Electronic System «All-Ukrainian Toloka Archival Card Index»: Structure, Tools, Prospects of Development

Oksana Tyshchenko[a] and Vladyslav Tyshchenko[b]

[a] *Institute of Ukrainian Language of the National Academy of Sciences of Ukraine, M. Hrushevsky str. 4, Kyiv, 01001 Ukraine*
[b] *National Pedagogical Drahomanov University, Pyrohov str. 9, Kyiv, 02000, Ukraine*

Abstract

The article covered the principles and tools of collective recognition of manuscripts of the Archival Card Index (ACI) – lexical and phraseological materials of the Commission for compiling the Dictionary of the living Ukrainian language of the All-Ukrainian Academy of Sciences. In 2018, the Institute of the Ukrainian Language of the National Academy of Sciences of Ukraine created an electronic system «Archival Card Index» (ESACI) – digital format of ACI. ACI (350 thousand units) has got a great importance in the context of the cultural and national revival in Ukraine in the early 20th century, as it plays an important role in the development of the Ukrainian language, the theory and practice of Ukrainian studies in the 20th – early 21th century. The ACI fragment (3000 units) was recognized manually: the texts were entered into the ESACI according to the fields of the microstructure of the card. Such recognition requires considerable the effort and the time, so the platform «All-Ukrainian Toloka Archival Card Index» (AUTACI) has been created on the ACI website, which provides unlimited simultaneous online participation of volunteers for manual card recognition. Collective access to the collection of the transcribed documents is accompanied by instructions and samples of execution. The form for filling in the card is simplified in contrast to the form in the ESACI, as we plan to involve non-specialists in the work. Access to the AUTACI is possible after registration and has no time limits. In the future, we plan to use it to create tools for future verification of ACI texts, were automatically recognized by the Transkribus software, and for the partition linguistic information in the appropriate fields.

Keywords 1

Archival Card Index (ACI); Electronic System «Archival Card Index» (ESACI); open platform «All-Ukrainian Toloka Archival Card Index» (AUTACI); Ukrainian Lexicography; Manual Handwriting Texts Recognition; Lexicographic Toloka (Crowdsourcing)

1. Introduction

Crowdsourcing involves obtaining work, information, or opinions from a large group of people who submit their data via the Internet, social media, and Smartphone apps while tapping into people with different skills or thoughts from all over the world. Participants work on a paid or free basis as volunteers. Crowdsourcing is becoming a popular method to raise capital for special projects, taps into the shared interests of a group. It usually involves taking a large job and breaking it into many smaller jobs that a crowd of people can work on separately, usually sourced via the Internet, it contributes to save time and money [9].

Problems in the field of open innovation have been studied since the late 80's of the 20th century. This is especially determined by the sources of innovation and their dynamics [20], finding ways to solving problems, in particular, resolve them into local tasks [5; 8; 19]. Today the specialized interest

COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine
EMAIL: tom-73@ukr.net (O. Tyshchenko); vvtyshchenko@gmail.com (V. Tyshchenko)
ORCID: [0000-0002-5709-1252] (O. Tyshchenko); [0000-0003-2449-4722] (V. Tyshchenko)

of researchers of open innovations is concentrated mainly in the field of business, in particular in the marketing [6; 7]. These are the benefits of engaging consumers in the product development and support, and also related activities. The such technology systems can improve the customer experience and can help companies to improve their innovation and the customer's relationship management capabilities [12]. Such models of open platforms began to be used (and quite effectively) in the field of public administration [10].

Open platforms are also used in scientific research, particularly in philological research. Thus, in the OpenCorpora project, scientists try to involve native speakers who do not have special linguistic knowledge in the annotation works. To do this, a method of organizing processes to support of high quality annotations has been created [4]. Another example: the merger of synsets (synonymous set) to integrate two lexicographic resources (RussNet and YARN) was implemented by the authors, particularly, through a combination of experts and crowdsourcing approaches. The developers emphasize that the crowdsourcing methodology is a new and relevant area of researches in many areas [3]. Crowdsourcing is used by the Transkribus automatic text recognition system team to solve such problems. For faster processing of texts, collective access to the collection of transcribed documents is provided in the Transkribus Web interface – a lightweight, convenient and easy to use version of Transkribus [14]. In Ukraine collective contributions, mainly from students, primarily enrich the corpus resources: GRAK (General Regional Annotated Corpus of the Ukrainian Language) regularly attracts volunteers to increase and diversify the text, particularly, during the distance practical training in applied linguistics [13].

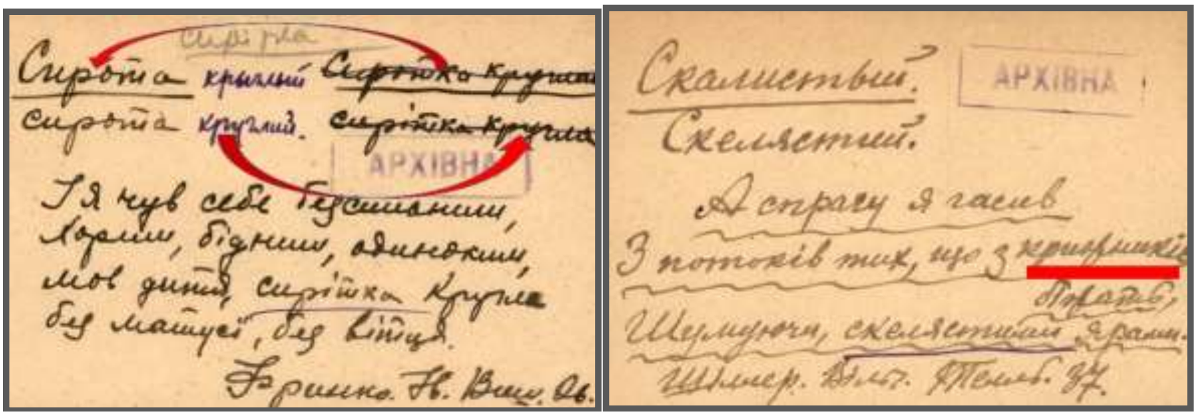
In the same way, everyone can join our common cause – the recognition of handwritten cards of the Archival Card Index (ACI). The stamp «ARCHIVAL», which marked the cards in the 50's of the 20th century, gave a conditional name to the Card Index itself. ACI was compiled in the 20s of the 20th century for the work of the Commission of the Dictionary of the Living Ukrainian Language Compiling of the All-Ukrainian Academy of Sciences. Several iconic Ukrainian dictionaries (first of all, «Russian Ukrainian Dictionary». V. I–III, Ch. edit A.Yu. Krymsky (I–II V.), S.O. Yefremov (III–IV V.), Kyiv 1924–1933) of the Golden Age of the Ukrainian lexicographic were compiled on the basis of this card index. In the 1930's, after the «purity» from «national junk» (the decline of Ukrainization and the beginning of mass repressions among all segments of the Ukrainian population, including the humanitarian sphere, from 1929 and, especially, in 1933-1937), the ACI was temporarily abandoned. In the 1950's it was combined with millions of cards of the new-created Lexical Card Index (hereinafter – LCI) of the O. Potebnya Institute of Linguistics of the UkSSR's Academy of Sciences. Later, the LCI (included the forgotten ACI) inherited the Institute of the Ukrainian Language of the National Academy of Sciences of Ukraine. ACI are materials with a lost and forgotten history. It should be updated in a modern field of question to find answers to common problems of restoring and strengthen of the Ukrainian language identity [17].

