

УДК 681.3

**О. В. Коваль, О. В. Карабчук**

Інститут проблем реєстрації інформації НАН України  
вул. М. Шпака, 2, 03113 Київ, Україна  
e-mail: svr@cki.ipri.kiev.ua

## **Реалізація методу потокового введення документів, що зберігаються в електронному вигляді**

*Розглянуто питання потокового введення інформації. Запропоновано реалізацію методу потокового введення документів, що перебувають в електронному вигляді.*

**Ключові слова:** системи автоматизації документообігу, реєстраційна карта документа, потокове введення документів, електронний архів, інформаційне сховище.

### **Вступ**

Потоки вхідної інформації корпоративних інформаційно-аналітичних систем (КІАС) надзвичайно різноманітні за джерелами і формами її подання. Вхідну інформацію за формою збереження можна умовно класифікувати на електронні документи і паперові. Під терміном «електронний документ» зазвичай розуміється текстовий файл, документ MS Word, електронна таблиця MS Excel, графічний файл, кілька взаємозалежних файлів різних форматів тощо. Прийом вхідної інформації в КІАС базується, як правило, на використанні засобів електронного документообігу [1, 2], у тому числі систем автоматизації документообігу (САД) [3] і систем потокового сканування-розпізнавання.

САД здійснюють інтеграцію в інформаційне сховище (ІС) КІАС електронних документів, які надходять із використанням електронної пошти, Інтранет-технологій та інше. Під інтеграцією у даному контексті слід розуміти виконання таких основних етапів автоматизованої обробки документів, що надходять до КІАС: прийом, реєстрація (формування реєстраційної картки) та розміщення даних про документ (метаданих) у документальній базі даних. Самі електронні документи зберігаються в спеціальних інформаційних сховищах чи у файлах операційної системи. На сьогоднішній день САД в Україні досить широко представлені як українськими розробниками програмних комплексів, так і закордонними [1]. САД надає широкі пошукові можливості для швидкого та зручного одержання будь-якої

інформації про електронні документи, накопичені й збережені в інформаційному сховищі КІАС.

Однак, якщо документи надходять у паперовому вигляді, то інтеграція документів у ІС КІАС реалізується за допомогою систем потокового сканування-розпізнавання, які базуються на технології потокового введення документів (ПВД). Потокове введення документів — це процес отримання графічного образу паперового документа у форматі електронного документа та подальшої роботи з ним у залежності від поставлених задач [4]. На відміну від САД системи потокового сканування-розпізнавання забезпечують можливість збереження графічного образу документа в базі даних. Процес інтеграції документів із використанням ПВД в ІС КІАС [5] включає в себе наступну послідовність дій:

- 1) підготовку до сканування — розшивку, сортування документів, підготовку листів нестандартних розмірів і пошкоджених документів;
- 2) сканування — підбір параметрів сканування для отримання максимально чіткого зображення, фільтрацію фону для старих архівних документів, сортування різноформатних документів, вирівнювання листів, повторне сканування у випадку недостатньої якості зображення;
- 3) обробку — повнотекстове розпізнавання, зональне розпізнавання, ідентифікацію форм;
- 4) атрибутування — формування метаданих (атрибутів) про документ;
- 5) верифікацію — перевірку атрибутів, що вводяться, незалежну перевірку двома операторами;
- 6) контроль якості — на цьому етапі проходить ручний розбір документа, при неможливості автоматичної обробки, або зміна документа;
- 7) експорт — формування шаблону документа (отримання графічного образу паперового документа у форматі електронного документа) в залежності від системи управління базами даних або САД, що використовуються, та розміщення його у відповідне інформаційне сховище із забезпеченням доступу до нього.

На сьогоднішній день вже існує багато систем, в яких інтеграція документів у ІС КІАС реалізується за допомогою систем потокового сканування-розпізнавання, які базуються на технології ПВД. Серед найбільш відомих, що були досліджені, можна назвати такі як АВВУУ FineReader Банк, FormReader, система «ДЕЛО», система групи компаній «Ланит». Але наявні системи орієнтовані лише на роботу з паперовими документами, повністю виключаючи можливість роботи з електронними копіями паперового документа.

