

**УКАРАНЕННЕ БЕЛАРУСКАМОЎНЫХ ТЭРМІНАЛАГІЧНЫХ
АДЗІНАК ТЭЗАЎРУСА AGROVOC У РАБОТУ АЎТАМАТЫЗАВАНАЙ
БІБЛІЯТЭЧНАЙ СІСТЭМЫ БЕЛАРУСКАЙ
СЕЛЬСКАГАСПАДАРЧАЙ БІБЛІЯТЭКІ**

**INTRODUCTION OF BELARUSIAN-LANGUAGE
TERMINOLOGICAL UNITS OF AGRICULTURAL TESAUROS
AGROVOC IN THE AUTOMATED LIBRARY SYSTEM
OF THE BELARUSIAN AGRICULTURAL LIBRARY**

Гарбачова Алена Ўладзімірайна – старэйшы навуковы супрацоўнік аддзела навуковага фармавання інфамацийных рэсурсаў, кандыдат сельскагаспадарчых навук, дацэнт, дзяржаўная ўстанова «Беларуская сельскагаспадарчая бібліятэка ім. І.С. Лупіновіча» Нацыянальнай акадэміі навук Беларусі (Рэспубліка Беларусь), e-mail: levigo@tut.by

Хальвіта Кацярына Станіславаўна – загадчык аддзела навуковага фармавання інфамацийных рэсурсаў, дзяржаўная ўстанова «Беларуская сельскагаспадарчая бібліятэка ім. І.С. Лупіновіча» Нацыянальнай акадэміі навук Беларусі (Рэспубліка Беларусь), e-mail: khalvita@belal.by

Варановіч Святлана Іванаўна – навуковы супрацоўнік аддзела навуковага фарміравання інфармацыйных рэсурсаў, дзяржаўная ўстанова «Беларуская сельскагаспадарчая бібліятэка ім. І.С. Лупіновіча» Нацыянальнай акадэміі навук Беларусі, (Рэспубліка Беларусь), e-mail: s.voropovich@belal.by

Дзямбоўская Жанна Віктараўна – навуковы супрацоўнік аддзела навуковага фарміравання інфармацыйных рэсурсаў, дзяржаўная ўстанова «Беларуская сельскагаспадарчая бібліятэка ім. І.С. Лупіновіча» Нацыянальнай акадэміі навук Беларусі, catalogue@belal.by

Самец Юлія Васільеўна – малады навуковы супрацоўнік аддзела навуковага фарміравання інфармацыйных рэсурсаў, дзяржаўная ўстанова «Беларуская сельскагаспадарчая бібліятэка ім. І.С. Лупіновіча» Нацыянальнай акадэміі навук Беларусі (Рэспубліка Беларусь), e-mail: catalogue@belal.by

Gorbachova Elena Vladimirovna – senior researcher of the Department of Scientific Formation of Information Resources, candidate of agricultural sciences, associate professor, State Institution «I.S. Lupinovich Belarusian Agricultural Library» of the National Academy of Sciences of Belarus (Belarus), e-mail: levigo@tut.by

Khalvita Ekaterina Stanislavovna – Head of the Department of Scientific Formation of Information Resources, State Institution «I.S. Lupinovich Belarusian Agricultural Library» of the National Academy of Sciences of Belarus (Belarus), e-mail: khalvita@belal.by

Voronovich Svetlana Ivanovna – researcher of the Department of Scientific Formation of Information Resources, State Institution «I.S. Lupinovich Belarusian Agricultural Library» of the National Academy of Sciences of Belarus (Belarus), e-mail: s.voronovich@belal.by

Dziambouskaya Zhanna Viktorovna – researcher of the Department of Scientific Formation of Information Resources, State Institution «I.S. Lupinovich Belarusian Agricultural Library» of the National Academy of Sciences of Belarus (Belarus), e-mail: catalogue@belal.by

Samets Yuliya Vasilievna – junior researcher, Department of Scientific Formation of Information Resources, State Institution «I.S. Lupinovich Belarusian Agricultural Library» of the National Academy of Sciences of Belarus (Belarus), e-mail: catalogue@belal.by

Анамацыя. У артыкуле выкладзены асаблівасці і асноўныя метадычныя падыходы да выкарыстання тэзаўруса AGROVOC у якасці лінгвістычнага інструмента пры аналітыка-сінтэтычнай апрацоўцы дакументаў у Беларускай сельскагаспадарчай бібліятэцы; разгледжана метадыка індэксавання дакументаў у бібліятэцы; адзначаюцца перспектывы перавагі выкарыстання ўказанага тэзаўруса для спецыялістаў, карыстальнікаў і прадстаўлення беларускамоўных дакументаў у базах даных аграрнай тэматыкі.

