

E-dictionary for Asian Languages

Marijana Janjić, Marko Požega, Dario Poljak, Sara Librenjak, Kristina Kocijan
Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučić 3, Zagreb, Croatia
marijanajanjic@yahoo.com, mpozega1@ffzg.hr, dpoljak1@ffzg.hr,
sara.librenjak@gmail.com, krkocijan@ffzg.hr

Summary

In this paper, authors present an e-dictionary of Asian languages designed for students whose primary language is Croatian. The project started with six Asian languages (Hindi, Japanese, Chinese, Korean, Persian and Sanskrit) all of which are using Croatian as a meta-language. Entries in Asian languages include several attributes (translation, Latin letters, grammatical notations, sentence with examples) that mainly depend on the language chosen for translation. Croatian entries are linked to the Croatian language portal where the meaning, word forms and categories are provided. The growth of the dictionary is largely dependent on students' participation.

Key words: e-dictionary, student's dictionary, Asian languages, Croatian language

Introduction

In order to get any work done, a student of Asian languages in Croatia must carry around several heavy dictionaries to each of the classes s/he takes. The survey we conducted among Croatian students of Asian languages at the beginning of this project revealed that they mostly use dictionaries with English-Asian language combinations. Apart from English-Croatian dictionary, not one of the used dictionaries includes Croatian. The reason behind the usage of English oriented dictionaries is (next to the fact that there are no Croatian-Asian languages dictionaries¹) that most of the study materials are also available in English rather than Croatian language, although we are talking about Croatian students studying in Croatia.

This situation inspired us to create an online dictionary for beginner learners of Asian languages, primarily including those Asian languages that are being studied at the University of Zagreb such as Hindi, Japanese, Chinese, Korean, Persian and Sanskrit.² We have opted for the e-version rather than the printed

¹ Apart from the Croatian-Japanese dictionary (2006) and Croatian-Turkish dictionary (2014).

² Turkish is also available at the University, but, so far, we do not have any resources for it.

version of such a dictionary for several reasons. The first reason was the time needed to produce such a book. The time needed to produce the e-version of a dictionary was relatively shorter compared to the printed dictionary. In addition, the e-version of a dictionary allows us to add, edit and delete as many entries as needed and to add new languages as well, all of which cannot be done with the printed version. Its on-line availability and free access allows students to ‘carry it around’ and have it available wherever and whenever they need, either via the website or as an android application.

In the following sections, we will shortly describe the content of this dictionary from the user’s and administrator’s perspectives.

The e-Dictionary Content

User’s perspective

We have opened the dictionary with 5.953 Croatian entries. However, this number is not evenly distributed among Asian languages. The languages with the most entries at the moment are Hindi (2 132) and Japanese (1 324) while Persian has the least (156). In between them are Sanskrit (1.028), Chinese (762) and Korean (668). At the moment, some words have been translated into only one and some into more Asian languages. It is possible to compare translations across different Asian language. This can be especially useful for learners of similar languages like Hindi and Sanskrit (Figure 1).

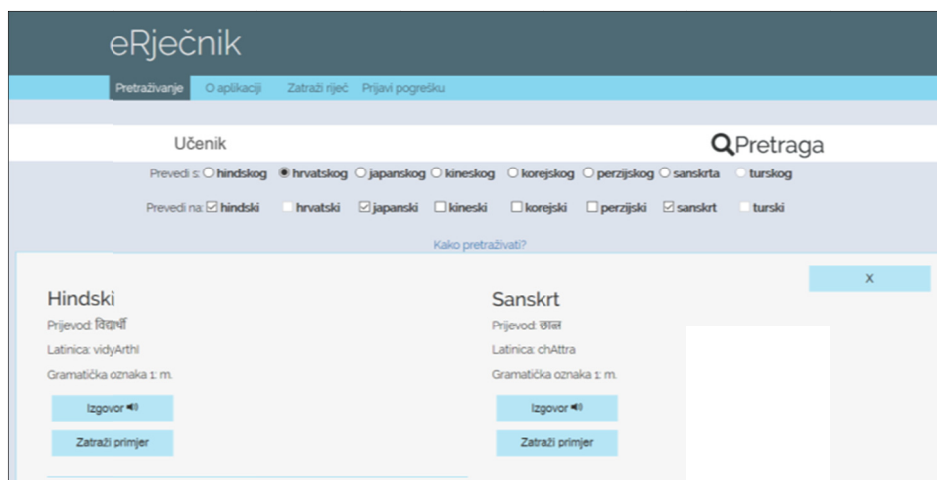


Figure 1. Comparing the Croatian word ‘student’ (hr. *učenik*) in Hindi and Sanskrit

Envisioned as a learner’s dictionary (beginner’s stage), it consists of elementary information that a student would need: (1) entry in one of Asian languages, (2)

grammatical information, if needed or available, (3) Croatian equivalent, (4) example(s) of use for each meaning.

