EasyChair Preprint
№ 6766

LingvoDoc: Extralinguistic Information and
Maps

Natalia Koshelyuk

October 5, 2021

LingvoDoc: extralinguistic information and maps[1]

Kosheliuk Natalia
ORCID 0000-0002-5833-7971
Ivannikov Institute for System Programming of the RAS, Moscow (Russia)
NKoshelyuk@yandex.ru

Abstract. The paper is a continuation of a series of articles about work and user experience on the LingvoDoc linguistic platform. The methods of adding extralinguistic information to dictionaries and corpora are presented step by step, and the possibilities of working with dialectological maps on the platform are also considered.

Keywords. LingvoDoc, data mining, linguistics, maps, extralinguistic information, maps

1 INTRODUCTION

Extralinguistics is a field of linguistics that studies the total combination of social, historical, geographical and cultural factors associated with the development of language. According to F. de Saussure, one of the founders of structuralism in linguistics, the cultural environment in which a nation develops, its history and customs may influence the development of language. Extralinguistic information, which can now be added to LingvoDoc, in turn, serves as a source of additional information about a language or dialect. On the basis of this information it is possible to judge the conditions of functioning and development of these languages and also this information often helps linguists to ascertain the dialectal affiliation of ancient writing monuments, when this cannot be done by means of classical linguistic analysis. In addition, knowing the area of distribution of a particular linguistic phenomenon, one can build the dialectological isoglosses[2] that are necessary for drawing up detailed linguistic maps.

This article presents a sequential description of the steps for working with extralinguistic information and isogloss maps in Lingvo-Doc.

2 HOW TO ADD EXTRALINGUISTIC INFORMATION INTO SOURCE

Work with sources on the LingvoDoc platform always begins with registration in the system or login to your account. After this step, the user will be able to work with the full linguistic base of the platform (in some cases with the prior permission of the author of a dictionary or corpus) and use all of its options.

Expanding a dictionary or corpus with the necessary information is carried out according to a single pattern, the only difference is the following:

1) filling in the appropriate fields of LingvoDoc when creating the resource[2] (Pic. 1);

2) entering extralinguistic information by editing the available dictionaries and corpora in the personal account (Pic. 2).
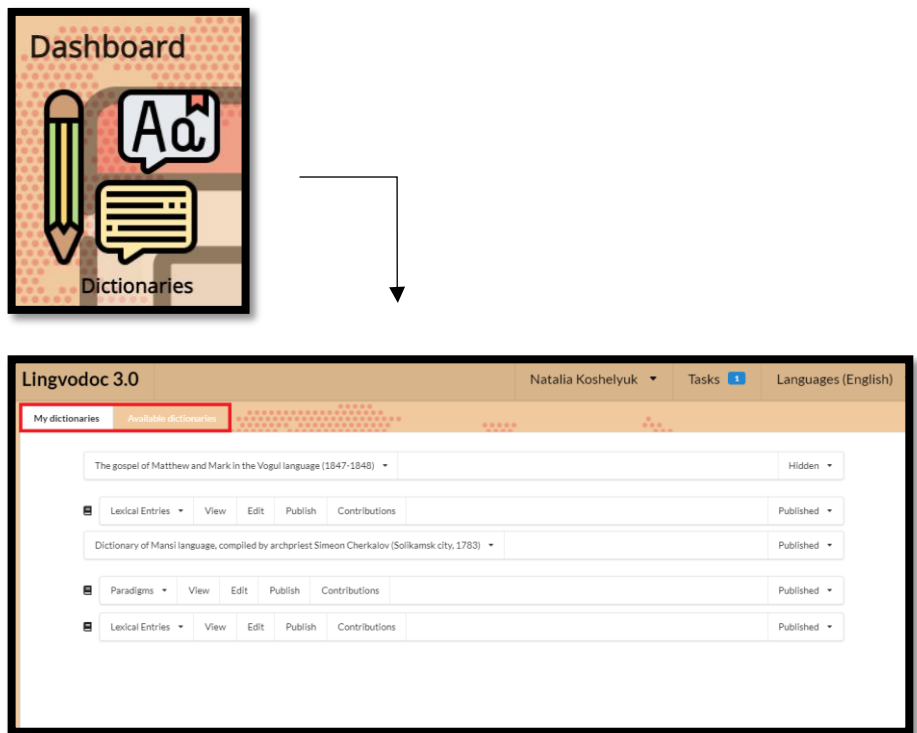
---

[1] Isogloss is a line on the map connecting places with the same linguistic features.