ACI contains two types of the cards: monolingual cards (with one title word – Ukrainian; it is 1/3 of the ACI) and bilingual cards (translated – Ukrainian-Russian, Russian-Ukrainian; 2/3 of ACI). These materials were prepared mainly for translated dictionaries. Headwords are usually accompanied by a quote and by a passport of the source. First, they are the working materials of the lexicographer, which shows the dynamics of scientific research. Secondly, the valuable linguistic facts recorded both in the title words and in the quotations, which can be explored from the point of view of both the individual and the collective linguistic creativity (Figure 1).

These are rare materials for linguistic researches in many areas, they need to be preserved and involved in modern linguistic processes. Therefore, it is logical to digitize the ACI and make it available on the Internet. The preparatory stage of ACI digitization became possible thanks to the Toloka (a toloka – ukr. *толока* 'crowdsourcing'): hundreds of volunteers joined the all-Ukrainian action «Preserve the Archival Lexical Card Index!» and manually processed about 6 million LC cards in order to choose 350 thousand cards with the «ARCHIVAL» stamp.

2. Scientific Novelty

In 2018, the Institute of the Ukrainian Language of the National Academy of Sciences of Ukraine created an Electronic System «Archival Card Index» (ESACI) – digital format of ACI (Figure 2).



сирітка

Сирота круглый ~~Сиротка круглая~~

сирота круглий. ~~сирітка кругла~~

АРХІВНА

Я чув себе безталанним, хорим, бідним, одиноким, мов дитя, сирітка кругла без матусі, без вітця.

Грінченко. Н. Вірш. Ол.

Скалистый.

Скелястий.

АРХІВНА

А спрагу я гасив З потоків тих, що з ~~кремнистих~~ Порогів, Шумуючи, скелястими ярами.

Шимер. Вил. Твор. 97.

Figure 1: Cards of Archival Card Index

A fragment of the ACI (about 3000 cards) has already been recognized, that is the texts of the cards have been manually transcribed. If the card has not recognized yet, it would only be viewed as an image. So, today the search for a given word for unrecognized cards (and this array of ACI predominates) is impossible (Figure 3).

The Card recognition is the entering of text in the appropriate fields that reflects the microstructure of the card, e.g., the headings and additional words or descriptive constructions: rus. *title*: ***свернуться,*** rus. additional: ***свернувшийся*** – ukr. *title*: ***скéплений*** (Figure 4). We could also sequentially record all corrections in the cards or later added items, e.g., to ***стемнеть*** added ***повечереть***. Such inserts we fix as a*dditional to the title word of the unit*, in this case – a synonym (in detail the structure of the ESACI is described in [15]).

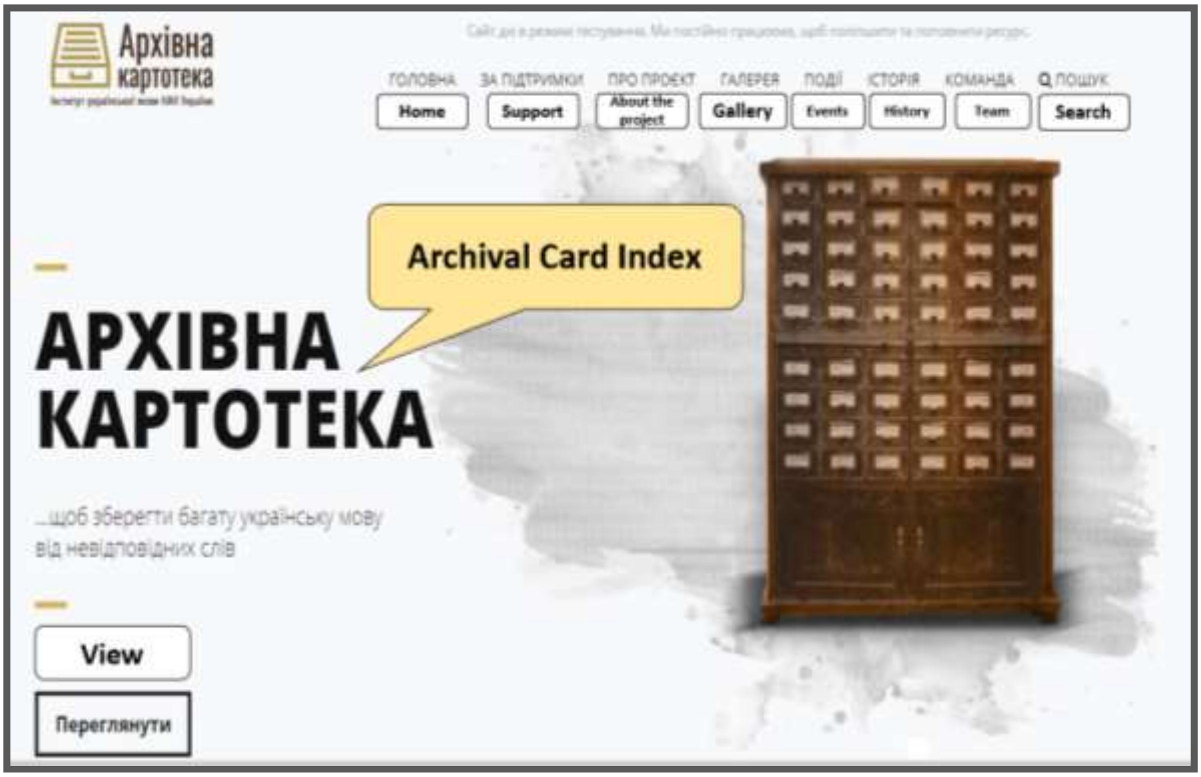


Figure 2: Electronic System «Archival Card Index»: web interface (2018–2021)

The manual recognition requires considerable effort and time. From 350,000 ACI cards, about 3,000 have been transcribed. The first fragment of the Archival Card Index was recognized within two months by four project participants in the Electronic System «Archival Card Index» (ESACI) in the offline mode, before the launch of the system «Toloka» [15]. Therefore, we see the point in using technologies that speed up this process. For example, there is a System of Automatic Text Recognition – Transkribus [14] (we have prepared the article «Synopsis: text, context, media» about machine recognition of handwritten text and preparation of archival data cards for it in «Ukrainian Language» journal (publication is expected in May, 2021). Of course, the description will need to check the recognition of automatic text and to enter information in the appropriate fields of the ACI system (to automate the work with the materials of the card index).

