

ПРОГНОЗУВАННЯ ЯКОСТІ ПРОЦЕСУ ОЦИФРОВУВАННЯ
СТАРОДРУКІВ ЗАСОБАМИ НЕЧІТКОЇ ЛОГІКИ

О. Б. Гаврилишин, В. Б. Репета, І. З. Миклушка

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

На основі встановлення таких пріоритетних факторів якості процесу оцифрування стародруків, як стан стародруку, особливість скануючого обладнання та функціональність програмного забезпечення, сформовано базу знань із виконанням умови «якщо – то» та розроблено модель логічного виведення. Побудова моделі логічного виведення, нечітких логічних рівнянь, встановлення універсальної множини та термів оцінювання уможливили формування кількісного показника якості процесу оцифрування стародруків. Сформовану базу знань було перевірено під час моделювання за допомогою системи Fuzzy Logic Toolbox середовища технологічних розрахунків Matlab за принципом Мамдані з використанням методу «Центр ваги» для операції дефазифікації. Запропонована методика розрахунку якості процесу оцифрування стародруків у кількісному вигляді уможливорює розробку імітаційної моделі для його прогнозування та статистичного оцінювання.

Ключові слова: оцифрування стародруків, модель, фактори якості, лінгвістичні змінні, нечіткі логічні рівняння.

Постановка проблеми. Цифрові технології сьогодні стали звичними. Серед них окремою сферою є створення та використання цифрових документів. У «доцифрову» еру людство накопичувало величезні обсяги паперових документів, видань, які зберігаються в бібліотеках і архівах. Зараз технології сканування паперових документів широко застосовуються у всіх сферах діяльності, а не тільки в архівній і бібліотечній справі, вони оснащені досконалыми апаратними та програмними засобами, зокрема мобільними пристроями, які стали повсякденними помічниками людини.

Проте коли йдеться про оцифрування оригіналів, вік яких перевищує одне десятиліття або і століття, звичайні методи, пристрої, а отже, і технології мають певні обмеження. Вік та умови використання чи зберігання таких «старих» оригіналів мають різний вплив на документ, що «дійшов» до наших часів.

У кожному разі, якщо необхідно отримати якісну цифрову копію такого документа, виникає задача, яка рідко має однотипний характер, а сам технологічний процес сканування не підлягає стандартизованій параметризації. Кількість та якість даних, які треба враховувати під час проектування процесів сканування старих документів, не є сталою величиною. Особливо це притаманно процесам

оцифрування рукописних або стародрукованих видань, бо тут ми взаємодіємо не з одиничним документом, а з багатьма сторінками, які зібрані у зошити і зброшуровані у вигляді книжки. І навіть те, що всі ці сторінки зберігалися палітуркою як єдиний інформаційний носій, не є гарантією, що сторінки виглядають однаково і можуть опрацьовуватися за однотипними технологічними схемами та операціями.

У цих умовах виникає потреба в застосуванні не тільки традиційних статистичних підходів, а й інших, нестандартних методів аналізу, що ґрунтуються на отриманні якісного оцінювання даних і здійсненні технологічних розрахунків на цій основі. Одним із методів інтелектуального аналізу даних є застосування математичного апарату нечіткої логіки, що може бути використаний під час оброблення технологічних показників процесу сканування рукописних та стародрукованих видань.

Вивчаючи фактори впливу на процес сканування таких документів, ми можемо передбачити значення цих факторів, які змінюють свої показники у проміжних значеннях, тому виникає змога використати логічні операції над нечіткими множинами параметрів сканування і сформулювати поняття лінгвістичної змінної, якою є нечіткі множини.

Аналіз останніх досліджень та публікацій. У цьому напрямі відомі праці, які стосуються застосування нечіткої логіки для забезпечення якості технологічних процесів [1, 2], комплексного оцінювання якості флексографічних друкарських відбитків [3], а також прогнозування якості та проектування книжкових видань [4, 5].

Мета статті — розробити модель прогнозування якості процесу оцифрування стародруків, спираючись на встановлені пріоритетні фактори процесу та використовуючи засади нечіткої логіки.

Виклад основного матеріалу дослідження. Розробка систем, здатних адекватно відтворювати міркування людини, вимагає математичного апарату, який переводить неоднозначні твердження на мову чітких і формальних математичних формул. Це було реалізовано в роботі Латфі Заде, яка створила основи напрямку нечіткої логіки і ввела поняття деякої універсальної множини для всієї проблемної області [6].