[1] Qv. preprint [Koshelyuk 2021].

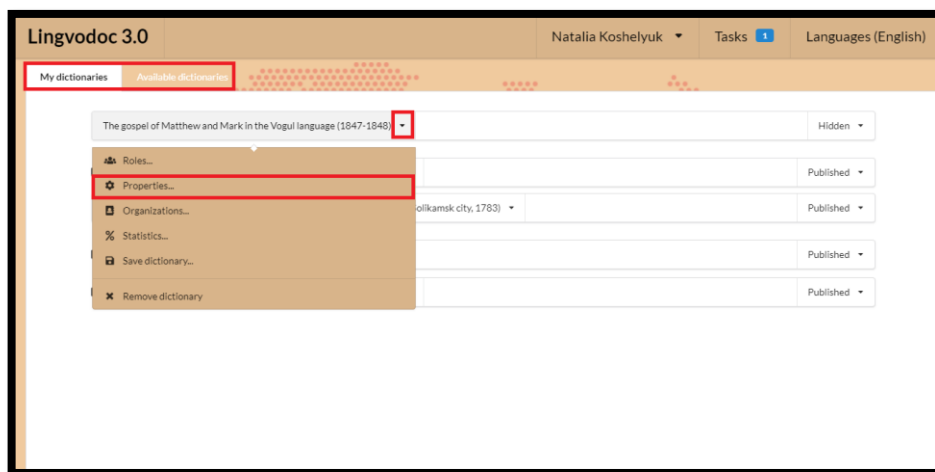*Pic. 1. Adding extralinguistic information when creating a resource*

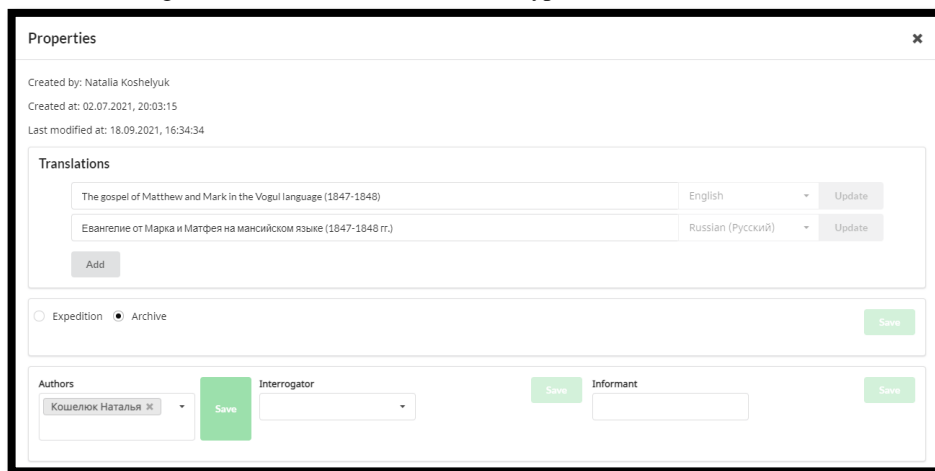*Pic. 2. Adding extralinguistic information by editing*

After the creation or selection of a desired resource, you need to go to the Properties tab, after which a field will open for filling in or

correcting additional information about the source (Pic. 3-5): title, authors, field or archive data, year of publication/collection of the material, genre, subject, localization, etc.
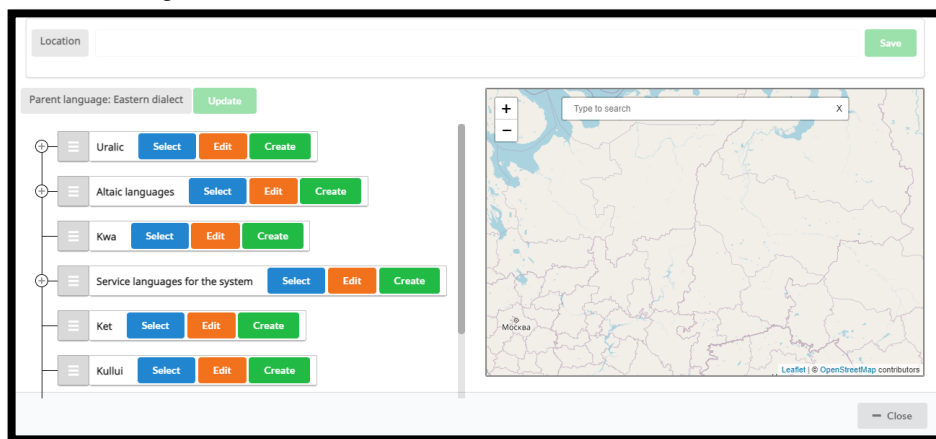
Pic. 3. Switching to fields with extralinguistic data

Pic. 4. Extralinguistic information. Title, author, type of source.