Abstract. The article outlines the features and main methodological approaches to using the AGROVOC thesaurus as a linguistic tool for the analytical and synthetic processing of documents in the Belarusian Agricultural Library; the methodology for indexing documents in the library is considered; the promising advantages of using the specified thesaurus for specialists, users and the presentation of Belarusian-language documents in agricultural databases are noted.

Ключавыя словы: тэзаўрус AGROVOC, інфармацыйныя рэсурсы, аналітыка-сінтэтычная апрацоўка дакументаў, інфармацыйна-пошукавая мова, электронны каталог.

Keywords: AGROVOC thesaurus, information resources, analytical and synthetic processing of documents, information retrieval language, electronic catalogue.

Найважнейшай функцыянальнай асаблівасцю тэзаўруса AGROVOC з'яўляецца магчымасць яго выкарыстання для індэксавання дакументаў і інфармацыйных (бібліяграфічных) рэсурсаў, якое, у сваю чаргу, уяўляе сабой базавую аперацыю аналітыка-сінтэтычнай апрацоўкі дакументаў для мэт інфармацыйнага пошуку. Паспяхова рэалізацыя мэт аналітыка-сінтэтычнай апрацоўкі дакументаў залежыць ад выбару інфармацыйна-пошукавай мовы (ШМ), якая ўключае ў сябе тэрміны індэксавання.

З'яўляючыся асноўнай ІІМ Міжнароднай інфармацыйнай сістэмы па сельскагаспадарчай навуцы і тэхналогіях AGRIS, тэзаўрус AGROVOC быў створаны і вядзецца ў мэтах адназначнага прадстаўлення паняццяў індэксуемых дакументаў і вырашэння праблемы неадназначнасці выказвання сэнсу пры дапамозе натуральнай мовы пры інфармацыйным пошуку.

У бягучы перыяд AGROVOC змяшчае больш за 40 900 канцэптаў і больш за 986 000 тэрмінаў, даступных на 42 мовах [1]. Адметнай якасцю тэзаўруса AGROVOC з'яўляецца забеспячэнне рэlevantнага тэматычнага пошуку ў міжнароднай базе дадзеных AGRIS на любой з прадстаўленых у ім моў, а таксама сэнсавая эквівалентнасць і спалучанасць лексічных адзінак нацыянальных версій. Ён знайшоў шырокае ўжыванне для індэксавання, пошуку і арганізацыі дадзеных у сельскагаспадарчых інфармацыйных сістэмах, на вэб-старонках і актыўна выкарыстоўваецца для гэтых мэт спецыялістамі [2].

Варта адзначыць, што стварэнне беларускамоўнага сегмента ў тэзаўрусе AGROVOC і выкарыстанне яго пры аналітыка-сінтэтычнай апрацоўцы нацыянальных дакументаў і выданняў на аснове стандартызаванай навуковай тэрміналогіі па праблематыцы АПК дазваляе пашырыць магчымасці доступу да беларускага кантэнту ў AGRIS не толькі айчынных, але і замежных даследчыкаў у задавальненні адпаведных інфармацыйных запытаў.

Аналітыка-сінтэтычная апрацоўка дакументаў уключае выкананне аперацыі іх індэксавання, якая ўяўляе сабой выражэнне зместу дакумента і/ці сэнсу інфармацыйнага запыту на інфармацыйна-пошукавай мове [3].

У сваю чаргу, інфармацыйна-пошукавая мова – гэта фармалізаваная штучная мова, якая прызначана для індэксавання дакументаў, інфармацыйных запытаў і апісання фактаў з мэтай далейшага захоўвання і пошуку.