Загалом нечітка логіка — це логіка, що оперує лінгвістичними змінними за допомогою правил, зрозумілих для людини і наближених за структурою до звичайної розмовної мови. Перевага систем із нечіткою логікою — можливість оперувати нечіткими вхідними даними.

Оцінювання якості процесу оцифрування стародрукованих видань за допомогою засобів нечіткої логіки містить такі етапи:

- встановлення універсальної терм-множини значень та відповідних їй лінгвістичних термів пріоритетних факторів якості (лінгвістичних змінних);
- побудова матриць попарних порівнянь для множини лінгвістичних термів відповідного інтервалу значень універсальної множини та отримання для кожної з матриць функцій належності;

- розроблення нечіткої бази знань із використанням нечітких логічних висловлювань типу «якщо – то»;
- побудова нечітких логічних рівнянь на підставі матриці знань і функцій належності, які визначають зв'язок між функціями належності вхідних та вихідних даних;
- дефазифікація нечіткої множини, суть якої полягає в розрахунку числового показника прогнозованої якості, наприклад, за методом центра ваги плоскої фігури [7, 8, 9].

Якість процесу оцифрування стародруку Q залежить від характеристики об'єкта, який сканується, характеристики обладнання і технологічного процесу сканування та обробки цифрового зображення. Відповідно, якість процесу оцифрування стародруку визначається за формулою:

$$Q = f(X, Y, Z), \quad (1)$$

де X — лінгвістична змінна, яка характеризує сканований об'єкт; Y — лінгвістична змінна, яка характеризує скануюче обладнання; Z — лінгвістична змінна, яка характеризує програмне забезпечення.

Лінгвістичну змінну, яка описує вплив сканованого об'єкта, визначають за залежністю:

$$X = f(x), \quad (2)$$

де x — характеристика стану стародруку.

Лінгвістична змінна, яка описує вплив параметрів обладнання, визначають за залежністю:

$$Y = f(y), \quad (3)$$

де y — особливість скануючого обладнання (планшетне чи V-подібна конструкція).

Лінгвістичну змінну, яка описує особливість технологічного процесу сканування й обробки, визначають за залежністю:

$$Z = f(z), \quad (4)$$

де z — функціональність програмного забезпечення.

Відповідно до результатів експертних суджень і експериментальних досліджень сформуємо схему логічного формування якості процесу оцифрування стародруків (рис. 1).

Оцінювання значень лінгвістичних змінних проведемо за допомогою системи якісних понять. Кожне з цих понять становить відповідну нечітку множину, тобто деяку властивість, яка розглядається як лінгвістичний терм. Лінгвістичні змінні, які забезпечують якість процесу оцифрування стародруків, та оціночні терми наведено в табл. 1.

Спираючись на експертні висловлювання щодо впливу факторів на якість процесу оцифрування стародруків, побудуємо функції належності. Відповідно, значення показника «Стан стародруку» визначимо на універсальній множині: $u_1 = 40$ од.; $u_2 = 50$ од.; $u_3 = 60$ од.; $u_4 = 70$ од.; $u_5 = 80$ од.; $u_6 = 90$ од.; $u_7 = 100$ од.