Pic. 5. Extralinguistic information. Localization

The place of collection of a field material or an archive source can be ascertained exactly by coordinates or conventionally by typing the name of a settlement or region in the search line.
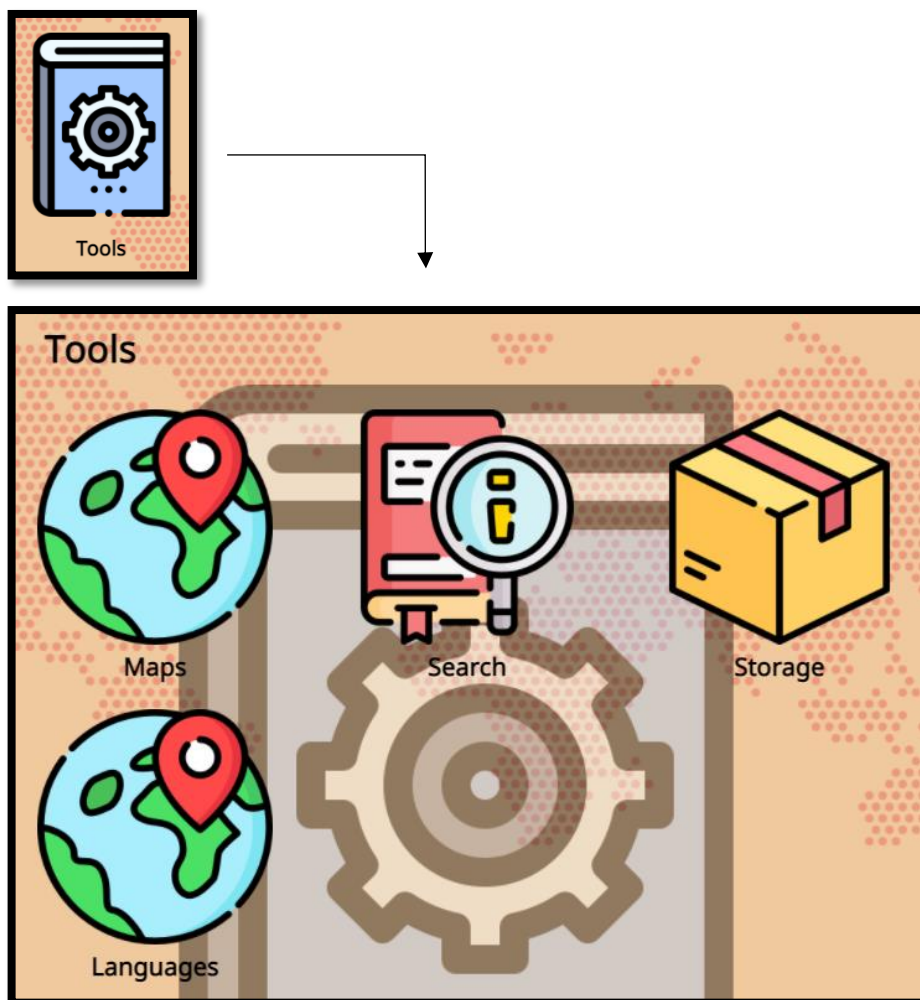
3 HOW TO WORK WITH MAPS

When working with maps in LingvoDoc, one can be offered a wide range of options:
✓ Maps
✓ Search
✓ Storage

In order to go to the desired option, you need to go to the Tools tab, after which all the icons will be displayed on the platform (Pic. 6).