Інфармацыйна-пошукавыя мовы адрозніваюцца ў залежнасці ад вядучага прынцыпа індэксавання. Пры гэтым яны могуць быць класіфікацыйныя, прадметызацыйныя і дэскрыптарныя інфармацыйна-пошукавыя. У сваю чаргу гэта вызначае спосаб індэксавання:

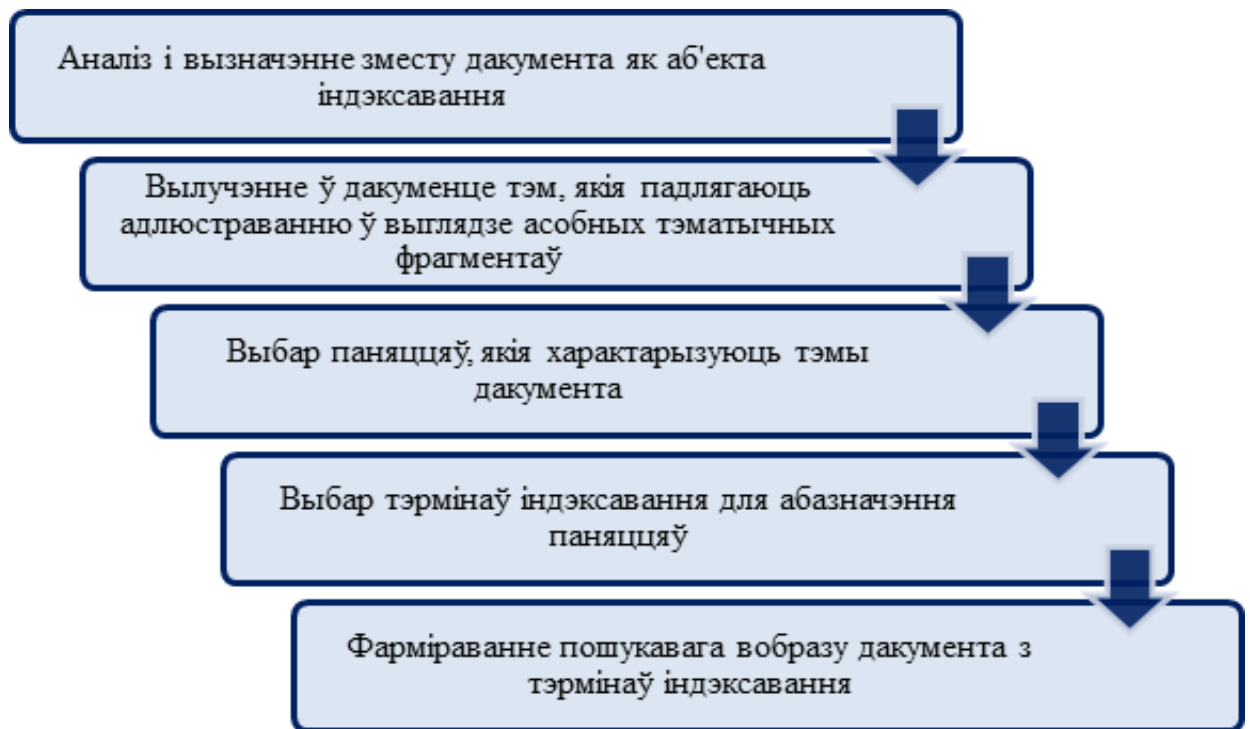
- прадметызацыя, калі змест дакумента і (або) запыту можа быць выражаны прадметнай рубрыкай у адпаведнасці з правіламі прадметызацыйнай ІІМ;
- сістэматызацыя, якая прадугледжвае выражэнне зместу дакумента і (або) запыту класіфікацыйнымі індэксамі ў адпаведнасці з правіламі якой-небудзь класіфікацыйнай ІІМ (класіфікацыйнай сістэмы);
- каардынатнае індэксаванне, калі выконваецца шматаспектны выраз асноўнага сэнсавага зместу дакумента і (або) інфармацыйнага запыту мноствам ключавых слоў ці дэскрыптараў на аснове прызначанай дэскрыптарнай інфармацыйна-пошукавай мовы.

Абраны спосаб прадугледжвае выкарыстанне адпаведных тэрмінаў індэксавання: прадметныя рубрыкі, класіфікацыйныя індэксы

(на класіфікацыйнай ІПМ), дэскрыптары і ключавыя словы (на дэскрыптарнай ІПМ).

