

УДК 004.9

МОДЕЛІ ФОРМУВАННЯ ТА ПРОГНОЗУВАННЯ ЯКОСТІ ПРОЦЕСУ ОЦИФРУВАННЯ СТАРОДРУКІВ

О. Б. Гаврилишин, В. Б. Репета, І. З. Миклушка, Д. П. Лозовий, В. Я. Филь

Українська академія друкарства,
вул. Під Голоском, 19, Львів, 79020, Україна

Спираючись на попередні результати і дослідження, розроблено структурну і алгоритмічні моделі формування та прогнозування якості процесу оцифрування стародруків. Структурна модель процесу вміщує три блоки: створення структурної моделі процесу оцифрування стародруків, створення моделей формування якості процесу для реалізації доповненої реальності, прогнозування якості процесу оцифрування засобами нечіткої логіки. Кожен із запропонованих блоків є завершеною ланкою виконання технологічних й інформаційних операцій та отриманих при цьому результатів. Основою формування алгоритмічної моделі прогнозування стали вхідні дані: множина лінгвістичних змінних; універсальна терм-множина значень змінних; модель логічного формування якості процесу оцифрування; лінгвістичні терми; матриці знань лінгвістичних змінних; нечіткі логічні рівняння розрахунку значень функцій належності лінгвістичних змінних; функції належності; аналітичне вираження операції дефазифікації для розрахунку комплексного показника якості процесу оцифрування стародруків. Запропонована алгоритмічна модель розрахунку значення комплексного показника якості процесу оцифрування дозволить реалізувати на її основі прогностичну імітаційну модель.

Ключові слова: стародрук, процес оцифрування, структурна модель, алгоритмічна модель, доповнена реальність.

Постановка проблеми. Згідно з аналізом, проведеним аналітичною компанією Fact.MR, світовий обсяг реєстрації документів за період 2022–2032 роки зросте на 12,6 % [1], що передбачає також зростання ринку й у частці оцифрування стародрукованих видань. До кінця вищезазначеного періоду ринок буде оцінений у понад 336,7 мільярдів доларів США [1].

Питаннями отримання та зберігання цифрових копій стародрукованих видань займаються лідери цифрової індустрії Google, Amazon та Internet Archive, значні ресурси і кошти вкладають компанії Adobe, ABBYY, Apple, Microsoft, doo GmbH. У відомих бібліотеках світу створюються спеціальні підрозділи і залучаються провідні фахівці для творення та розповсюдження цифрових ресурсів. Розпорядженням Кабінету Міністрів України схвалено Стратегію розвитку бібліотечної справи на період до 2025 року «Якісні зміни бібліотек для забезпечення сталого розвитку України», в якій описуються сучасні вітчизняні бібліотечні ресурси і пропонуються невідкладні заходи з розвитку цифрових архівів, що передбачає фінансування

і розробку технологій проведення оцифрування фондів та їх зберігання [2]. Питання якості оцифрування книжок є популярними темами міжнародних наукових конференцій та форумів [3–5].

Інноваційною стала технологія доповненої реальності, яка створила можливість віддаленого доступу користування фондами бібліотек і підвищила зацікавленість до розвитку технологій оцифрування [6]. Тому створення цифрових копій стародрукованих видань з наступною реалізацією в системах доповненої реальності є сучасним розвитком архівно-бібліотечної справи.