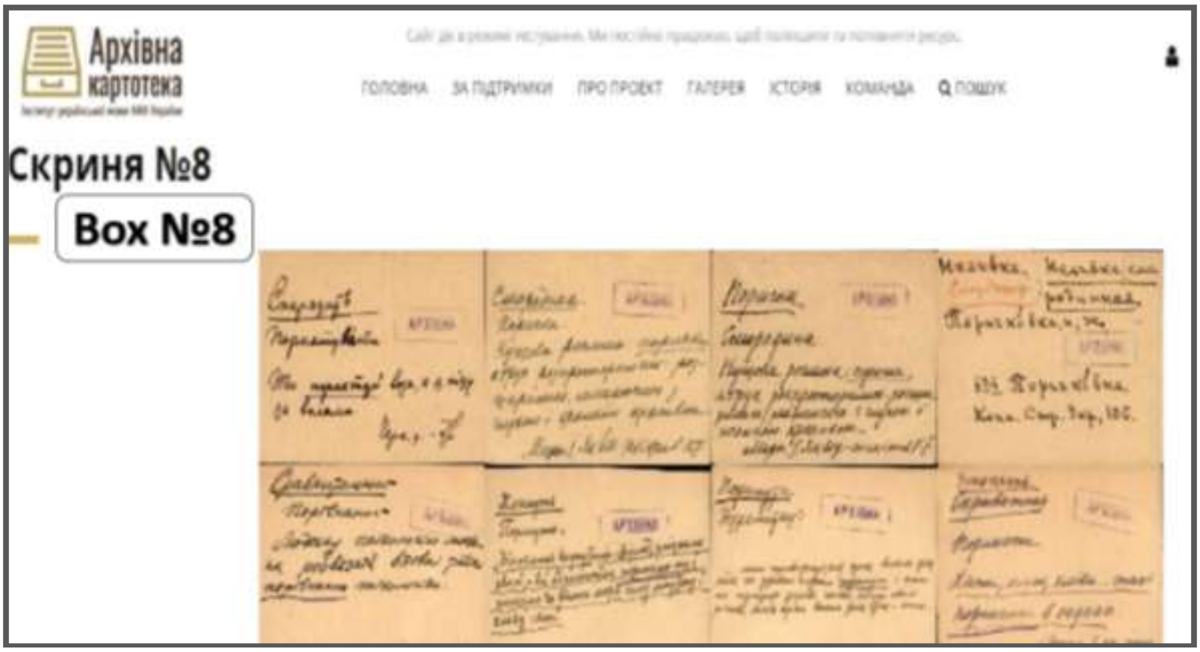
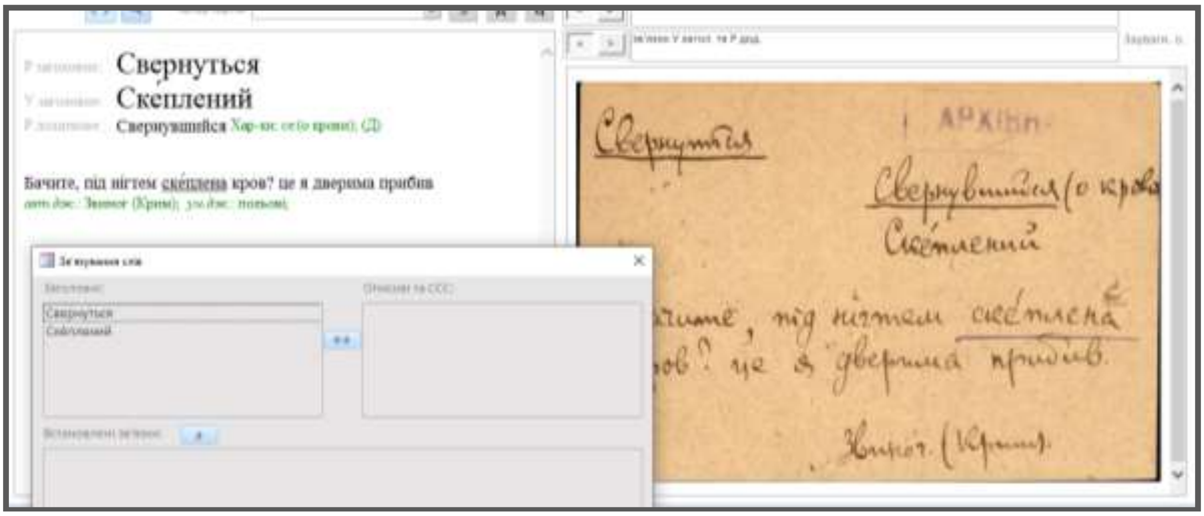


Figure 3: Box №8: the images of scanned ACI cards, which are contained in the same order in the box №8, the bookcase №1 on paper form



Свернуться

Скеплений

Свернувшийся Хар-кс ст(о крови); (Д)

Бачите, під нігтем скéплена кров? це я дверима прибив

Свернуться

Скеплений

Свернувшийся (о крови)

Скéплений

Бачите, під нігтем скéплена кров? це я дверима прибив.

Живот. (Крим).

Figure 4: The ACI Card, manually recognized in ESACI

We can speed up the correction of cards within the framework of our new project **«All-Ukrainian Toloka Archival Card Index» (AUTACI)** – an online platform on the ACI website for manual recognition of card texts by everyone interested [1] (Figure 5). Interested persons can register and take part in the manually recognition of card texts of the ACI within the Lexicographical Toloka. As already mentioned, the collective addition of volunteers to the affairs of the ACI – the selection and sorting of paper cards – has become traditional.

3. Basic Information about the «All-Ukrainian Toloka Archival Card Index» (AUTACI)

Consider the structure of AUTACI, its content, tools. Toloka is available at http://work.iul-nasu.org.ua/web/. The main page (Figure 5) contains a description of this toolkit: it states its purpose, gives instructions for work and prospects for the application of results. We present these aspects step by step.

ALL-UKRAINIAN TOLOKA.

What? Recognition of the Archival Lexical Card Index:

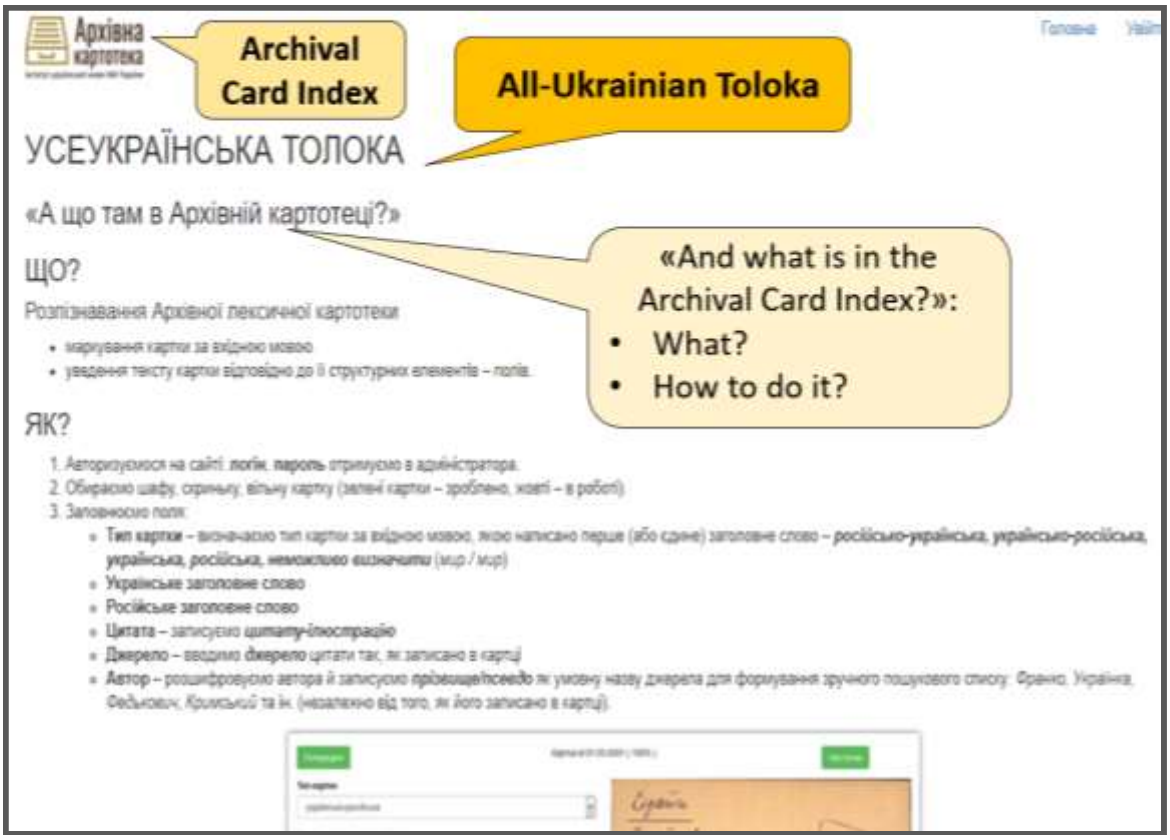
• card marking for the input language;

• entering the text of the card according to the fields (elements of the card microstructure).

How? There are six basic steps you need to take.

Step 1. Register on the site: login, password (received from the administrator, «Login» button).

Step 2. Select *a bookcase* (Figure 6), *a box* (Figure 7), *a card* (Figure 8). There will be two bookcases in total. The structure of the electronic file match to its real state in paper form.



УСЕУКРАЇНСЬКА ТОЛОКА

«А що там в Архівній картотеці?»

ЩО?

Розпізнавання Архівної лексичної картотеки

- маркування картки за вхідною мовою
- уведення тексту картки відповідно до її структурних елементів – полів.

ЯК?

1. Авторизуюся на сайті: **логін**, **пароль** отримую в адміністратора.
2. Обираю шафу, скриньку, вільну картку (зелені картки – зроблено, жовті – в роботі)
3. Заповнюю поля:
   - **Тип картки** – визначаю тип картки за вхідною мовою, якою написано перше (або єдине) заголовне слово – ***російсько-українська, українсько-російська, українська, російська, неможливо визначити*** (мур / мур)
   - **Українське заголовне слово**
   - **Російське заголовне слово**
   - **Цитата** – записую **цитату-ілюстрацію**
   - **Джерело** – вводимо **джерело** цитати так, як записано в картці
   - **Автор** – розшифровую автора й записую **прізвище/псевдо** як умовну назву джерела для формування зручного пошукового списку: Франко, Українка, Федькович, Кримський та ін. (незалежно від того, як його записано в картці).

Figure 5: Section «All-Ukrainian Toloka» on website ACI: The Manual

During operation, the downloaded cards receive the status displayed on the interface with color: *blank* (gray), *completed* (green), *uncompleted* (yellow), *in operation* (blue) (Figure 8). By selecting the desired card (usually *blank*), the operator recognizes it by filling in the specified fields. If during the transcription there are questions and the card needs to be finalized, the operator will select the *Difficult* check box and the card will receive the status *uncompleted*. If the transcription is successful, he will select *Done* → *Save* and the card will receive the status *completed*. Now the site has a simple **filter to go to the next card**: *next / previous (all in a row)*; *next / previous completed*; the *next / previous uncompleted* (Figure 9). Subsequently, we plan to distinguish between the transition to the next card for the administrator and the operator (see below).

Step 3. Enter the following **text** in the formula for recognizing ACI cards in the appropriate fields (Figure 9).