Виходячи із зазначених вище процесів інтеграції документів у ІС КІАС із використанням САД і ПВД, можна зробити висновок, що вони не забезпечують інтеграцію графічного образу паперового документа у форматі електронного документа, якщо такий електронний документ вже існує до процесу інтеграції. А, враховуючи те, що в сучасних КІАС практично стандартом стає забезпечення доступу до графічного образу паперового документа, що інтегруються в КІАС, тобто створення інформаційного сховища документів — програмно-апаратного комплексу, що вирішує задачі введення і зберігання електронних графічних образів документів, а також доступу до них, то постає задача створення засобів інтеграції графічного образу паперового документа, що існує в форматі електронного документа, в КІАС.

## Метод потокового введення документів, що зберігаються в електронному вигляді, до корпоративних інформаційно-аналітичних систем

Вирішити зазначену задачу можливо, використовуючи ті підходи, що реалізовані в САД і ПВД. Пропонується удосконалений метод ПВД, суть якого полягає в тому, що такі етапи ПВД як підготовка до сканування, сканування та обробка документа виключаються, а натомість впроваджується розширений етап автоматизованої обробки електронних документів, подібний наявному в САД.

За удосконаленим методом ПВД здійснюється обробка електронних документів, які існують у вигляді файлів наступних форматів: doc, docx, xls, xlsx, txt, rtf, rar, pdf, djvu, jpg, gif, bmp та ін. Для кожного формату файлів розроблені відповідні сценарії їхньої обробки. Під сценарієм ми розуміємо послідовність виконання дій, що необхідні для реалізації визначених завдань, у нашому випадку, виконання наступних операцій (рис. 1):

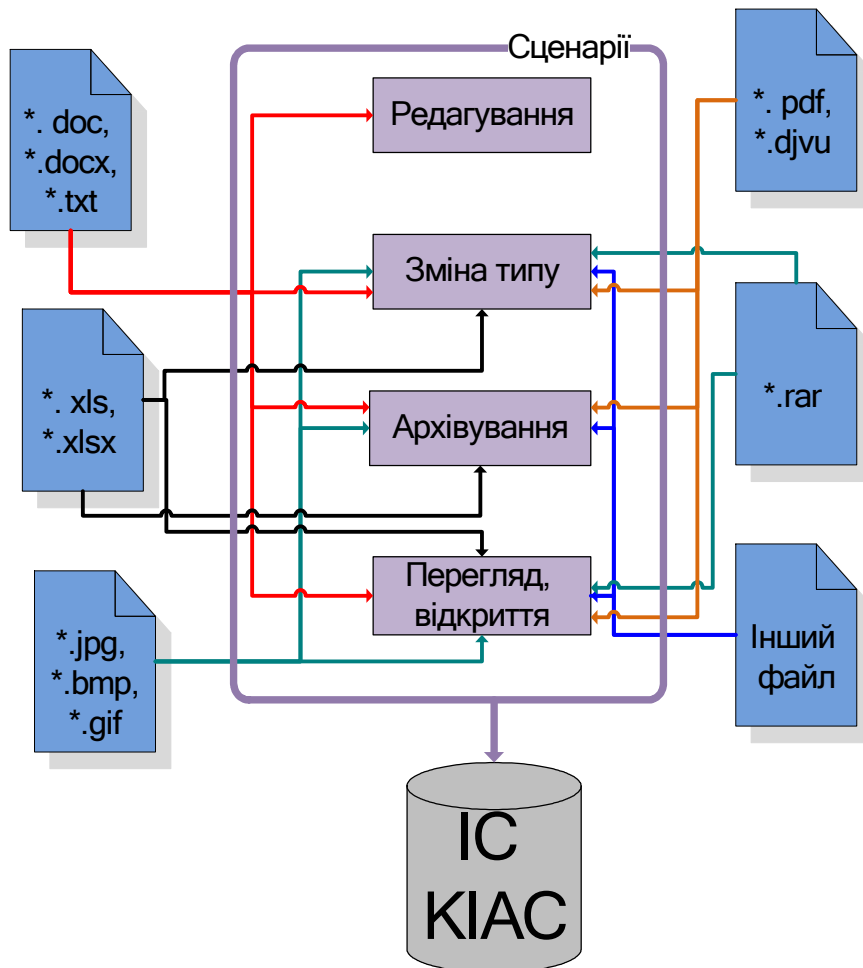


Рис. 1. Схема автоматизованої обробки електронних документів

- 1) архівування — архівація файлів, розархівування файлів;
- 2) редагування — відкриття документа для внесення змін;
- 3) перегляду — відкриття файлу, зчитування метаданих файлу та занесення їх та електронного графічного образу документа до ІС КІАС;
- 4) зміни типу — зміна формату файлу.