*Pic. 6.* Tools. Maps, Search, Storage, Languages

Maps is an interactive map of Eurasia, which displays the entire set of loaded dictionaries and corpora for each language and dialect in dotted form (Pic. 7). By selecting the desired source from the suggested list, you can see the location of its collection in detail.
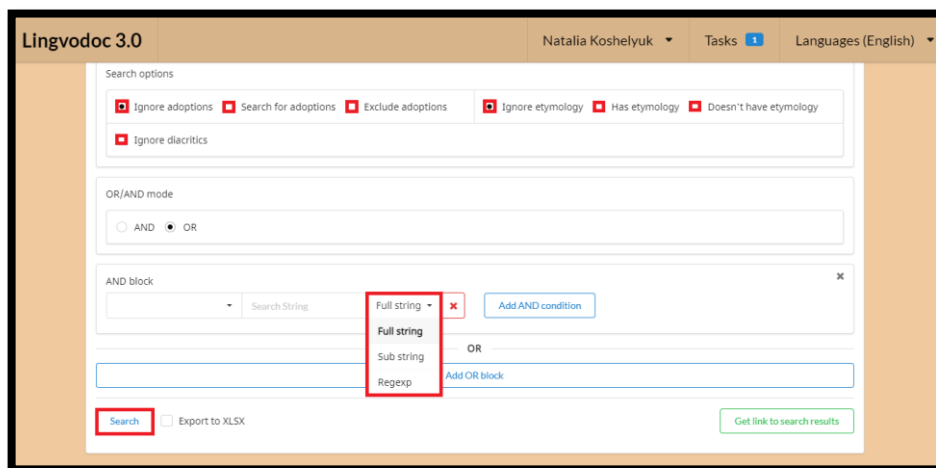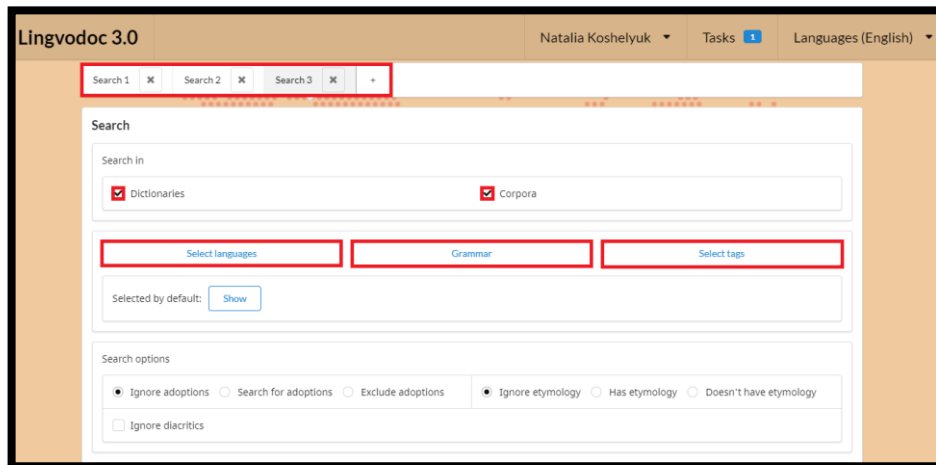
Pic. 7. Maps
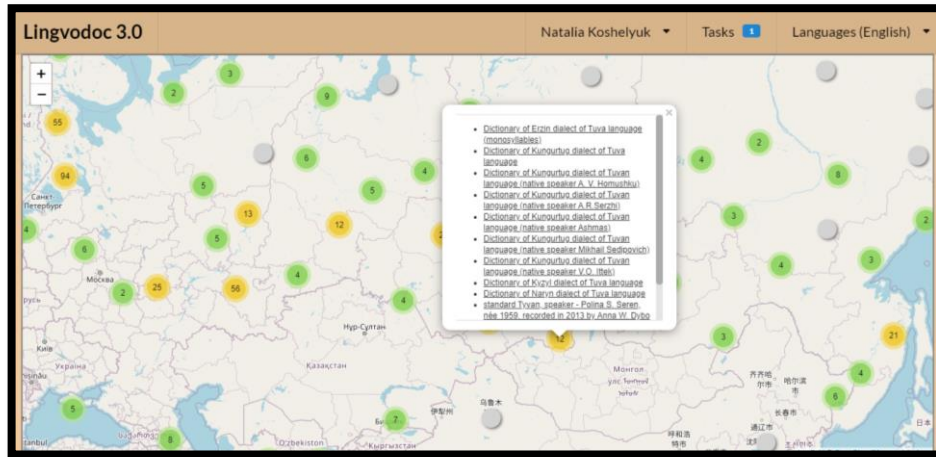