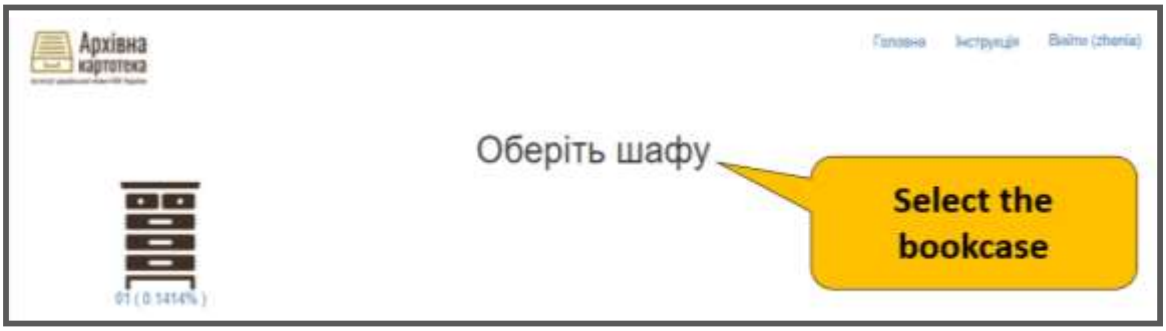


Figure 6: Select the bookcase

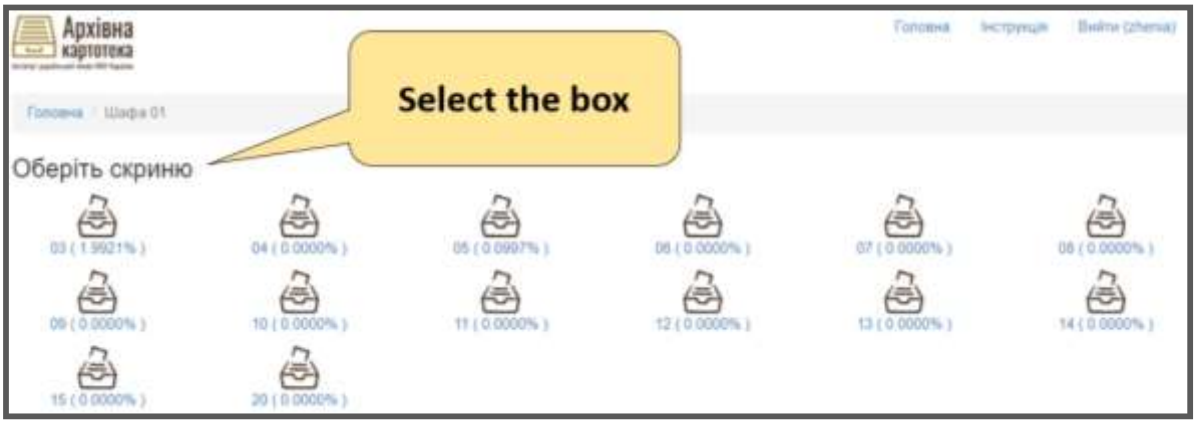
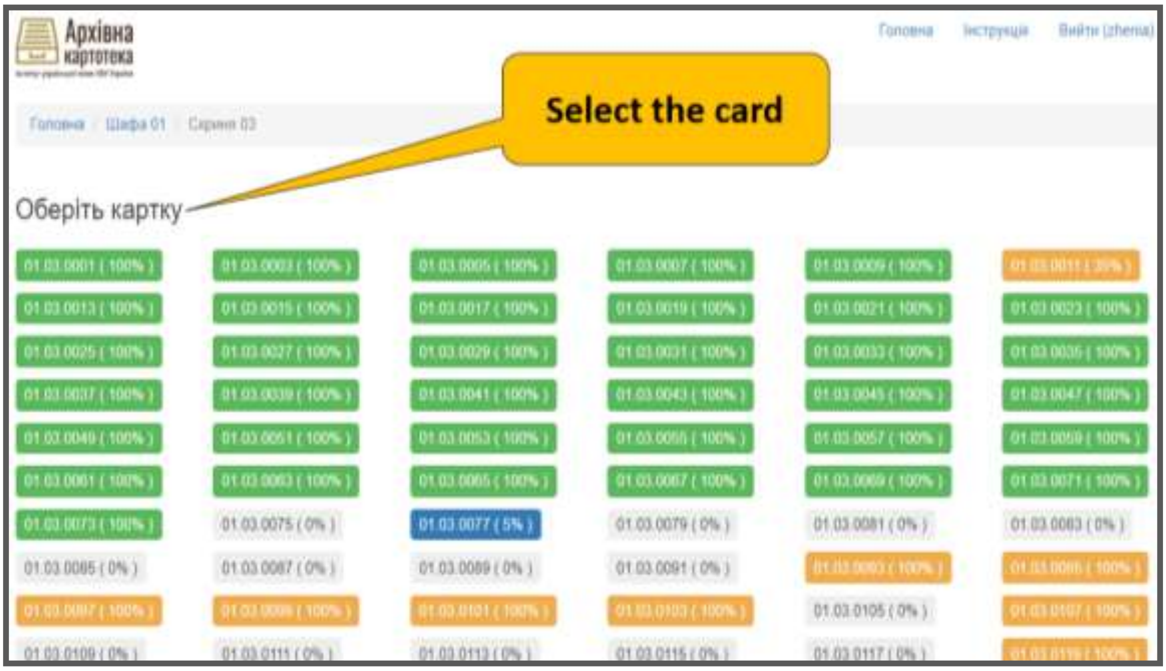


Figure 7: Select the box

- **Card type** – determine the existing input language that is written with the first (or only) title word. We distinguish the following **types**: *Russian-Ukrainian, Ukrainian-Russian, Ukrainian, Russian, it is impossible to determine* (e.g., in the case of *мир* / *мир*). In Figure 5 – *Ukrainian-Russian* card.
- **Ukrainian title word:** *Сідати.*
- **Russian title word:** *Садиться.*
- **Quote** – enter **the quote-illustration:** *Семен зрадів так прохав брата сідати....*
- **Source** – enter the source (passport) of the quote as written on the card: *Коцюб. І. 120. Ціпов.*
- **Author** – decipher the source and write the **name / pseudo / title of the publication** as a condition of the name of the source to form a convenient search list: *Франко, Українка, Федькович, Кримський* and others. (regardless of how it is written on the card). The author of the card in Figure 5 – *Коцюбинський.*
- We focus the operators' attention on the important conditions of the ACI cards transcription: **accurately and truthfully** reproduce the text of the card: with all elements, signs, abbreviations, as well as errors (if any), format (italics, underlines, strikethroughs and inserts, uppercase / lowercase). We remember the pre-reform elements in the Russian part of some cards: *і, ъ, ѣ: семь лѣтъ, Семилѣтній*, about the older Ukrainian spellings, dialectal variants: *лїс* instead of the modern normative *ліс, життє* instead of *життя*, etc. Technically, this feature is provided by a panel in each text input field, which contains special characters and means of a text editor (Figure 9).

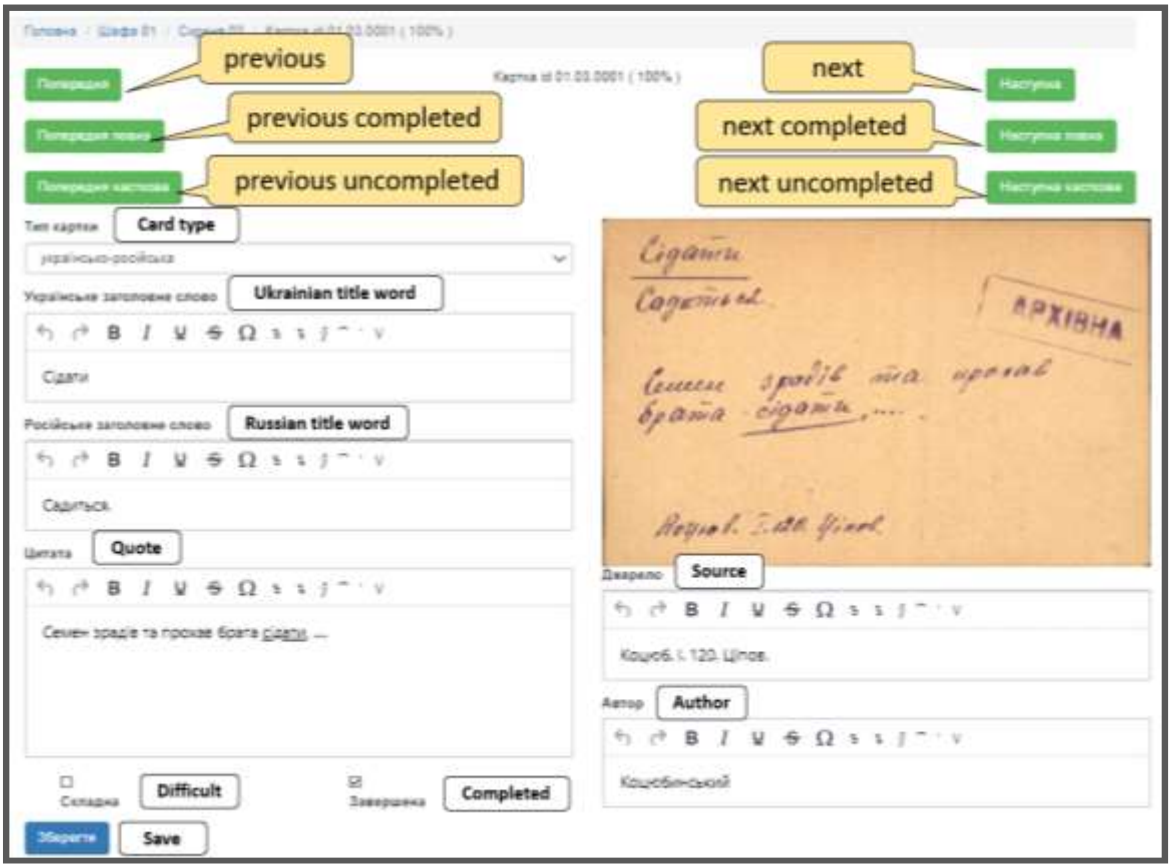
If you have got the difficulty decoding text, when it is heavily to understand what is written, then you should use the following tips:

- **CTRL+** – the card image can be enlarged.
- **DIDN'T HELP? Google** for help: you look for the title of the work by the author or quote, the author's name – by the quote, etc. It is possible to copy and paste the text in the appropriate field.
- **DIDN'T SAVE YOU?** Denote the unclear fragment by dots in double square brackets [[...]].
- **DO YOU DOUBT**, that everything was done correctly? Press the «Difficult» button – the card receives the status **uncompleted** (yellow), it will be checked and completed later.



01.03.0001 (100%)	01.03.0003 (100%)	01.03.0005 (100%)	01.03.0007 (100%)	01.03.0009 (100%)	01.03.0011 (35%)
01.03.0013 (100%)	01.03.0015 (100%)	01.03.0017 (100%)	01.03.0019 (100%)	01.03.0021 (100%)	01.03.0023 (100%)
01.03.0025 (100%)	01.03.0027 (100%)	01.03.0029 (100%)	01.03.0031 (100%)	01.03.0033 (100%)	01.03.0035 (100%)
01.03.0037 (100%)	01.03.0039 (100%)	01.03.0041 (100%)	01.03.0043 (100%)	01.03.0045 (100%)	01.03.0047 (100%)
01.03.0049 (100%)	01.03.0051 (100%)	01.03.0053 (100%)	01.03.0055 (100%)	01.03.0057 (100%)	01.03.0059 (100%)
01.03.0061 (100%)	01.03.0063 (100%)	01.03.0065 (100%)	01.03.0067 (100%)	01.03.0069 (100%)	01.03.0071 (100%)
01.03.0073 (100%)	01.03.0075 (0%)	01.03.0077 (5%)	01.03.0079 (0%)	01.03.0081 (0%)	01.03.0083 (0%)
01.03.0085 (0%)	01.03.0087 (0%)	01.03.0089 (0%)	01.03.0091 (0%)	01.03.0093 (100%)	01.03.0095 (100%)
01.03.0097 (100%)	01.03.0099 (100%)	01.03.0101 (100%)	01.03.0103 (100%)	01.03.0105 (0%)	01.03.0107 (100%)
01.03.0109 (0%)	01.03.0111 (0%)	01.03.0113 (0%)	01.03.0115 (0%)	01.03.0117 (0%)	01.03.0119 (100%)

Figure 8: Select the card: blank (gray), completed (green), uncompleted (yellow), in operation (blue)



Тип картки: українсько-російська

Українське заголовне слово: Сідати

Російське заголовне слово: Садиться

Цитата: Семен зрадів та просив брата сідати, ...

Джерело: Коцюб. І. 120. Цнов.

Автор: Коцюбинський

□ Складна ☑ Завершена

Figure 9: The form for ACI card recognition

Step 4. Completed – press when you are sure that everything was done correctly. The card receives the status **completed** (green).

Step 5. Save – save the recognized card.

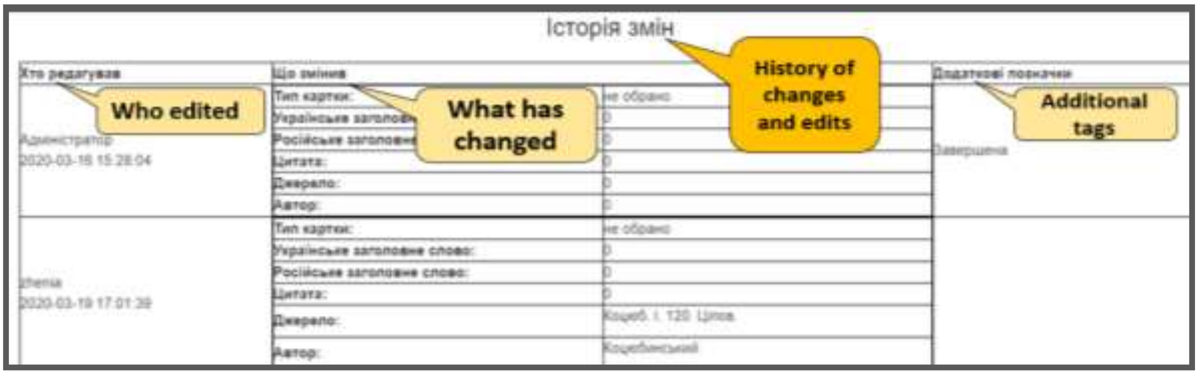
Step 6. Next – choose a new card: **next / previous** (all in a row); **next / previous completed; next / previous uncompleted.**

We remind you that the structure of the form for recognition in AUTACI is simpler than the structure of the card in ESACI. If the operator is occurred a very difficult card that contains a lot of different information not provided by the form (e.g., *additional words and sources*, *grammatical notes*, *more difficult structure*, *many edits*), he will skip it and goes to the next card, where everything is clear and simple. Difficult cards are for the next level of work in ESACI, more professional. In the AUTACI, they will retain the status of *blank*.

Access to the site is possible for two types of users: administrator and operator, respectively, with different levels of rights.

1. ***The administrator*** may: register operators; add cards; monitor the status of work performed – recognize and check the cards recognized by other participants, monitor statistics, history of changes in each card (Figures 10, 11, 12, 13).

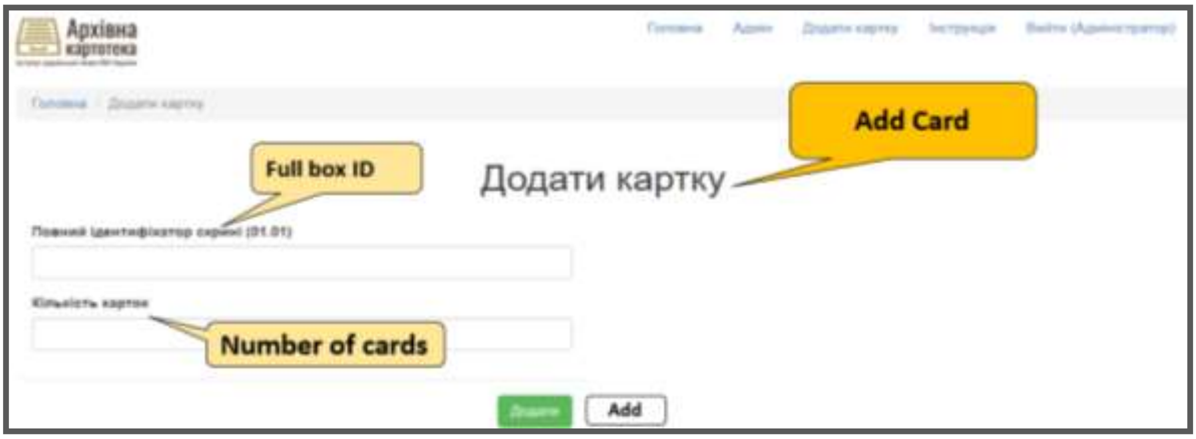
2. ***The operator*** may: enter information into the recognition form; check the work done, make changes to his cards; review the work done by other participants (Figures 6, 7, 8, 9).



