 человек (абхазский) |
| **Википедия** | • Специалисты по языку переводят страницы Википедии, в среднем 50 страниц формата А4 в месяц.  • Редакторы проверяют переводы.  • IT-специалист технической поддержки. | 1800 страниц формата А4 переведенного текста (адыгский)  1800 страниц формата А4 переведенного текста (абхазский) |

**Бюджет**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Мероприятие** | **Участники** | **Количество** | **Месячная зарплата** | **Сумма** |
| **Комон Войс** | Помощник | 6 | 25 тысяч | 1,800,000₽ |
| Специалист по языкам | 2 | 5 тысяч | 120,000₽ |
| IT специалист | 1 | 20 тысяч | 240,000₽ |
| Дополнительно транспорт и связь | 1 | 10 тысяч | 120,000₽ |
| **Википедия** | Специалист по языкам | 6 | 25 тысяч | 1,800,000₽ |
| Редактор | 2 | 20 тысяч | 480,000₽ |
| IT специалист | 1 | 20 тысяч | 240,000₽ |
| **Оба проекта** | PR-менеджер | 1 | 15 тысяч | 180,000₽ |
| **Общая Сумма** |  |  |  | ₽ |