Аналіз останніх досліджень та публікацій. У роботі [7] запропоновано організацію технологічної інфраструктури цифрового комплексу на основі взаємопов'язаних етапів: створення цифрових копій; адміністрування та зберігання цифрового фонду; опису та обліку цифрових ресурсів; обробки й управління цифровими ресурсами; пошуку та публікації цифрових ресурсів. У роботі [8] проаналізовано особливості процесу оцифрування стародрукованих видань як одного із сучасних методів зберігання світової культурної спадщини, а також встановлено фактори, які впливають на показники отриманої цифрової копії, та варіанти усунення дефектів як на рівні спеціалізованого обладнання, так і на рівні програмного забезпечення. Було встановлено такі пріоритетні фактори якості процесу оцифрування, як стан стародруку, особливість скануючого обладнання та функціональність програмного забезпечення. У праці [9], використовуючи інструменти нечіткої логіки, визначено базу знань із виконанням умови «якщо-тоді» та розраховано якість процесу оцифрування стародруків у кількісному вигляді. Представлені дослідження відображають окремі етапи, які стосуються процесів оцифрування стародруків, відповідно постає завдання розробки комплексної технології, функціонування якої забезпечить формування і прогнозування якості процесу оцифрування для наступного відтворення засобами доповненої реальності.

Мета статті – розроблення комплексної моделі формування і прогнозування якості процесу оцифрування стародруків.

Виклад основного матеріалу дослідження. Результати, отримані у попередніх дослідженнях [8–10], покладено в основу для створення моделей формування та прогнозування якості процесу оцифрування стародруків. Для початку зазначимо основні функціональні блоки розробленої технології. Кожен із блоків є завершеною ланкою виконання технологічних й інформаційних процедур та отриманих при цьому результатів.

Блок 1. Створення структурної моделі процесу оцифрування стародруків.

1.1. Аналіз технологій оцифрування друкованих зображень та програмного забезпечення для оцифрування та обробки цифрового зображення.

1.2. Особливості технологічних процесів оцифрування стародруків. Залежно від типу стародруку, його стану та поставленої мети оцифрування відбувається зміна операцій процесу оцифрування. Можливе застосування ручної реставрації стародруку. Якщо стародрук у вигляді аркушів, то достатньо оцифрування на планшетних пристроях; якщо стародрук у вигляді книжки, то виникає необхідність використання скануючих пристроїв, особливістю якого є розміщення стародруку на

V-подібному тримачі. Результат сканування вимагає додаткової програмної обробки щодо вирівнювання та виправлення кривизни сторінок, якої не вдалося уникнути при скануванні.

1.3. Аналіз проблем при оцифруванні стародруків та варіанти компенсації одержаних дефектів. При отриманні цифрової майстер-копії, згідно з вимогами щодо збереження оцифрованих документів, подальша цифрова обробка документа заборонена. Інша річ, коли цифрове зображення стародруку призначене для ознайомлення, видавничо-поліграфічного відтворення чи відтворення засобами доповненої реальності. З цією метою оцифроване зображення піддається цифровій обробці для покращення чіткості зображення, збільшення контрастності, усунення з документа сторонніх зображень та компенсації відсутності елементів зображення.

1.4. Побудова структурної моделі процесу оцифрування стародруків. Модель охоплює підготовчі операції, що вміщують вибір стародруку, встановлення його типу (ідентифікація), прийняття рішення про ручну реставрацію, якщо є така можливість, передання стародруку на оцифрування, що вимагає обліку та документального супроводу, зокрема електронного.

Блок 2. Створення моделей формування якості процесу оцифрування стародруків.

2.1. Згідно з проведеним анкетуванням експертів встановлено перелік факторів якості процесу оцифрування стародруків та визначено їхню важливість. Обробка результатів проводиться із застосуванням методики теорії нечітких множин. Фактор, який отримав максимальне значення, вважатиметься найбільш пріоритетним. Встановлення факторів якості процесу оцифрування стародруків.

2.2. Побудова орієнтованого графа відповідно до переліку отриманих факторів, вершини якого позначають множину факторів, а ребра ідентифікують впливи та залежності між ними.

2.3. Встановлення пріоритетності факторів серед вибраної множини методом ранжування та можливість на цій основі синтезування моделі їхнього пріоритетного впливу на якість процесу оцифрування стародруку.

2.4. Розроблення імітаційної моделі компенсації відсутності елемента на цифровому зображенні як інструменту формування якості оцифрування для наступного відтворення технологіями доповненої реальності. Використання при цьому запатентованого способу компенсації відсутності елементів зображення та відповідних фільтрів.