Варта адзначыць, што калі індэксаванне дакументаў адбываецца з прымяненнем тэзаўруса, неабходна прытрымлівацца пэўнай метадыкі, з дапамогай якой можна адэкватна і цалкам раскрыць змест дакументаў і тым самым павысіць магчымасці і эфектыўнасць пошуку дакументаў інфармацыйна-пошукавай сістэме.

Працэс індэксавання складаецца з шэрагу паслядоўна выкананых этапаў, кожны з якіх мае важнае значэнне (Малюнак 1).



Малюнак 1 – Этапы працэса індэксавання дакументаў

Практыка індэксавання паказвае, што выкананне першых чатырох этапаў адбываецца практычна адначасова. Таму, па сутнасці, спачатку павінна быць выяўлена ўся асноўная інфармацыя, якая змешчана ў дакуменце і якую патрэбна адлюстраваць ключавымі словамі, а потым, на заключным этапе, трэба сфарміраваць пашуковы вобраз дакумента. Для гэтага выконваецца замена ключавых слоў дэскрыптарамі тэзаўруса з улікам устаноўленых у ім сувязей.

Пры аналізе зместу дакумента неабходна ўлічваць усе яго элементы, якія дазваляюць усталяваць змест, накіраванасць, форму і прызначэнне. Пры гэтым трэба прааналізаваць і зрабіць ацэнку ўсіх складовых частак дакумента, уключаючы загаловак, анатацыю, рэферат, уводзіны, змест, заключэнне, а так сама ілюстрацыі, табліцы і г.д.

Змест дакументаў аналізуецца па шэрагу аспектаў:

- прадмет і тэма даследавання;
- бок, з якога аналізуецца (даследуецца, апісваецца) прадмет або яго ўласцівасці, прыкметы і заканамернасці (прыродныя, фізічныя, біялагічныя, хімічныя і інш.);
- вобласць ужывання або выкарыстання прадмета;
- від даследавання ўласцівасцяў прадмета;
- канкрэтны мета даследавання;
- методыка правядзення даследавання і спецыяльнае абсталяванне, якое выкарыстоўваецца для вывучэння прадмета і даследавання яго ўласцівасцяў;
- умовы, у якіх праведзены даследаванні ўласцівасцей прадмета.

Этап фарміравання пашуковага вобразу дакумента ўключае дзве складовыя часткі. Першая з іх – гэта дакладнае і найбольш поўнае адлюстраванне галоўных семантычных аспектаў дакумента ключавымі словамі. Другая частка заключаецца ў замене ключавых слоў дэскрыптарамі тэзаўруса, якія найбольш дакладна суадносяцца з ключавымі словамі, адлюстроўваюць паняцці, з мэтай мінімізацыі скажэння інфармацыі, змешчанай у дакуменце.