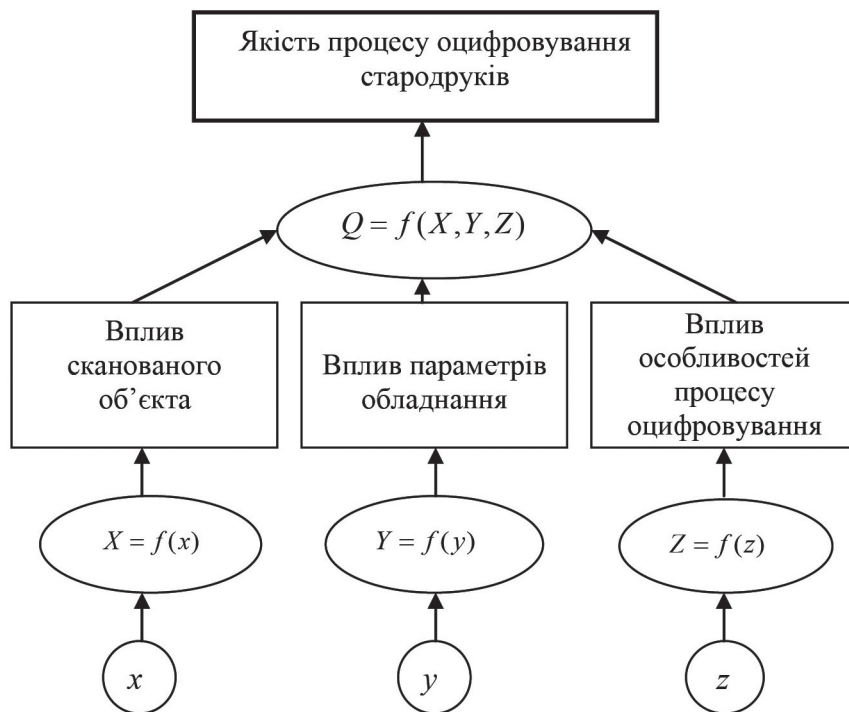


Рис. 1. Схема логічного формування якості процесу оцифрування стародруків

Таблиця 1

Лінгвістичні змінні параметрів якості процесу оцифрування стародруків

№	Назва змінної	Універсальна множина	Терми оцінювання
1	Стан стародруку	40–100 од.	Поганий Задовільний Відмінний
2	Особливість скануючого обладнання	50–100 од.	Планшетний V-подібний
3	Функціональність програмного забезпечення	10–100 од.	Низька Середня Висока

Для лінгвістичної оцінки цього показника використовуємо сукупність нечітких термів: $T(x) = \langle \text{поганий, задовільний, відмінний} \rangle$. Відповідно до вказаних термів отримаємо функції належності лінгвістичної змінної «Стан стародруку» (рис. 2а).

Для змінної «Особливість скануючого обладнання» показник визначимо на універсальній множині: $u_1 = 50$ од.; $u_2 = 60$ од.; $u_3 = 70$ од.; $u_4 = 80$ од.; $u_5 = 90$ од.; $u_6 = 100$ од.

Для лінгвістичної оцінки змінної використовуємо сукупність нечітких термів: $T(y) = \langle \text{планшетний, V-подібний} \rangle$. Отже, щодо показника «Особливість скануючого обладнання» за відношенням до зазначених термів отримаємо функції належності (рис. 2б).