Блок 3. Прогнозування якості процесу оцифрування засобами нечіткої логіки.

3.1. Формування універсальної терм-множини значень та відповідних їй лінгвістичних термів пріоритетних факторів у вигляді лінгвістичних змінних. Побудова моделі логічного виведення, яка дозволяє отримати комплексний показник якості процесу оцифрування як результат використання засад нечіткої логіки.

3.2. Побудова матриць попарних порівнянь та розрахунок функцій належності лінгвістичних змінних для множини лінгвістичних термів відносно інтервалів значень універсальної терм-множини.



Рис. 1. Структурна модель формування та прогнозування якості процесу оцифрування стародруків

3.3. Розроблення бази знань із використанням нечітких логічних висловлювань типу «якщо <умова>, тоді <висновок>», що відтворює алгоритм реалізації якості процесу оцифрування залежно від значень лінгвістичних термів.

3.4. Побудова нечітких логічних рівнянь з матриці знань та функцій належності, які визначають зв'язок між функціями належності вхідних та вихідних даних.

3.5. Проведення операції дефазифікації за принципом «центр ваги» для розрахунку комплексного прогностичного показника якості процесу оцифрування. При цьому використовуються значення функцій належності лінгвістичних змінних, область існування яких визначена універсальною множиною.

Відповідно до опису трьох вищевказаних блоків синтезуємо структурно-функціональну модель інформаційної технології формування та прогнозування якості процесу оцифрування стародруків (рис. 1). Модель відображає послідовність реалізації та суть етапів і віднесених до них функціональних процедур інформаційної технології формування та прогнозування якості процесу оцифрування стародруків, а також взаємозв'язки між ними.

До однієї з основних компонент розробленої інформаційної технології належать спосіб [10] та імітаційна модель, яка уможливує проведення цифрової обробки зображення, суть якої полягає у компенсації відсутності елемента зображення шляхом заповнення цієї ділянки пікселями згідно з вказаною математичною моделлю.

Результати, отримані при застосуванні нечіткої логіки для прогнозування якості процесу оцифрування, дають можливість розробки алгоритму розрахунку числового значення комплексного показника якості процесу оцифрування та розробки імітаційної моделі. Основою формування алгоритму будуть вхідні дані: множина лінгвістичних змінних (вибрані пріоритетні фактори); універсальна терм-множина значень лінгвістичних змінних; модель логічного формування якості процесу оцифрування [9]; лінгвістичні терми лінгвістичних змінних; матриці знань лінгвістичних змінних; нечіткі логічні рівняння розрахунку значень функцій належності лінгвістичних змінних; функції належності терм-множин лінгвістичних змінних; аналітичне вираження операції дефазифікації як формули розрахунку комплексного показника якості процесу оцифрування стародруків.

На рис. 2 показано алгоритмічну модель показника якості процесу оцифрування стародруків для прогностичного оцінювання. При реалізації алгоритму передбачається функціонування програмних елементів за наступними етапами: введення елементів універсальної множини, які стосуються інтервалів можливих значень пріоритетних факторів (лінгвістичних змінних); формування бази знань; введення технологічних параметрів факторів на заданій універсальній терм-множині; побудова функцій належності та матриць з пронормованими їхніми значеннями у точках аналізу інтервалів параметризації лінгвістичних змінних; виконання умови «якщо-тоді»; дефазифікація, результатом якої є комплексний показник якості процесу оцифрування стародруків.

Алгоритмічна модель містить можливість механізму зворотного зв'язку, суть якого полягає у наданні користувачеві можливості уточнення та введення інших

значень вхідних даних у разі одержання низького прогностичного показника якості процесу.