Для прикладу розглянемо сценарій перегляду файлів. Для даного сценарію був розроблений відповідний алгоритм його реалізації, тобто кінцевий опис конкретної послідовності дій, що потрібно виконати за певний час для досягнення результату вирішення задач [6].

Спочатку користувач вибирає необхідний файл, після чого в автоматичному режимі зчитується інформація про цей файл, а саме: дата створення, розмір файлу, дата останнього відкриття файлу, формат файлу, його фізична адреса, та відображається ця інформація на екрані автоматизованого робочого міста користувача. Це досягається завдяки розробленому алгоритму зчитування атрибутів файлу. У разі, якщо користувач відкриває файл для перегляду, файл відкривається автоматично у тій програмі, в якій файли даного формату відкриваються за замовчанням. Для реалізації даної можливості був розроблений алгоритм завантаження необхідних бібліотек для відкриття файлу перед користувачем. У разі, якщо формат файлу невідомий, користувачеві пропонується самостійно вибрати необхідну програму для відкриття та перегляду файлу.

На базі удосконаленого методу ПВД розроблений програмний комплекс, що працює згідно такого алгоритму (рис. 2):

- 1) користувачеві пропонується вибрати необхідний файл (див. рис. 3);
- 2) перевіряється наявність вибраного файлу, зчитуються його атрибути, визначається формат файлу;
- 3) користувач може переглянути атрибути вибраного файлу або перейти до вибору сценаріїв (див. рис. 4);
- 4) визначається сценарій, який можна застосувати до цього формату файлу;
- 5) користувач обирає доступні сценарії обробки файлу, або може вибрати інший файл;
- 6) у тому випадку, якщо був вибраний один із сценаріїв обробки файлу, проводиться обробка файлу за заданим сценарієм, записується інформація до бази даних і виводиться службова інформація про хід виконання того чи іншого сценарію;
- 7) після успішної чи невдалої роботи сценарію, пропонується вибрати інший файл чи закінчити роботу.

Нижче наведені приклади екранних форм інтерфейсу користувача розробленого програмного комплексу, а саме: екранна форма вибору файлу для подальшої інтеграції до ІС КІАС (рис. 3) та екранна форма відображення атрибутів вибраного файлу (рис. 4).