Історія змін

Хто редагував	Що змінив		Додаткові позначки
Адміністратор 2020-03-18 15:28:04	Тип картки:	не обрано	Завершена
	Українське заголовне слово:	0	
	Російське заголовне слово:	0	
	Цитата:	0	
	Джерело:	0	
	Автор:	0	
zhenia 2020-03-19 17:01:39	Тип картки:	не обрано	
	Українське заголовне слово:	0	
	Російське заголовне слово:	0	
	Цитата:	0	
	Джерело:	Коцеб. І. 120 Цнов.	
	Автор:	Коцебинський	

Figure 10: History of changes



Додати картку

Повний ідентифікатор скриньки (01.01)

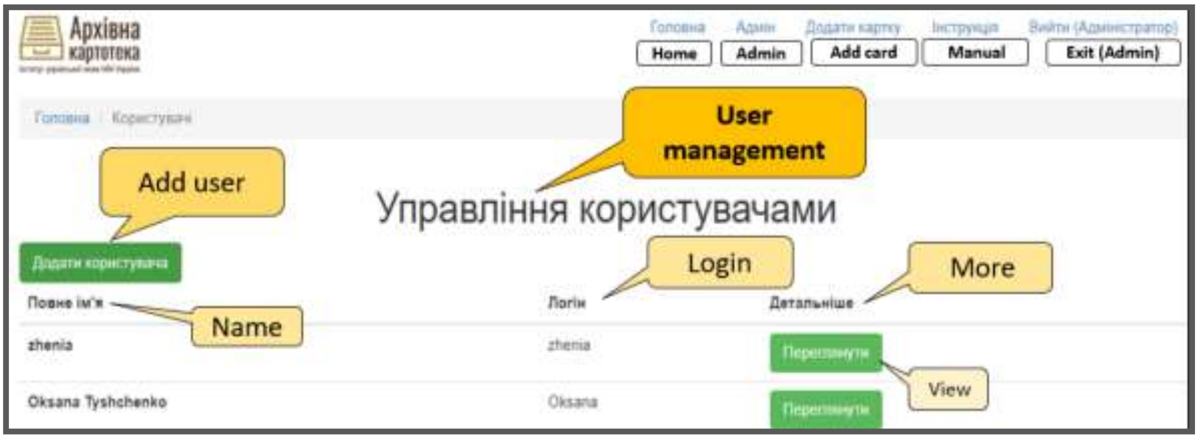
Кількість карток

Figure 11: Administrator’s excess: Add Card

4. Conclusion and Future Work

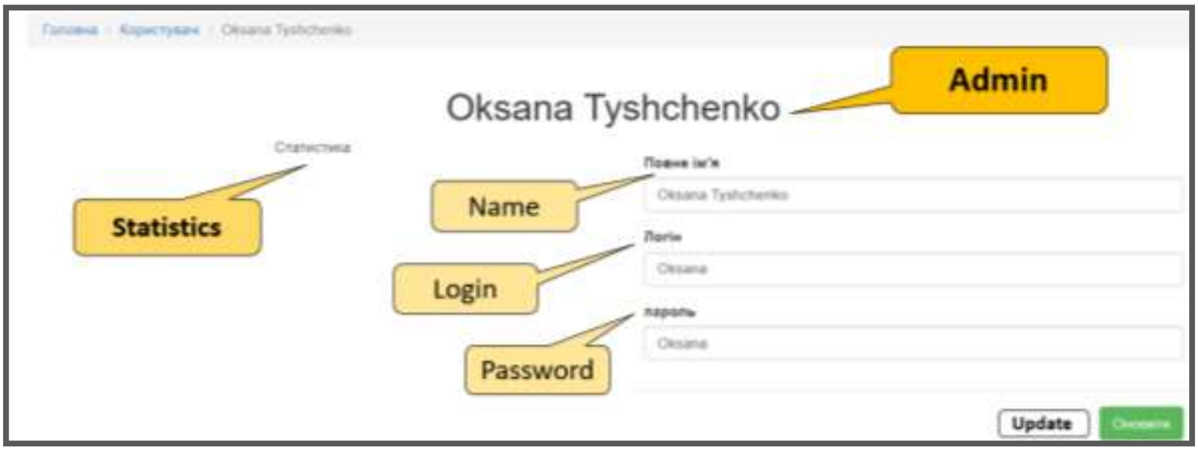
Subsequently, the interface will be improved. The pages of ESACI and Toloka (AUTACI) will have mutual hyperlinks. We will also improve the pass to the next page for the administrator and the operator to check the selected type of cards (completed and uncompleted) and general monitoring of the work. Also will be created new filters:

- ***for the administrator*** – ***transition*** with the right ***to view and make changes to the cards***: a) for *the operator*: Petrenko, Ivanenko, Sydorenko, etc. or all in a row; b) for *readiness*: completed; uncompleted; blank; all in a row.
- ***for the operators*** – ***transition*** with the right to ***work with cards*** with different degree of completion: a) to *view* the cards (*own / other operator*); b) to *make changes* to the cards (*only their own*).



Повне ім'я	Логін	Детальніше
zhenia	zhenia	Переглянути
Oksana Tyshchenko	Oksana	Переглянути

Figure 12: Administrator's excess: Users (Operators) Management



Повне ім'я: Oksana Tyshchenko

Логін: Oksana

пароль: Oksana

Figure 13: Administrator's excess: Statistics

There is no doubt that the trend towards «open innovation» has revived interest in using external sources of innovation. Different societies, institutions, firms purposefully open their models for connection of internal and external ideas, for joint creation of values with their partners and users. Internet platforms for Toloka (Crowdsourcing) and Co-creation have changed the way open innovation is introduced. They have provided new ways to work together to solve problems and create values. Toloka cause considerable interest in the community and gives real results [11].