Search allows you to track various linguistic phenomena (grammatical, phonetic, semantic, etc.) both in individual Lingvodoc dictionaries and corpora, and throughout the entire database. Each search can be carried out in a separate tab and the desired parameters can be selected (Pic. 8).
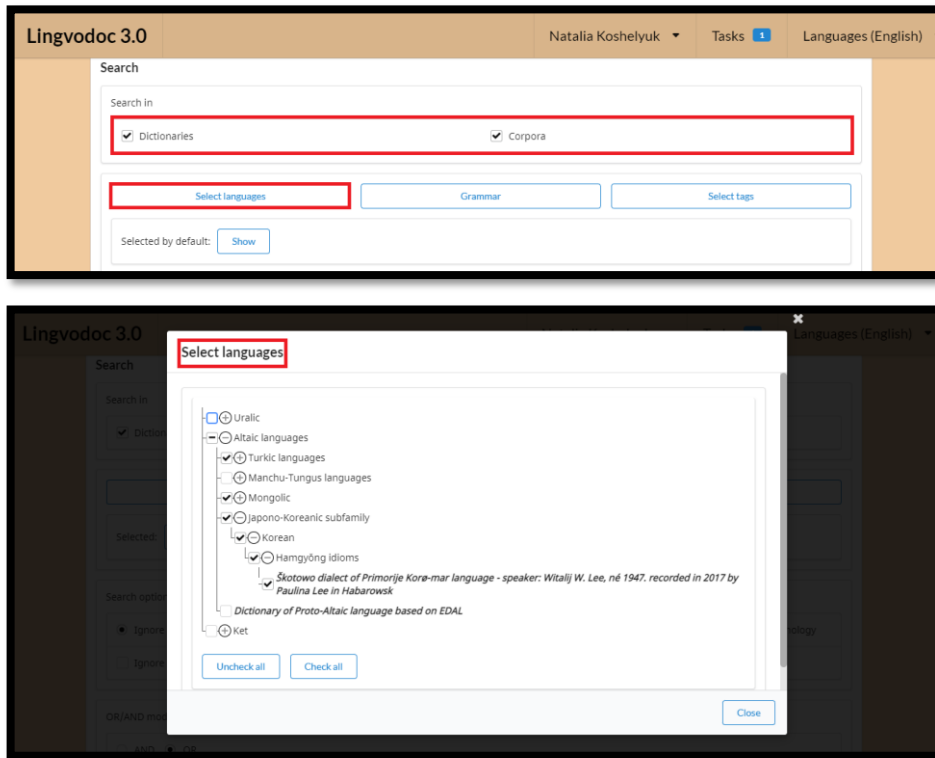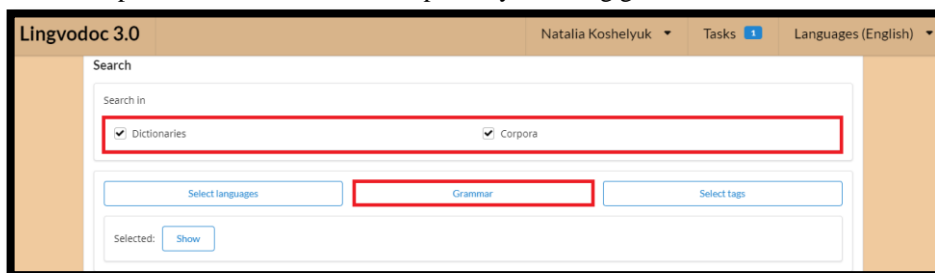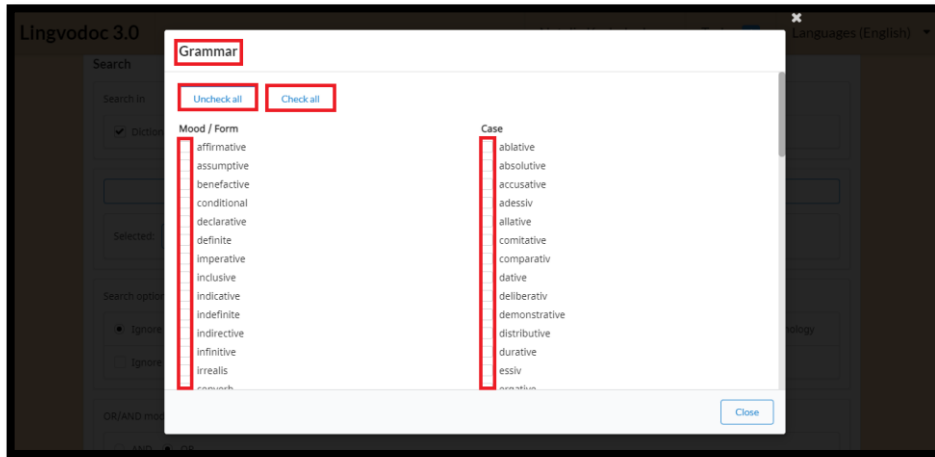
Pic. 8. Search
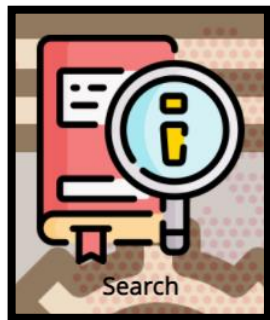
We display in more details how to carry out a search by individual languages, grammatical indicators and tags (Pic. 9-11).