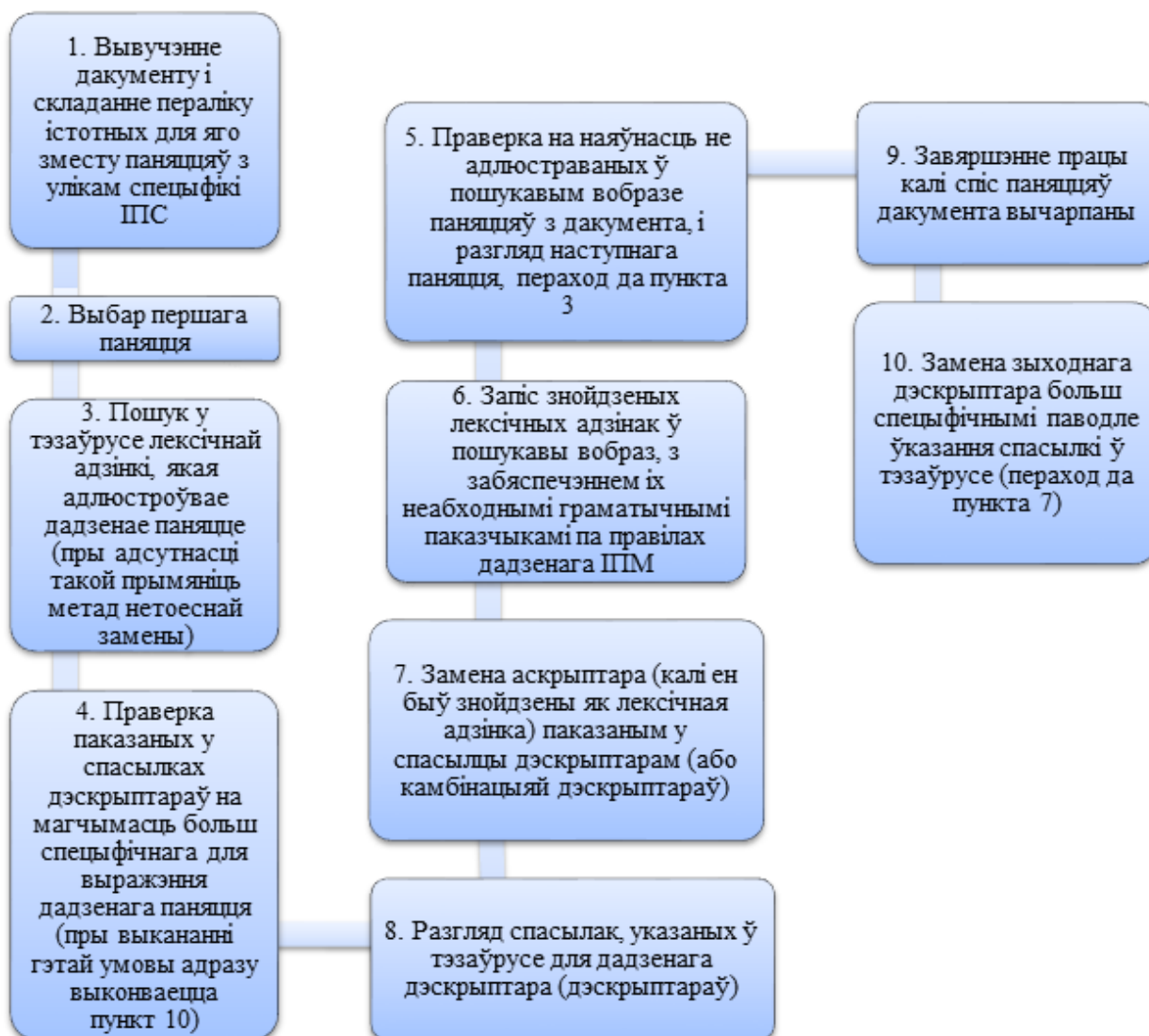
Пры замяшчэнні ключавых слоў дэскрыптарамі тэзаўруса спачатку выконваецца тоесны замен, калі падбіраецца дэскрыптар, які супадае па назве і зместу з ідэнтычным ключавым словам. Пры немагчымасці дакладнага адлюстравання зместу дакументаў дапускаецца нетоесная замена ключавых слоў дэскрыптарамі. Такая замена прыводзіць к частковай страце сэнсу, скажэнню інфармацыі. У гэтым выпадку пошук найбольш прыдатнага дэскрыптара пачынаюць з вызначэння дэскрыптарнай вобласці, потым дэскрыптарнай групы, а толькі потым – вылучаюць дэскрыптар, які па сэнсе найбольш блізкі ключавому слову.

У цэлым алгарытм індэксавання з выкарыстаннем тэзаўруса мае наступны выгляд (Малюнак 2).

Практыка аналітыка-сінтэтычнай апрацоўкі дакументаў у дзяржаўнай установе «Беларуская сельскагаспадарчая бібліятэка ім. І.С. Лупіновіча» Нацыянальнай акадэміі навук Беларусі (далей – БелСГБ) прадугледжвае выкарыстанне такіх лінгвістычных інструментаў, як:

- БД поўных табліц УДК;
- галіновыя працоўныя табліцы УДК для аграпрамысловага комплексу Цэнтральнай навуковай сельскагаспадарчай бібліятэкі (ЦНСГБ), алфавітна-прадметны паказальнік да іх;
- электронны інфармацыйна-пошукавы тэзаўрус па сельскай гаспадарцы і харчаванні (ІПТ) ЦНСГБ (Масква).

Складанне анатацыі да артыкула рэгламентуе ДАСТ 7.9-95 (ИСО 214-76) «Система стандартов по информации, библиотечному и издательскому делу. Реферат и аннотация. Общие требования».