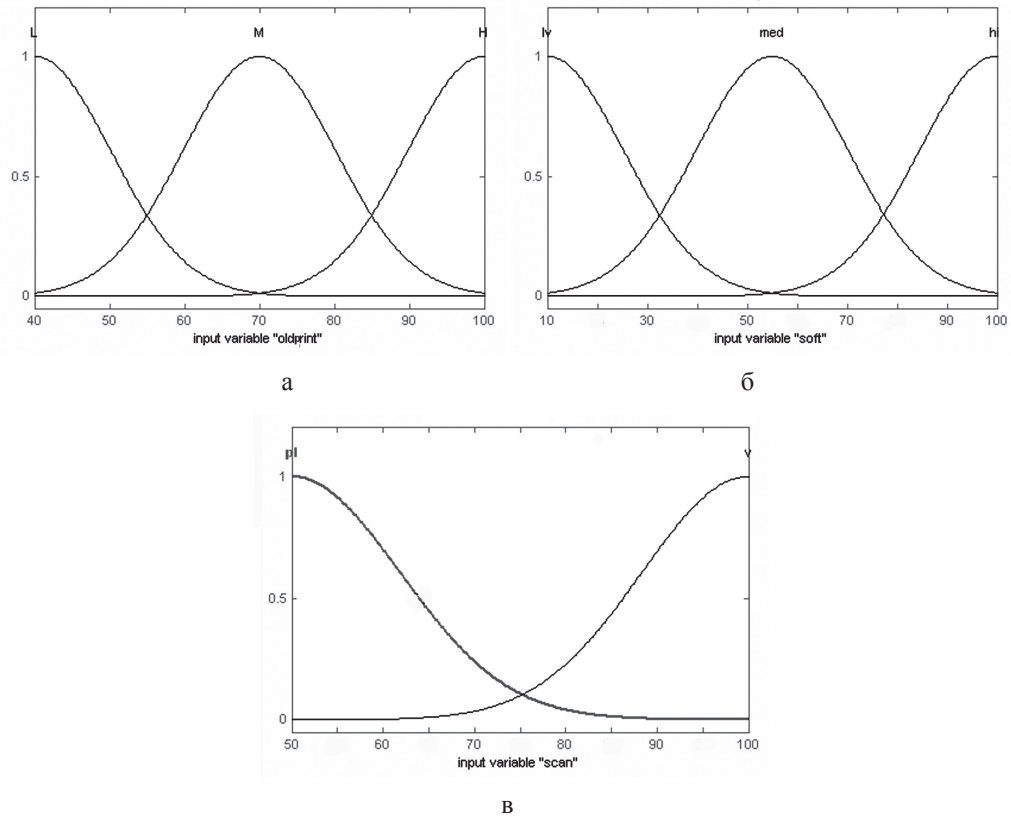


Рис. 2. Функції належності лінгвістичних змінних:

а — стан стародруку; б — функціональність програмного забезпечення;

в — особливість скануючого обладнання

Побудуємо функції належності для лінгвістичної змінної «Функціональність програмного забезпечення» як один із факторів якості процесу (рис. 2в). Функціональність програмного забезпечення визначимо на універсальній множині: $u_1 = 10$ од.; $u_2 = 30$ од.; $u_3 = 50$ од.; $u_4 = 70$ од.; $u_5 = 90$ од.; $u_6 = 100$ од.

Для лінгвістичної оцінки показника використовуємо сукупність нечітких термів: $T(L) = \langle \text{низька, середня, висока} \rangle$.

Сформуємо нечітку базу знань щодо вибраних параметрів якості процесу оцифрування:

1. Для терму якість процесу «низька»:

- якщо стан стародруку «поганий», особливість скануючого обладнання «V-подібний», функціональність програмного забезпечення «низька», то якість процесу «низька», або
- якщо стан стародруку «задовільний», особливість скануючого обладнання «планшетний», функціональність програмного забезпечення «низька», то якість процесу «низька».

2. Для терму якість процесу «середня»:

- якщо стан стародруку «задовільний», особливість скануючого обладнання «V-подібний», функціональність програмного забезпечення «середня», то якість процесу «середня», або
- якщо стан стародруку «поганий», особливість скануючого обладнання «планшетний», функціональність програмного забезпечення «середня», то якість процесу «середня».

3. Для терму якість процесу «висока»:

- якщо стан стародруку «відмінний», особливість скануючого обладнання «V-подібний», функціональність програмного забезпечення «висока», то якість процесу «висока», або
- якщо стан стародруку «відмінний», особливість скануючого обладнання «V-подібний», функціональність програмного забезпечення «середня», то якість процесу «висока».

Логічні рівняння забезпечення якості процесу оцифрування стародруків виглядають так:

$$\mu^n = \mu^n(x) \wedge \mu^{v-nod}(y) \wedge \mu^n(z) \vee \mu^3(x) \wedge \mu^{nl}(y) \wedge \mu^n(z); \quad (5)$$

$$\mu^c = \mu^3(x) \wedge \mu^{v-nod}(y) \wedge \mu^c(z) \vee \mu^n(x) \wedge \mu^{nl}(y) \wedge \mu^c(z); \quad (6)$$

$$\mu^6 = \mu^6(x) \wedge \mu^{v-nod}(y) \wedge \mu^6(z) \vee \mu^6(x) \wedge \mu^{v-nod}(y) \wedge \mu^c(z). \quad (7)$$

Позначення \wedge і \vee — це операції визначення мінімуму та максимуму у логічних рівняннях.

Скориставшись функціями належності і підставивши ступені належності в систему нечітких логічних рівнянь, можна розрахувати один із варіантів впливу пріоритетних факторів на процес оцифрування.

Для проведення операції дефазифікації за методом «Центр ваги» [8] встановимо верхню і нижню межу якості процесу Q : нижня — 1 у. о., верхня — 100 у. о. Провівши дефазифікацію отриманих нечітких значень процесу, отримаємо кількісну прогностичну оцінку якості процесу.

Сформовану базу знань було перевірено під час моделювання за допомогою системи Fuzzy Logic Toolbox середовища технологічних розрахунків MATLAB за принципом Мамдані [10, 11] (рис. 3).