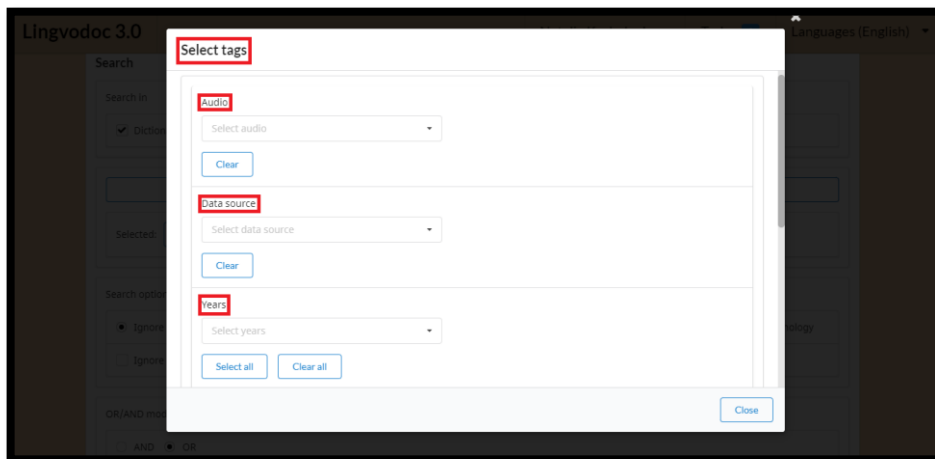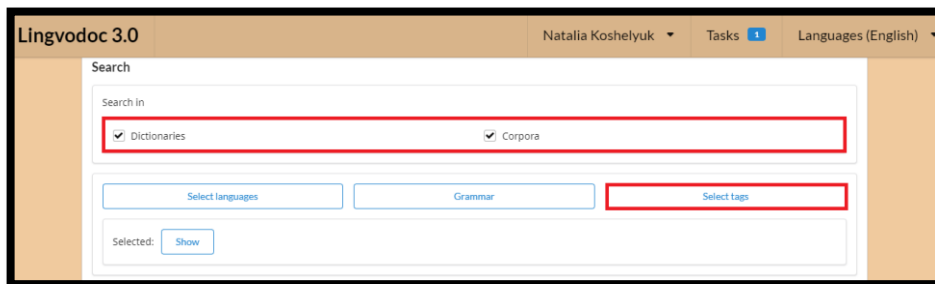*Pic. 9.* Implementation of the Search option by selecting language