Малюнак 2 – Алгарытм індэксавання па інфармацыйна-пошукавым тэзаўрусе

Навуковая апрацоўка дакументаў у БелСГБ выконваецца з выкарыстаннем сістэмы аўтаматызацыі бібліятэк (САБ) «ИРБИС64+» АРМ «Каталогизатор». Пры гэтым аналітыка-сінтэтычная апрацоўка дакумента ажыццяўляецца шляхам запаўнення асноўных палёў у закладцы «Систематизация» АРМ «Каталогизатор» САБ «ИРБИС64+»:

- поле 675 – «индексы УДК»;
- поле 610 – «ненормированные ключевые слова» (праз паўтаральнае поле);
- поле 331 – «аннотация».

Цяпер у электронным каталогу «Belal» БелСГБ прадстаўлена больш за 800 назваў кніг, а таксама 21 назва часопісаў і каля 300 навуковых артыкулаў на беларускай мове. Пры іх індэксаванні выкарыстоўваўся тэзурус ФДБНУ ЦНСГБ, а адпаведна дакументы заіндэксаваныя рускамоўнымі тэрмінамі.

На пачатковым этапе выканання ў БелСГБ навукова-даследчай работы па тэме «Беларускамоўная сельскагаспадарчая тэрміналогія як складнік шматмоўнага тэзаўруса Харчовай і сельскагаспадарчай арганізацыі Аб'яднаных Нацый: сістэматызацыя, упарадкаванне і навуковае апісанне»

(далее – НДР) для вивучэнне тэматычнага склада фонда бібліятэкі ажжыццаўся адбор дакументаў з электроннага каталога паслядоўным пошукам праз аўтаматызаваную бібліятэчную сістэму «ИРБИС64+» АРМ «Каталогизатор». Адабраны масіў дакументаў быў сістэматызаваны на падставе індэксавання рускамоўнымі дэскрыптарамі тэзаўруса AGROVOC. Профільныя бібліяграфічныя запісы навуковай галіновай беларускай літаратуры адбіраліся па паступленням ў фонд бібліятэкі за 2021 год. Усяго было абрана 14 900 тэрмінаў. Атрыманы масіў быў адсартаваны па назве ключавых слоў (у алфавіце ад А да Я) і прааналізаваны на прадмет паўтараў. Па выніках аналізу выяўлена, што 1429 дэскрыптараў не паўтараюцца, 1526 тэрміна паўтараюцца ад 2 да 10 раз у масіве, 549 дэскрыптараў паўтараюцца больш за 11 разоў у масіве. Пералік дэскрыптараў, якія найбольш часта паўтараюцца, дазволіў выявіць прыярытэтныя прадметныя катэгорыі (напрамкі) тэрміналагічнага даследавання, неабходнага для стварэння беларускамоўнай версіі тэзаўруса AGROVOC.

Праведзеныя даследаванні паказалі, што існуюць значныя адрозненні ў структуры і прынцыпах пабудовы названага слоўніка тэрмінаў і тэзаўруса AGROVOC. Так можна адзначыць адсутнасць у тэзаўрусе AGROVOC некаторых важных тэрмінаў, якія неабходны для складання пошукавай выявы дакументаў і ў наступстве для пошуку інфармацыі. Не хапае звыклых для нас назваў раслін (яны прадстаўлены на лацінцы, за выключэннем малога адсотка раслін). Напрыклад: у тэзаўрусе ЦНСГБ “пшеница – В1 зерновые культуры”, а ў тэзаўрусе AGROVOC прадстаўлена як “Triticum – Poaceae”, г.зн. толькі на лацінскай мове. І такая сітуацыя характэрна для ўсіх назваў раслін. Досвед працы бібліятэкі паказвае, што як для рускамоўнага, так і для беларускамоўнага карыстальніка адсутнасць звыклых тэрмінаў (рускамоўных ці беларускамоўных), у дадзеным выпадку назваў раслін, прыводзіць да страты інфармацыі, адбору няпоўнай інфармацыі па тэме, якая цікавіць. У той жа час, работа індэксатара накіравана, перш за ўсё, на задавальненне інфармацыйных патрэб і аблягчэнне пошуку інфармацыі ў электронных каталогах бібліятэк. Таму ў электронным каталогу БелСГБ назвы раслін прадстаўлены як на лацінскай, так і рускай мове. Такую практыку плануецца захаваць. Пасля заканчэння выканання названай навукова-даследчай працы для беларускамоўных выданняў мэтазгодна прыводзіць таксама тэрміны на трох мовах, у тым ліку на беларускай.