The first three elements are what any reader would expect to find in a dictionary. The last element is, what we believe, the key element that separates this e-dictionary from others. Among many e-dictionaries available for the Asian languages that we came across, examples are not included in their structure. Our survey, however, showed that examples are of great importance to students. Even as experienced users of these languages, we are often at a loss when learning a new word from a dictionary trying to actually use it in a sentence.

The dictionary also provides a translation to Croatian if the searched item is in one of the offered Asian languages. However, we did not provide any additional information for Croatian words. With the permission of Croatian language portal³ that has an online database of Croatian, we have linked the words to their page where all the existing information for that word can be found.

The important role in building a learner's dictionary must also be given to learners themselves. For that reason, we have enabled to our users two-way communication with language administrators. This way they can ask for a new entry, new example for particular entry or suggest their own example for an existing entry.

Administrator's perspective

Administration of our e-dictionary allows us several possibilities: adding, editing and learning about the dictionary usage. The new words can be added either as single entries, mainly upon the user's request, or as a batch prepared by language administrators. All the entries that have already been stored in our database can be accessed and edited (i.e. fixing the errors, adding additional meanings, attributes, examples). The third section of the administration portal enables us an insight into the dictionary usage statistics, making sure that no personal data is collected or stored. We are thus familiar with the data on how many searches were made each day, what words were searched for, in what languages, if the search was successful or not. Additionally, we are using Google Analytics to check user's demographic as well as how many of them are new and how many returning users.

The Technology behind the e-Dictionary?

Developing a web application that is useful, intuitive and visually compelling is a big undertaking and is therefore commonly divided up in frontend and backend development. Our process was no different while developing e-dictionary as it encompasses a lot of information that has to be displayed in a suc-

³ <http://hjp.znanje.hr/>

cinct and readable manner whilst also implementing a number of features for both users and administrators.

Frontend development is commonly done with HTML, CSS and JavaScript. By adhering to some of the newest standards and technologies, we have been able to speed up the development and debugging time. In that regard, we have used CSS pre-processor SASS (Syntactically Awesome Stylesheet) which extends CSS functionality by providing many features from the simplest ability in creating reusable variables to creating whole snippets of extendable code.

To make the e-dictionary as fluid as possible, we have also opted for asynchronous bi-directional calls to the server. In this way, the user's page does not need to refresh each time the search is made. We have accomplished this by using AJAX (Asynchronous JavaScript) calls that were handled with a precompiled JavaScript library, jQuery. It has simplified the convoluted approach of writing reusable AJAX functions and has allowed us to efficiently request and receive data from the server in real time. With jQuery, we are able to send the searched item, source and target languages to the server. The results are returned in the form of an encoded JSON.

The entire e-dictionary system is built on the top of the Model–View–Controller (MVC) software architecture. The Model is the application object usually related with database entity, the View is its screen presentation, and the Controller defines the way the user interface reacts to the user input. Because e-dictionary uses pivot language, database must contain transfer tables between every language and pivot language. This kind of relational model requires smooth data manipulation using models and MVC.

Open source PHP framework, Laravel, offers full stack development environment following MVC architectural pattern, which we used to design data access layer (backend) and presentation layer (frontend) of our application. Furthermore, by using Laravel router we easily implemented AJAX web development technique resulting in zero page reloadings. Such approach increases user experience and leaves prepared setup for light JSON application programming interface (API) implementation. Our API is currently being used by android mobile application which is still in a development stage. Also, our API could be used by other applications which want to implement results of our dictionary search into their own products.

Conclusion

The development of this small e-dictionary has showed how cooperation of experts in different fields (linguists, lexicographers, information scientists) can be fruitful. As a result of scientific excellence, a new free on-line tool had been devised for Croatian students while the research team will gain further insight into students' habits and needs. The project has also showed how students can be and should be included in the development of tools oriented towards their needs.