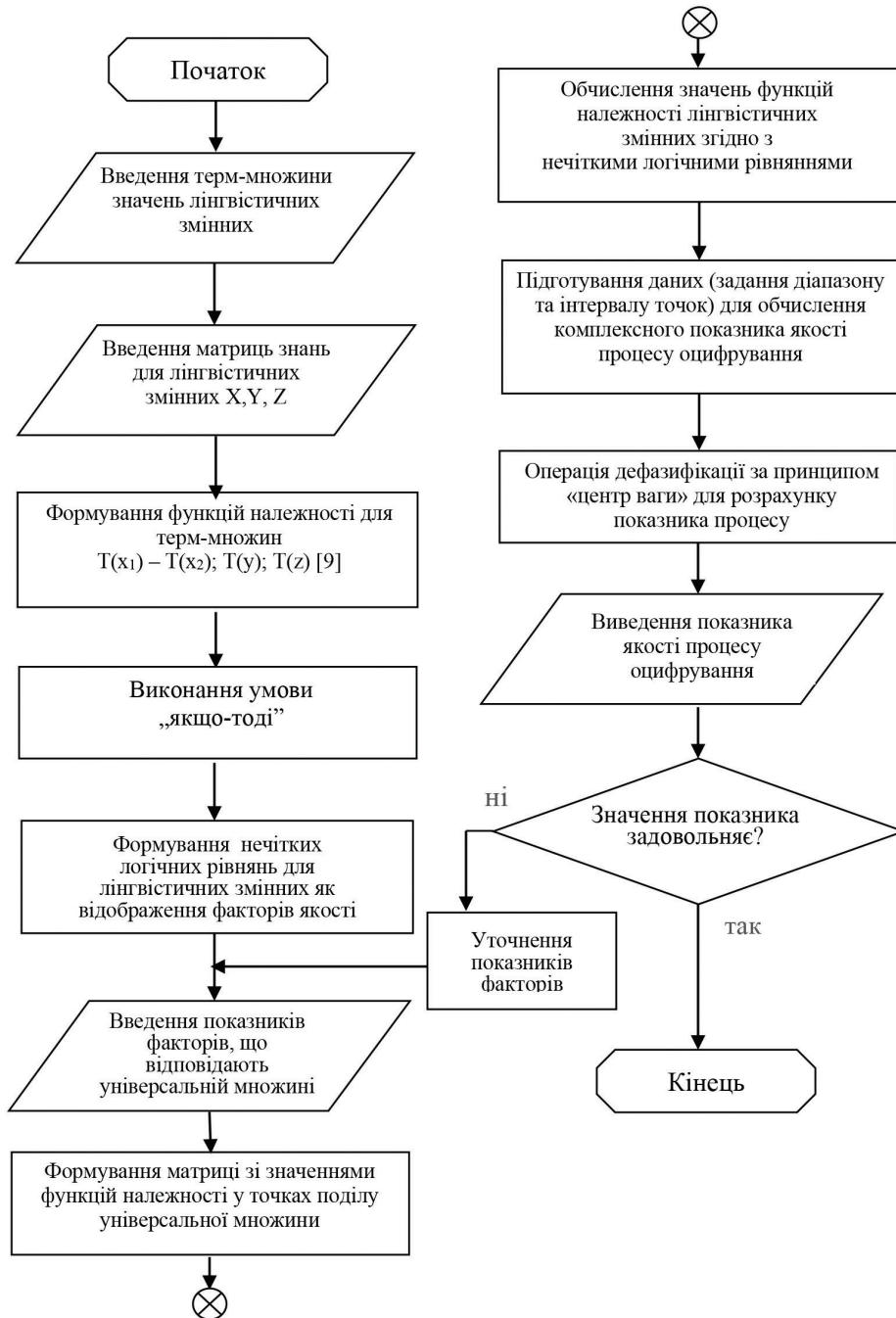


Рис. 2. Алгоритмічна модель прогнозування значення показника якості процесу оцифрування на основі засад нечіткої логіки