Таксама не хапае тэрмінаў, злучаных, напрыклад, з праблемамі аховы навакольнага асяроддзя. У прыватнасці, можна адзначыць адсутнасць у тэзаўрусе AGROVOC прамых тэрміна, якія тычацца аварыі на Чарнобыльскай АЭС (напрыклад, «Чернобыльская АЭС», «зоны экологического бедствия», «экологические катастрофы»). У тэзаўрусе AGROVOC мае месца тэрмін «антропогенные катастрофы». Але гэта прыватныя пытанні, якія можна вырашыць шляхам уключэння ў тэзаўрус

неабходных тэрмінаў. Акрамя таго, індэксаванне на рускай мове беларускамоўных дакументаў не заўсёды дае магчымасць у поўнай меры і карэктна выканаць іх аналітыка-сінтэтычную апрацоўку з-за адсутнасці поўнай эквівалентнасці дэскрыптараў, што абумоўлена моўнымі асаблівасцямі.

У рамках выканання адзначанай вышэй НДР прадугледжваецца ўкараненне сфарміраванага корпуса беларускамоўных тэрміналагічных адзінак у работу аўтаматызаванай бібліятэчнай сістэмы БелСГБ з мэтай выкарыстання беларускамоўных нарміраваных тэрмінаў тэзаўруса AGROVOC у якасці лінгвістычнага сродка для аналітыка-сінтэтычнай апрацоўкі беларускамоўных публікацый (дакументаў).

Варта адзначыць, што распрацоўка як базавай англамоўнай версіі, так і іншых нацыянальных варыянтаў тэзаўруса AGROVOC – гэта працяглы, працаёмкі і карпатлівы працэс. Стварэнне беларускамоўнай версіі знаходзіцца толькі на пачатковым этапе. Таму ў рамках названай НДР для апрабацыі магчымасцяў тэзаўруса AGROVOC для індэксавання дакументаў і для мэт інфармацыйнага пошуку у САБ «ІРБИС64+» была створана тэставая база дадзеных «AGROVOC_Тестовая БД_НИР», якая дазволіць распачаць працэс выкарыстання дадзенага лінгвістычнага сродку ў працы БелСГБ, а таксама не парушыць цэласнасць і структуру яе электроннага каталога.

Працэс індэксавання з выкарыстаннем беларускамоўнай версіі тэзаўруса AGROVOC будзе ажыццяўляцца ў адпаведнасці з існуючым алгарытмам (малюнак 2). Ва ўказанай тэставай базе дадзеных САБ «ІРБИС64+» у закладцы «Систематизация» АРМ «Каталогизатор» будзе запаўняцца поле 610 – «ненормированные ключевые слова» (праз паўтаральнае поле).

У сувязі з тым, што ў Рэспубліцы Беларусь на заканадаўчым узроўні прызнана дзве дзяржаўныя мовы (руская і беларуская), а таксама ўлічваючы той факт, што пераважная колькасць навуковых публікацый, кніг, навучальных дапаможнікаў і іншых дакументаў у рэспубліцы выдаецца на рускай мове, індэксаванне беларускамоўнымі тэрмінамі будзе дублявацца рускамоўнымі (з выкарыстаннем тэзаўруса ЦНСХБ). Гэта значна пашырыць інфармацыйна-пошукавыя магчымасці карыстальнікаў.

СПІС ЛІТАРАТУРЫ

1. AGROVOC Multilingual Thesaurus [Электронный ресурс]. – Режим доступа: <https://agrovoc.fao.org/browse/agrovoc/ru/>. – Режим доступа: 20.09.2023. – Дата доступа 04.09.2023.
2. Орлова, С. А. AGROVOC: Многоязычный тезаурус сельскохозяйственной терминологии в новом формате представления [Электронный ресурс] / С. А. Орлова // Культура: теория и практика : электрон. науч. журн. – 2018. – Вып. 1. – Режим доступа: <http://theoryofculture.ru/issues/80/1030>. – Дата доступа 04.09.2023.
3. Информационно-поисковые языки. Термины и определения : СТБ 7.74-2002. – Введ. 01.07.2003. – Минск : БелГИСС, 2003. – 32 с. – (Система стандартов по информации, библиотечному и издательскому делу).