Результати моделювання засвідчують адекватність розробленої бази знань і можливості її використання для комплексного оцінювання якості процесу оцифрування стародруків задля його прогнозування.

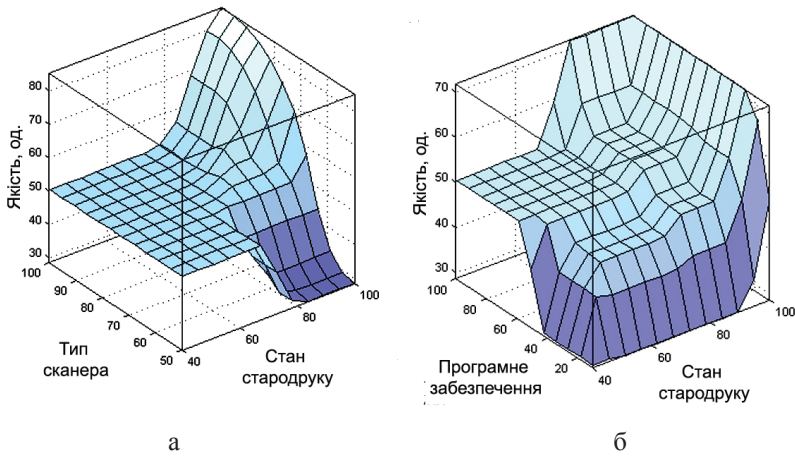


Рис. 3. Модель впливу факторів процесу оцифрування на його якість:
 а — вплив особливості скануючого обладнання і стану стародруку;
 б — вплив функціональності програмного забезпечення та стану стародруку

Висновки. Отже, у статті опрацьовано пріоритетні фактори якості процесу оцифрування стародруків. Проведений аналіз із використанням експертно-лінгвістичної інформації і правила «якщо – то» дав змогу отримати нечіткі логічні рівняння впливу лінгвістичних змінних на якість процесу і, відповідно, комплексно та кількісно оцінити якість технологічного процесу оцифрування стародруків. Запропонована методика розрахунку якості процесу оцифрування у кількісному вигляді уможливорює розробку імітаційної моделі його прогнозування та статистичного оцінювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Репета В. Б., Сеньківський В. М., Гургалъ Н. С. Прогнозування якості процесу УФ-флексографічного друку етикетки на основі нечіткої логіки. Наукові записки [Української академії друкарства]. 2014. № 1–2. С. 84–89.
2. Repeta V., Kukura Y. Quantitative Evaluation of Quality of Flexographic Prints by Means of Fuzzy Logic. Acta Graphica. 2016. № 1. Pp. 39–46.
3. Repeta V., Kukura Y., Kukura V. Analysis of the relationship of quality factors in the solventless lamination process. Journal of Print and Media Technology Research. 2018. № 1. Pp. 27–34.
4. Теоретичні основи забезпечення якості видавничо-поліграфічних процесів (Частина 4. Прогнозування та забезпечення якості засобами нечіткої логіки) / Піх І. В., Сеньківський В. М., Сеньківська Н. Є., Калиній І. В. Наукові записки [Української академії друкарства]. 2017. № 1 (54). С. 22–29.
5. Піх І. В., Сеньківський В. М. Інформаційні технології моделювання видавничих процесів : навч. посіб. Львів : УАД, 2013. 220 с.
6. Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений. Москва : Мир, 1976. 165 с.

7. Сявавко М. С. Інформаційна система «Нечіткий експерт». Львів : Видавничий центр ЛНУ ім. І. Франка., 2007. 320 с.
8. Ротштейн А. П. Интеллектуальные технологии идентификации: нечеткие множества, нейронные сети, генетические алгоритмы. Винница : УНІВЕРСАМ-Вінниця, 1999. 320 с.
9. Репета В. Б. Методологічні основи забезпечення якості процесу флексографічного друку етикетки УФ-фарбами : дис. ... д-ра техн. наук : 05.05.01. Львів, 2019.
10. Штовба С. Д. Проектирование нечетких систем средствами MATLAB. Москва : Горячая линия – Телеком, 2007. 288 с.
11. Mamdani E. H., Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*. 1975. Vol. 7. № 1. Pp. 1–13.