*Pic. 10.* Implementation of the Search option by selecting grammar indicators

Pic. 11. Implementation of the Search option through tags

In addition, the Search option makes it possible to combine the data requested in Search into separate areas. A special color palette can
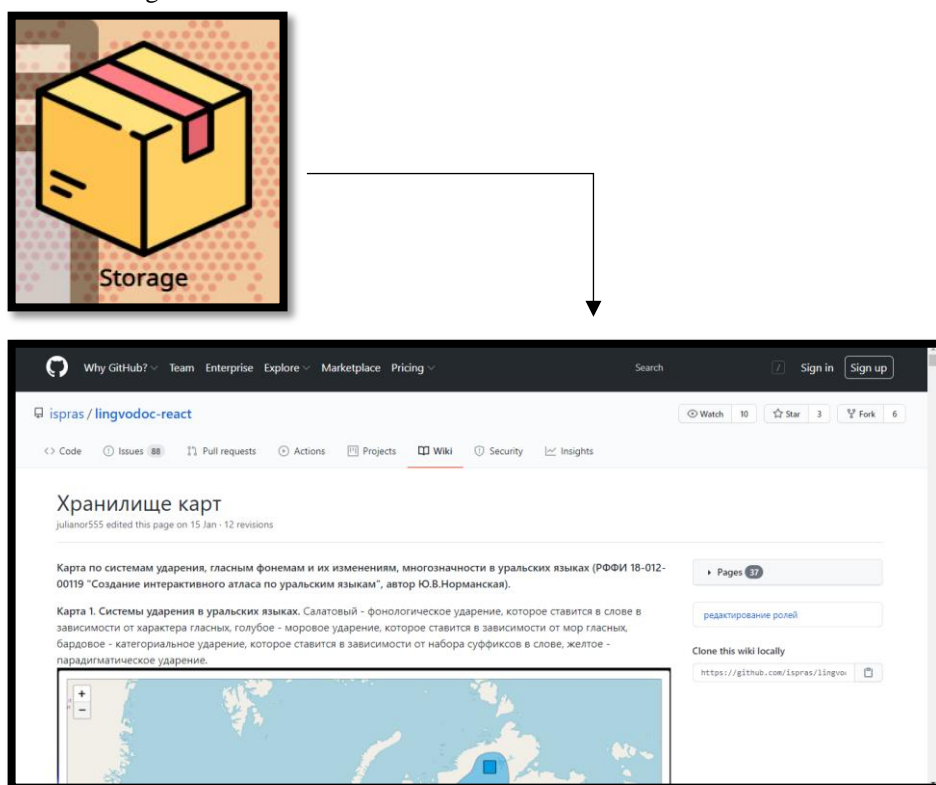
be selected for the borders of each of them (Pic. 12) and the received results can be converted into exe.

Pic. 12. Search. Geographical ranges
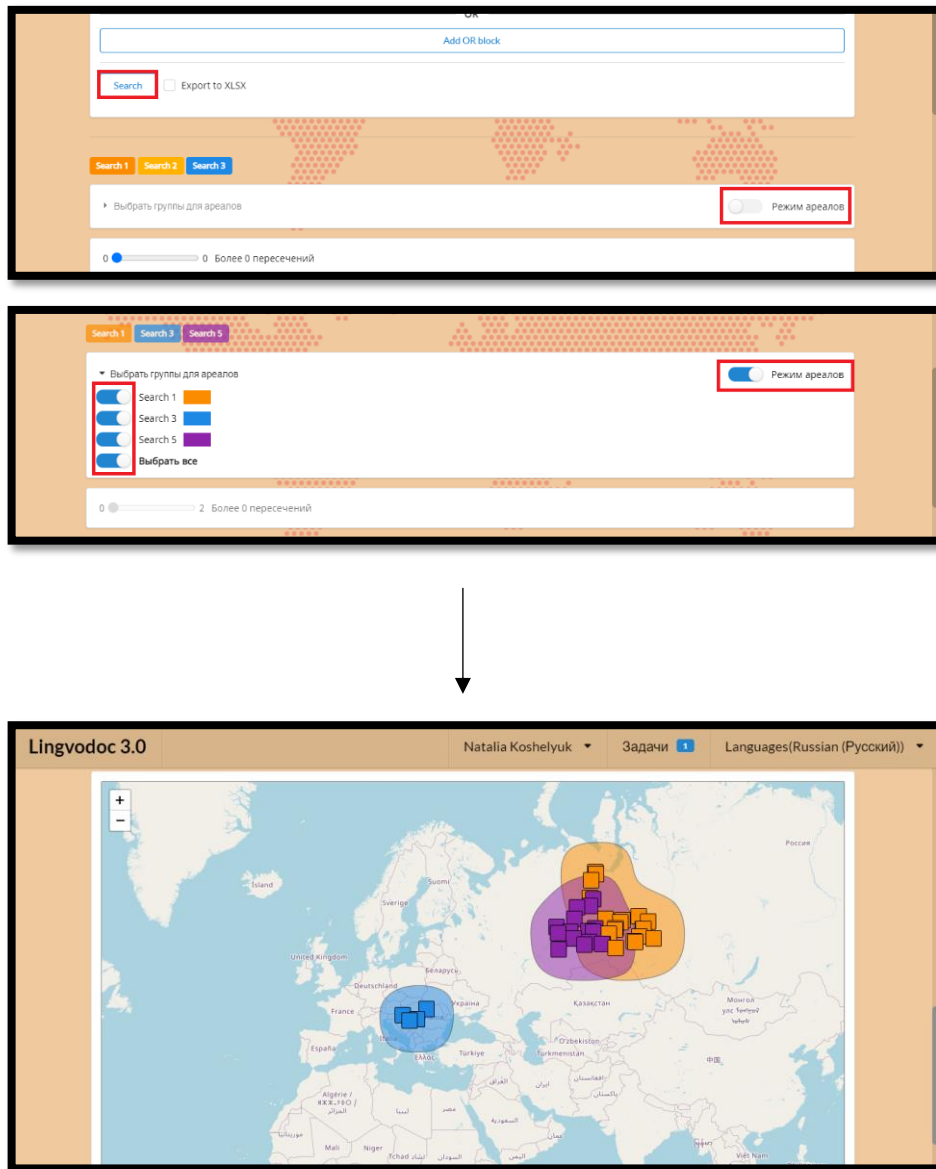
Storage is a collection of ready-made maps of various language features in various languages. At the moment, the collection consists of 46 maps in Russian (Pic. 13).