What will happen from this? We emphasize that ACI and its component AUTACI are only a tool for studying the lexicographic values of the Archival Card Index. Creating a text version of the ACI will give everyone the opportunity to work with words, quotes, sources. It will let understand what language material could be fill the 4th repressed volume of the «Russian-Ukrainian Dictionary» 1924-1933, ed. A.Yu. Krymsky and S.O. Yefremov and other dictionaries written but not published up at that time. This capability will be achieved through tools such as:

- fast and convenient search;
- sorting for different filters (by *word*, *phrase*, *author*, *work* and many others);
- creation of a register (*Russian*, *Ukrainian*, translated *Russian / Ukrainian*, *Ukrainian / Russian*);
- review of *edits* in the cards and changes in the sources of language material, understanding of the dynamics of the lexicographer's thinking;
- observation about finding the necessary *match* to the word and much more.

This will enrich the tools and will expand the language base of linguistic research, will contribute to the creation of modern dictionaries, guides, grammars. In general, the ACI and the electronic resources were created for its processing will play an important role in renewal and preserving the identity of the Ukrainian language. Therefore, the actual scientific studies of ACI in linguistic and lingua-cultural optics are promising and important.

5. Acknowledgements

We thank the volunteers who joined the Toloka-2018 – choosing the cards of the Archival Card Index: students and teachers of the National Pedagogical Dragomanov University, Kyiv B. Grinchenko University, Kyiv National T. Shevchenko University, speech therapists and musicians, publishers and photographers, philologists and aircraft builders, teachers and school students from Kyiv, Lutsk, Kherson, Boryspil. We express our gratitude for the help to Prof. P. Hrytsenko and Prof. E. Karpilovska (Institute of the Ukrainian Language of the National Academy of Sciences of Ukraine), to Assoc. Prof. V. Starko (Ukrainian Catholic University).

The project Electronic System «Archival Card Index» (ESACI) – digital format of ACI was supported by the Ukrainian Cultural Foundation, grant № 1109, September – November 2018. The work was attended by: PhD in Philology O. Tyshchenko (project manager) and DSc in Philology L. Kysliuk (researchers of the Institute of Ukrainian Language NAS of Ukraine); PhD in Philology Yu. Pozdran (senior lecturer at Vinnytsia National Technical University) and Yu. Vozniuk (graduate student of the Institute of the Ukrainian Language of the NAS of Ukraine). The electronic system was implemented with the participation of programmers led by M. Tkachenko.

6. References

[1] All-Ukrainian Toloka: Archival Card Index, 2020, URL: http://work.iul-nasu.org.ua

[2] Archival Card Index, 2018–2021, URL: https://ak.iul-nasu.org.ua

[3] I. Azarova, P. Braslavsky, V. Zakharov, Yu. Kiselev, D. Ustalov, M. Khokhlova, RussNet and YARN. In: Structural and Applied Linguistics, vol. 12, St. Petersburg, 2019, pp. 34–52.

[4] V. Bocharov, S. Alexeeva, D. Granovsky, E. Protopopova, M. Stepanova, & A. Surikov. Crowdsourcing morphological annotation. In: Computational linguistics and intelligent technologies, Bekasovo, 2013, pp. 109–114, URL: http://www.dialog-21.ru/digests/dialog2013/materials/pdf/BocharovVV.pdf

[5] H. W. Chesbrough, (Ed.), Open innovation. The new imperative for creating and profiting from technology. Harvard Business School Press, Boston, 2006, 227 p.

[6] H. W. Chesbrough, W. Vanhaverbeke, J. West (Eds.) Open innovation: Researching a new paradigm, Oxford Univ. Press, Oxford, 2006, pp. 1–12. URL: http://scholar.google.com/scholar_lookup?&author=H.W..%20Chesbrough&pages=1-12&publication_year=2006.

[7] K. Brockhoff, Customers' perspectives of involvement in new product development. Int. J. Technology Management 5/6, 2003.

[8] H. Chesbrough, A better way to innovate. Harvard Business Review, vol. 81(7):12–3, Boston, 2003, 115 p.

[9] General Regionally Annotated Corpus of the Ukrainian Language (GRAC). M. Shvedova, R. Von Waldenfels, S.Yarygin, M. Kruk, A.Rysin, V. Starko, M.Wozniak, Kyiv– Oslo–Yen, 2017–2019, URL: https://www.uacorpus.org

[10] G. Koch, J. Füller, S. Brunswicker: Online Crowdsourcing in the Public Sector: How to Design Open Government Platforms. In: International Conference on Online Communities and Social Computing OCSC 2011: Online Communities and Social Computing, 2011, pp. 203–212, URL: https://link.springer.com/chapter/10.1007/978-3-642-21796-8_22

[11] H. Marshall, Crowdsourcing. Investopedia. 2019, URL: https://www.investopedia.com/terms/c/crowdsourcing.asp

[12] S. Nambisan, P. Nambisan: How to profit from a better virtual customer environment. In: MIT Sloan Management Review, vol. 49, 2008, pp. 53–61, URL: https://sloanreview.mit.edu/article/how-to-profit-from-a-better-virtual-customer-environment/ last accessed 2021/01/31

[13] Students of the Institute of Philology joined the development of the General regionally annotated corpus of the Ukrainian language. Institute of Philology of Kyiv B. Hrinchenko University. URL: https://if.kubg.edu.ua/prouniversitet/podii/1418-studenty-instytutu-filolohii-doluchylysia-do-rozrobky-heneralnoho-rehionalno-anotovanoho-korpusu-ukrainskoi-movy.html

[14] Transkribus, 2021, URL: https://readcoop.eu/transkribus/ Transkribus | Handwritten Text Recognition | READ COOP

[15] O. Tyshchenko, Archival card index of the Ukrainian language in digital format: from a language monument to modern lexicographic tools. In: Rocznik Slawistyczny. vol. LXIX, Wrocław, 2020, pp. 185–197

[16] O. Tyshchenko, Electronic lexical card index: the way to create modern vocabulary tools. In: Ukrainian language, 2, 2019, pp 37-52

[17] O. Tyshchenko, The archival card index as the lexical and illustrative base of «Russian-Ukrainian dictionary» ed. A. Krymsky and S. Yefremov part 1. Lexical card index: history of creation and repression; II. Micro- and macrostructure of archival lexical card index. In: Ukrainian language, 2, 2016, pp. 44–71; 3, 2016, pp. 57–78

[18] O. Tyshchenko, V. Tyshchenko, Metadata of the Linguistic Sourcesin Lexicographic Electronic Tool. In: Computational Linguistics and Intelligent Systems. Proceedingsof the 4th International Conferenceon Computational Linguistics and Intelligent Systems (COLINS). vol. I: Main Conference (Lviv, Ukraine, April 23–24), 2020, URL: http://ceur-ws.org/Vol-2604/paper24.pdf

[19] E. von Hippel, Sticky information and the locus of problem solving. Implications for innovation. In: Management Science, vol. 4, 1994, pp. 429–439

[20] E. von Hippel, The sources of innovation. Oxford University Press, New York, 1988, 221 p

[21] Material for the SANU Dictionary. Institute for Serbian Language SANU, 2018, URL: http://www.isj.sanu.ac.rs/2018/10/04/gradja-za-recnik-sanu-blago-koje-treba-sacuvati/