REFERENCES

1. Repeta, V. B., Senkivskyi, V. M., & Hurhal, N. S. 2014. Prohnozuvannia yakosti protsesu UF-fleksografichnoho druku etyketky na osnovi nechitkoi lohiky: Naukovi zapysky [Ukrainskoi akademii drukarstva], 1–2, 84–89 (in Ukrainian).
2. Repeta, V., & Kukura, Y. (2016). Quantitative Evaluation of Quality of Flexographic Prints by Means of Fuzzy Logic: *Acta Graphica*, 1, 39–46 (in English).
3. Repeta, V., Kukura, Y., & Kukura, V. (2018). Analysis of the relationship of quality factors in the solventless lamination process: *Journal of Print and Media Technology Research*, 1, 27–34 (in English).
4. Pikh, I. V., Senkivskyi, V. M., Senkivska, N. Ye., & Kalynii, I. V. (2017). Teoretychni osnovy zabezpechennia yakosti vydavnycho-polihrafichnykh protsesiv (Chastyna 4. Prohnozuvannia ta zabezpechennia yakosti zasobamy nechitkoi lohiky): Naukovi zapysky [Ukrainskoi akademii drukarstva], 1 (54), 22–29 (in Ukrainian).
5. Pikh, I. V., & Senkivskyi, V. M. (2013). Informatsiini tekhnolohii modeliuvannia vydavnychykh protsesiv. Lviv : UAD (in Ukrainian).
6. Zade, L. A. (1976). Ponjatie lingvisticheskoy peremennoy i ego primenenie k prinjatiyu priblizhennykh reshenij. Moskva : Mir (in Russian).
7. Siavavko, M. S. (2007). Informatsiina systema “Nechitkyi ekspert”. Lviv : Vydavnychiy tsentr LNU im. I. Franka (in Ukrainian).
8. Rotshtejn, A. P. (1999). Intellektual’nye tehnologi identifikacii: nechetkie mnozhestva, nejronnye seti, geneticheskie algoritmy. Vinnica : UNIVERSAM-Vinnicja (in Russian).
9. Repeta, V. B. (2019). Metodolohichni osnovy zabezpechennia yakosti protsesu fleksografichnoho druku etyketky UF-farbamy : dys. ... d-ra tekhn. nauk : 05.05.01. Lviv (in Ukrainian).
10. Shtovba, S. D. (2007). Proektirovanie nechetkih sistem sredstvami MATLAB. Moskva : Gorjachaja liniya – Telekom (in Russian).
11. Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller: *International Journal of Man-Machine Studies*, 7, 1, 1–13 (in English).

FORECASTING THE QUALITY OF OLD RAREBOOKS DIGITAZATION PROCESS BY FUZZY LOGIC

O. B. Havrylyshyn, V. B. Repeta, I. Z. Myklushka

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

In the “pre-digital” era, humanity has accumulated huge amounts of paper documents, books, which are stored in libraries and archives. With the development of information technology, the question arose of preserving such documents in digital form and familiarizing stakeholders, using various means of visualization. Every time it is necessary to obtain a high-quality digital copy of a certain document, there is a problem of high-quality technological process of digitization, which due to the variety of old rarebook is not subject to standardized parameterization. Based on the priority of quality factors of the process of digitization of old rarebooks, such as the state of pages, features of scanning equipment and software functionality, a knowledge base is formed with the fulfillment of the condition “if-then” and a model of logical inference is developed. The construction of a model of logical inference, fuzzy logical equations, the definition of a universal set and evaluation terms allow determining a quantitative indicator of the quality of the process of digitization of old rarebooks. The formed knowledge base has been checked at modelling by means of technological calculations system Fuzzy Logic Toolbox of MatLab program of the according to the Mamdani principle. The construction of membership functions for variables “Status of an old rarebook”, “Feature of scanning equipment” and “Functionality of software” has been checked by the method “Center of gravity” for the dephasification operation. Accordingly, a model of the influence of factors that determine the quality of the digitization process is obtained and the adequacy of the developed knowledge base is confirmed. The proposed method for calculating the quality of the old rarebooks digitization process in quantitative form allows the development of a simulation model for its forecasting and statistical evaluation.

Keywords: digitization of old rarebooks, logical model, quality factors, linguistic variables, fuzzy logical equations.

Стаття надійшла до редакції 27.02.2020.

Received 27.02.2020.