Pic. 13. Storage

Карта 1. Системы ударения в уральских языках. Салатовый - фонологическое ударение, которое ставится в слове в зависимости от характера гласных, голубое - моровое ударение, которое ставится в зависимости от мор гласных, бардовое - категориальное ударение, которое ставится в зависимости от набора суффиксов в слове, желтое - парадигматическое ударение.

- **Карта 6.8 *č в инлауте: Малин. *č > ʒ, *č > ž, Зелен. č > š, *č > ʃ, Жёлт. *č > t, *č > d, Голуб. č > s , č > z

- **Карта 6.9 *š в анлауте: Голубой *š > h, *š > x, Оранжевый *š > s , *š > z, Синий š > t, š > d

- Карта 3.7. э

Карты 4. Изменение прауральских гласных первого слога в современных уральских языках и диалектах. Карты 4.1 Изменения гласных первого слога в языках, расположенных в районе Уральских гор Карты 4.1.1. Огубление гласных

- *Карта 4.1.1.1 e > u*
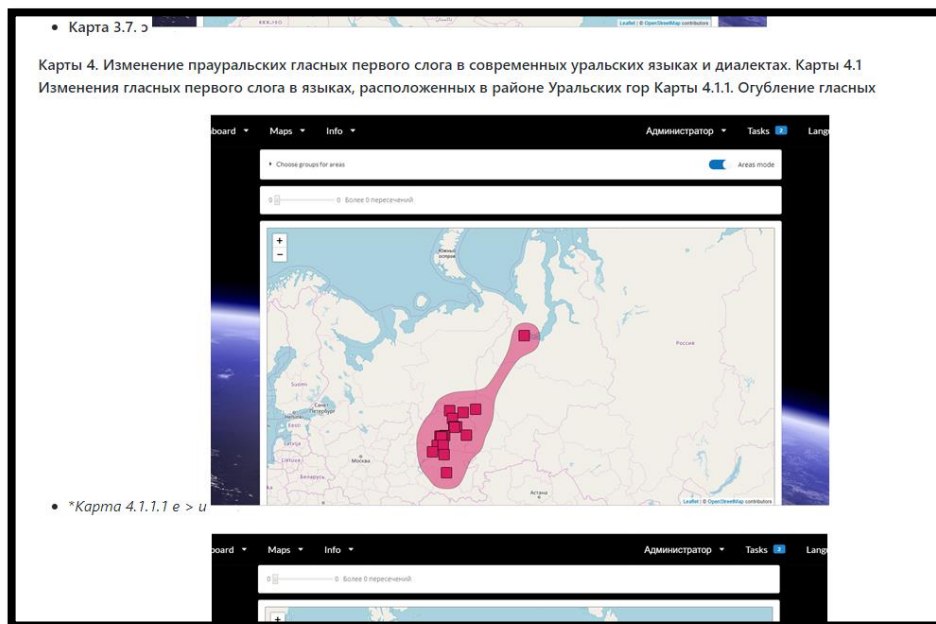
CONCLUSION

As can be seen from the overview, the search algorithms built into LingvoDoc can be used to analyze linguistic phenomena in terms of their geographical location, supplementing them with extralinguistic information, which in turn makes it possible to conduct research on the study of a linguistic material on a more meaningful and all-embracing level.

REFERENCES

1. LingvoDoc Homepage, http://lingvodoc.ispras.ru/. Last accessed 09/20/2021
2. Great Russian Encyclopedia, https://bigenc.ru/linguistics/text/2002556. Last accessed 09/20/2021