Висновки. На підставі попередніх досліджень та розробок було створено моделі формування та прогнозування якості процесу оцифрування стародруків, що включають програмний модуль компенсації відсутності зображень при цифровій обробці як складової формування якості цього процесу. Структурна модель формування та прогнозування якості процесу містить три основні блоки: розроблення структурної моделі процесу оцифрування стародруків, створення моделей формування якості процесу для реалізації доповненої реальності та прогнозування якості цього процесу за допомогою засобів нечіткої логіки. В сукупності модель відображає послідовність реалізації та суть етапів та віднесених до них функціональних процедур технології формування якості процесу оцифрування стародруків для наступної реалізації доповненої реальності з етапом прогнозування. Запропонована алгоритмічна модель для розрахунку значення комплексного показника якості процесу оцифрування дозволить створити на її основі прогностичну імітаційну модель.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Document Imaging Market. URL: <https://www.factmr.com/report/document-imaging-market>.
2. Розпорядження КМУ від 23 березня 2016 р. № 219 Про схвалення Стратегії розвитку бібліотечної справи на період до 2025 року «Якісні зміни бібліотек для забезпечення сталого розвитку України». 2016. URL: <https://zakon.rada.gov.ua/laws/show/219-2016-%D1%80#Text> (дата звернення: 21.07.2023).
3. Новий корпоративний науковий проект «Книжкові пам'ятки України» презентовано університетським бібліотекам. URL: <http://www.nbu.gov.ua/node/2425/>.
4. Костюченко О. В Острозькій академії реалізували грантовий проєкт із діджиталізації стародруків, 2022. URL: <https://www.oa.edu.ua/ua/info/news/2022/28-12-03/>.
5. Коцюба Є. Технологічні особливості оцифрування цінних та унікальних документів наукової бібліотеки. *Місце і роль бібліотек у формуванні національного інформаційного простору*: Міжнародна наукова конференція. 2014. URL: <http://conference.nbu.gov.ua/report/view/id/276/>.
6. Підсумки третього науково-практичного семінару «Оцифроване надбання: збереження, доступ, репрезентація». 2016. URL: <http://www.nbu.gov.ua/node/2932/>.
7. Іван Лобузин. Цифрові бібліотечні проєкти: технологічні рішення та управління життєвим циклом колекцій : монографія. НАН України, Нац. б-ка України ім. В. І. Вернадського. Київ, 2016. 216 с.
8. Гаврилишин О. Б., Репета В. Б., Миклушка І. З. Аналіз процесу оцифрування стародруків. *Науковий вісник НЛТУ України*. 2020. Т. 30. № 3. С. 130-134.
9. Гаврилишин О. Б., Репета В. Б., Миклушка І. З. Прогнозування якості процесу оцифрування стародруків засобами нечіткої логіки. *Видавнича справа і поліграфія*. 2020. № 1 (79). С. 11-19.
10. Пат. на корисну модель № 143254 Україна, МПК D21H 27/18. Спосіб компенсування відсутності елементів зображення у стародруках при їх оцифруванні / В. Б. Репета, О. Б. Цімер, І. З. Миклушка ; заявник та власник пат.: Укр. акад. друкарства ; заявл. 01.08.2019 ; заявка № u201909058 ; Бюл. № 14. 4 с.

