

**Проект: №423**  
**Направление: Приложни програми**  
**ТЕМА: Речник на българския архаизъм**

<https://github.com/OgiJr/Bulgarian-Archaism-Dictionary>

**АВТОРИ:**

**Създатели:**

Огнян Траянов, Американския Колеж, [o.trajanov22@acsbg.org](mailto:o.trajanov22@acsbg.org)

Йоана Станкова, Американския Колеж, [y.stankova22@acsbg.org](mailto:y.stankova22@acsbg.org)

**РЪКОВОДИТЕЛ:**

Д-р Паулина Иванова Тодорова [p.todorova@acsbg.org](mailto:p.todorova@acsbg.org),

Телефон: 089 619 4957

Информатика и Информационни Технологии,  
АКС

**РЕЗЮМЕ:**

**Идея на проекта:**

В българският език и литература се използват множество различни архаизми и диалекти. Те могат да затруднят учениците, които не ги разбират, но също така правят нашия език по-богат. Затова, те не могат да бъдат заменени, за да се улеснят произведенията.

Тази тема стана изключително актуална, когато през 2019г. бяха сменени 6000 архаични думи в „Под игото“ на Иван Вазов, което доведе до множество отзиви на различни новинарски компании и индивидуални създатели в Интернет пространството в полза на премахването на тези промени.

Нашето приложение иска да премахне нуждата за замяна на архаизма и диалекта в произведенията, които влизат в програмата на МОН, като направи дигитален речник, чрез който да се улесни изучаването на тези произведения.

До сега вече има няколко дигитални речници за българските архаизми и диалекти, но техният формат (PDF в повечето случаи или уебсайт с изредени всички термини) не им позволява да бъдат лесно използвани и ги доближава до хартиените речници.

Нашият проект ще се състои от четири части: база данни, в която са изписани необходимите думи и е достъпна за всички; приложение, чрез което може потребителят да потърси определена дума в базата данни; уебсайт, който ще има предишната функция, както и чрез него ще могат да се предложат нови думи за нашата база данни и хром добавка, с която може да се сканира дадена дума и да се даде нейната дефиниция от базата данни..

Един от аргументите срещу интеграцията на информационните технологии е необходимостта от много ресурси и средства. Но тук идват на помощ смартфоните на учениците - средство, което досега е използвано, като източник за разсейване.

---

Смартфоните обаче могат да се превърнат в източник за знание и учение. Почти всеки ученик има достъп до телефон и Интернет, което му позволява да използва тази технология, която от своя страна не носи разход за държавата и МОН да екипира училищата с лаптопи, таблети или хромбуци.

### Компоненти на проекта:

#### База данни:

Думите, които приложението „превежда“ се намират в наша базата данни. Тя съдържа тяхното значение и техните синоними. Базата данни е публично достъпна, т.е. ако се въведе нейния URL в Интернет може да се види нейното съдържание. Към самата база данни могат да се добавят или редактират думи по всяко време. За повече информация за това действие, прочетете разделът за Framework design на проекта.

#### Framework Дизайн:

Целта на проекта е потребителите сами да могат сами да вкарват свои *думи и тяхното значение* към базата данни. Това ще помогне разширяването на нашия архив от думи и ще направи приложението ни по-полезно за потребителите. Това се осъществява през уебсайта на проекта. Когато потребителят предложи да се качи нова дума към проекта, се изпраща имейл заявка на администраторите и те преценяват дали дадената дефиниция е правилна.

#### Уебсайт:

Освен миналоизброената функция, през уебсайтът може да се потърси някоя дума в базата данни, както и да се отворят *изцяло „преведени“ произведения* от българската литература. Т.е. има архив от различни творби, в които непознатите думи са заменени с познати, за да се улесни прочита на учениците, които желаят тази функция.

#### Chrome Extension:

Потребителите могат да изтеглят chrome extension към своят browser, с който като сканират в Интернет някоя непозната дума и изберат опцията за търсене в базата данни, ще им излезе дефиницията на тази дума. Това е изключително полезно при четенето на произведения от PDF документи или през някой уебсайт.

#### Режим OCR:

Решихме да използваме *OCR(Optical Character Recognition)*, за да улесним процеса на търсене на думи. С тази технология ученикът трябва само да сканира (снима) дадената дума, която не разбира или страницата, на която е написана и приложението ще сканира входните данни чрез Google Cloud Vision библиотеката, ще намери думата в базата данни и изпише дефиницията на дадената дума.

#### Употреба:

*Приложението може да влезе в употреба за:*

- 
- Книги и учебници – да се улесни четенето на произведение, в което има архаични думи или диалекти;
  - Ученици- в класната стая или от вкъщи;
  - Любители на българската литература;
  - Премахва нуждата за улесняване и редактиране на дадени произведения;
  - Други случаи, в които е използвана непозната за потребителя дума

### Стъпки на създаване на проекта:

- 
- Създаване на приложението в Xamarin Forms (Огнян)
  - Качване на приложението в Google Play Store и App Store
  - Създаване на база от данни (Огнян и Йоана)
  - Въвеждане на думите в базата данни използвайки различни източници и със съгласието на собственика на тези източници. (Огнян)
  - Създаване на уебсайт (Йоана)
  - Закупуване на domain на уебсайта
  - Създаване на Chrome Extension (Йоана)
  - Публикуване на extension-a

### Ниво на сложност:

Основни проблеми при реализацията на софтуера са основно архитектурни. С други думи това са трудности, възникнали при структуриране на необходима софтуерна архитектура за реализацията на проекта.

Един от най-големите проблеми е липсата на настояща база от данни с български архаизми. ***Ние трябва да използваме различни речници, PDF или на хартия и в повечето случаи да въведем тяхната информация на ръка.*** Също така трябва да имаме съгласието на собствениците на тези източници за тяхната употреба.

Друг голям приоритет на приложението е поддържането на различни устройства. Тъй като приложението трябва да поддържа максимално количество устройства, необходимо е софтуерът да бъде тестван на колкото се може повече смартфони.

Сложността на проекта е на *високо ниво*, тъй като изисква комбиниране на много различни технологии и модули, чиято координация е ключова за успеха на приложението.

### Реализация:

За да реализираме този проект използваме *Xamarin Forms, Visual Studio, Google Cloud Vision, PHP, Apache и MySQL*. Приложението се създава с *Xamarin Forms и Visual Studio*. Сканирането на думите става чрез *Google Cloud Vision*. Тази програма може да разпознава дадени думи от снимков материал. Тя предава информацията на приложението и то търси съответстваща дума в базата данни.

### Лиценз:

This project is licensed under the terms of the GNU GPLv3 License. For more information, see LICENSE.