REFERENCES

1. Document Imaging Market. Retrieved from <https://www.factmr.com/report/document-ima-ging-market> (in English).
2. Rozporiadzhennia KMU vid 23 bereznia 2016 r. № 219 Pro skhvalennia Stratehii rozvytku biblioteknoi spravy na period do 2025 roku «Iakisni zminy bibliotek dlia zabezpechennia staloho rozvytku Ukrainy». 2016. Retrieved from <https://zakon.rada.gov.ua/laws/show/219-2016-%D1%80#Text> (data zvernennia: 21.07.2023) (in Ukrainian).
3. Novyi korporatyvnyi naukovyi proekt «Knyzhkovi pamiatky Ukrainy» prezentovano uni-versytetskym bibliotekam. Retrieved from <http://www.nbuv.gov.ua/node/2425/> (in Ukrainian).
4. Kostiuchenko, O. (2022). V Ostrozhkii akademii realizuvaly hrantovyi projekt iz didzhytalizatsii starodrukiv. Retrieved from <https://www.oa.edu.ua/ua/info/news/2022/28-12-03/> (in Ukrai-nian).
5. Kotsiuba, Ye. (2014). Tekhnolohichni osoblyvosti otsyfruvannia tsinnykh ta unikalnykh do-kumentiv naukovoi biblioteki. Mistse i rol bibliotek u formuvanni natsionalnoho informatsii-noho prostoru: Mizhnarodna naukova konferentsiia. Retrieved from <http://conference.nbuv.gov.ua/report/view/id/276/> (in Ukrainian).
6. Pidsumky tretoho nauково-praktychnoho seminaru «Otsyfrovanie nadbannia: zberezhen-nia, dostup, reprezentatsiia». 2016. Retrieved from <http://www.nbuv.gov.ua/node/2932/> (in Uk-rainian).
7. Ivan Lobuzin. Tsyfrovi bibliotekni proekty: tekhnolohichni rishennia ta upravlinnia zhyttie-vym tsyklom kolektsii. NAN Ukrainy, Nats. b-ka Ukrainy im. V. I. Vernadskoho. Kyiv, 2016 (in Ukrainian).
8. Havrylyshyn, O. B., Repeta, V. B., & Myklushka, I. Z. (2020). Analiz protsesu otsyfruvannia starodrukiv: Naukovyi visnyk NLTU Ukrainy, 30, 3, 130–134 (in Ukrainian).
9. Havrylyshyn, O. B., Repeta, V. B., & Myklushka, I. Z. (2020). Prohnozuvannia yakosti pro-tsesu otsyfruvannia starodrukiv zasobamy nechitkoi lohiky: Vydavnycha sprava i polihrafiia, 1 (79), 11–19 (in Ukrainian).
10. Pat. na korysnu model № 143254 Ukraina, MPK D21H 27/18. Sposib kompensuvannia vidsutnosti elementiv zobrazhennia u starodrukakh pry yikh otsyfruvuvanni / V. B. Repeta, O. B. Tsimer, I. Z. Myklushka ; zaiavnyk ta vlasnyk pat.: Ukr. akad. drukarstva ; zaiavl. 01.08.2019 ; zaiavka № u201909058 ; Biul. № 14. 4 s. (in Ukrainian).

doi: 10.32403/2411-3611-2023-2-44-36-44

MODELS FOR THE FORMATION AND FORECASTING OF THE QUALITY OF THE OLD BOOKS DIGITIZATION PROCESS

O. B. Havrylyshyn, V. B. Repeta, I. Z. Myklushka, D. P. Lozovyi, V. Y. Fyl

*Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine
vreneta@gmail.com*

The article is devoted to the development of structural and algorithmic models of the formation and forecasting of the quality of the old books digitization process. The structural model of the process includes three blocks: creation of a structural model of the old books digitization process, creation of models of the process quality formation for the augmented reality implementation, forecasting the digitization process quality by means of fuzzy logic. At the same time, a method and a simulation model of compensation for the element absence on a digital image is used as a tool for forming the digitization process quality. Each of the proposed blocks is a completed link in the implementation of technological and informational operations and the results obtained. The input data become the basis for the formation of the algorithmic forecasting model: a set of linguistic variables; a universal term-set of variable values; a model of logic formation of the digitization process quality; linguistic terms; knowledge matrices of linguistic variables; fuzzy logic equations for calculating the values of the membership functions of linguistic variables; membership functions; analytical expression of the defuzzification operation for the calculation of a comprehensive quality indicator of the old books digitization process. The algorithmic model includes the possibility of a feedback mechanism, the essence of which is to provide the user with the opportunity to clarify and enter other values of the input data in case of receiving a low prognostic indicator of the process quality. The proposed algorithmic model for calculating the value of the complex indicator of the digitization process quality will allow implementing a prognostic simulation model based on it.

Keywords: *old books, digitization, algorithmic model, structural model, forecasting, augmented reality.*

Стаття надійшла до редакції 03.08.2023.

Received 03.08.2023.