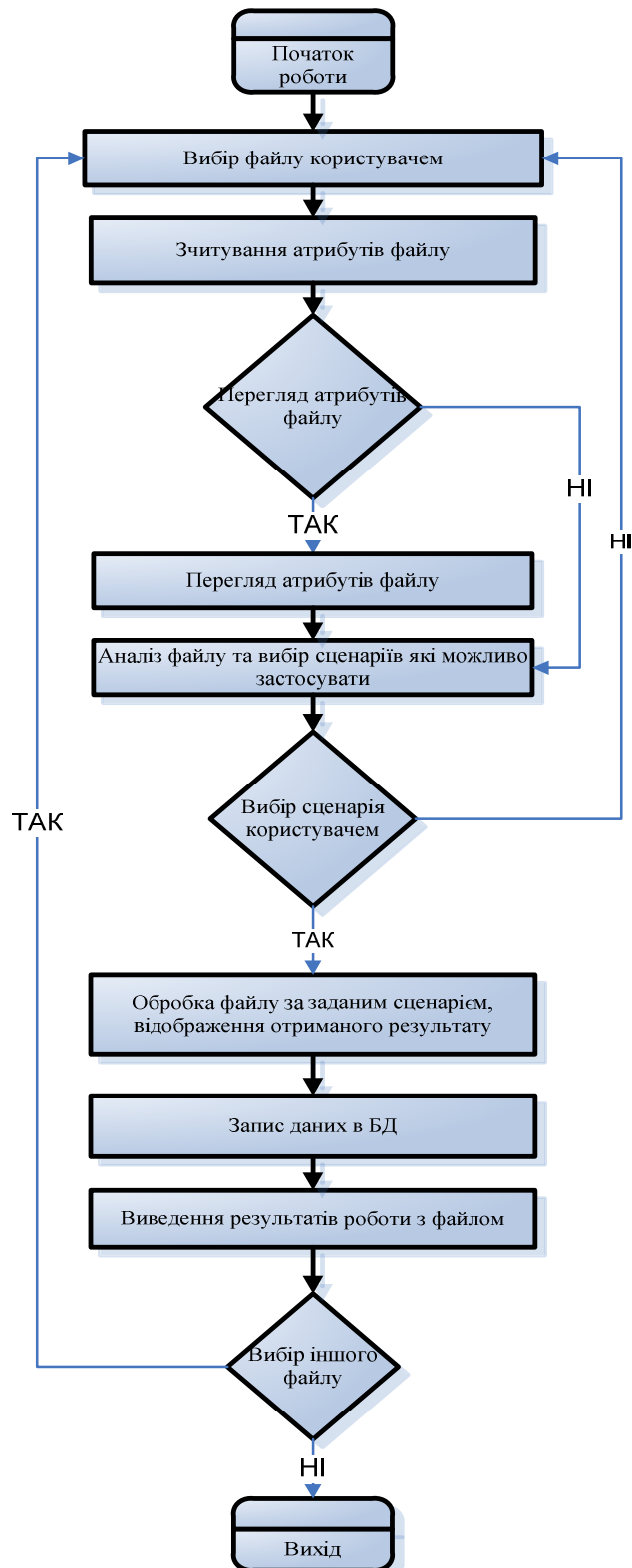


Рис. 2. Алгоритм функціонування програмного комплексу, що реалізує запропонований удосконалений метод ПВД

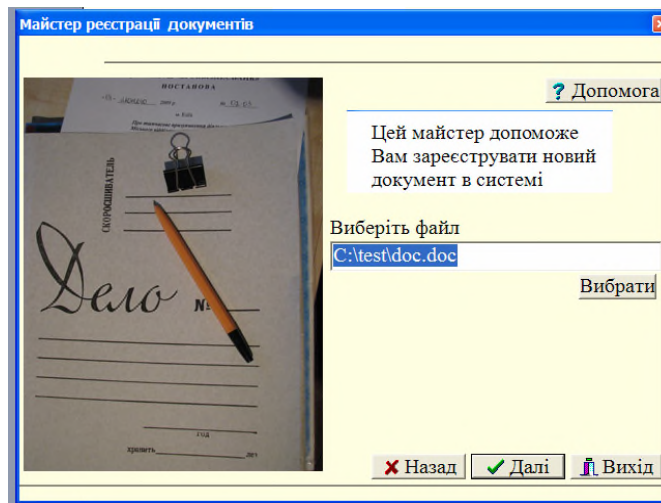


Рис. 3. Приклад екранних форм інтерфейсу користувача програмного комплексу.  
Вибір файлу користувачем

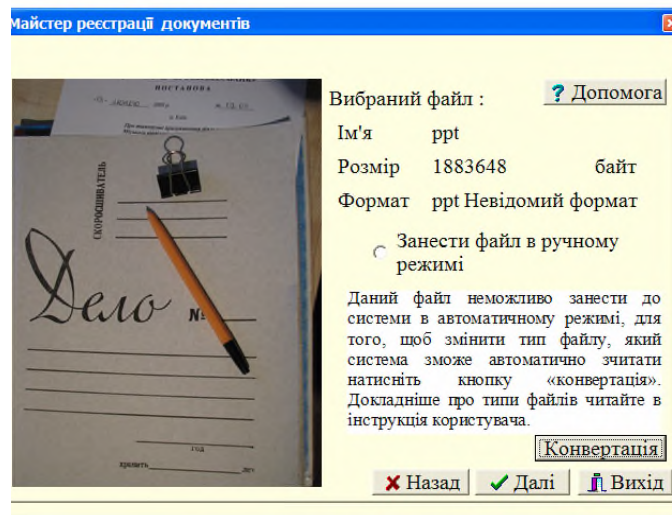


Рис. 4. Приклад екранних форм інтерфейсу користувача програмного комплексу.  
Відображення атрибутів вибраного файлу

Для реалізації програмного комплексу на базі удосконаленого методу ПВД була розроблена база даних (рис. 5), в якій у таблицях «Файл\_зображення», «Файл\_doc\_txt», «Файл\_gar», «Файл\_pdf», «Файл\_інше», зберігаються електронні копії документів. Таким чином, після визначення типу файлу, він автоматично заноситься до тієї таблиці, в якій зберігаються файли відповідного типу. Як приклад файл у форматі «doc», буде занесено в таблицю «Файл\_doc\_txt». Натомість у таблиці «Документ» зберігаються метадані документів. Такий варіант реалізації необхідний для реалізації сценаріїв обробки файлів і забезпечує оперативність пошуку необхідного документа, що відповідно зменшує навантаження на БД. Усі

інші таблиці БД є інформаційно-довідковими. Таблиці «Логін» і «Пароль» забезпечують регламент доступу до програмного комплексу.

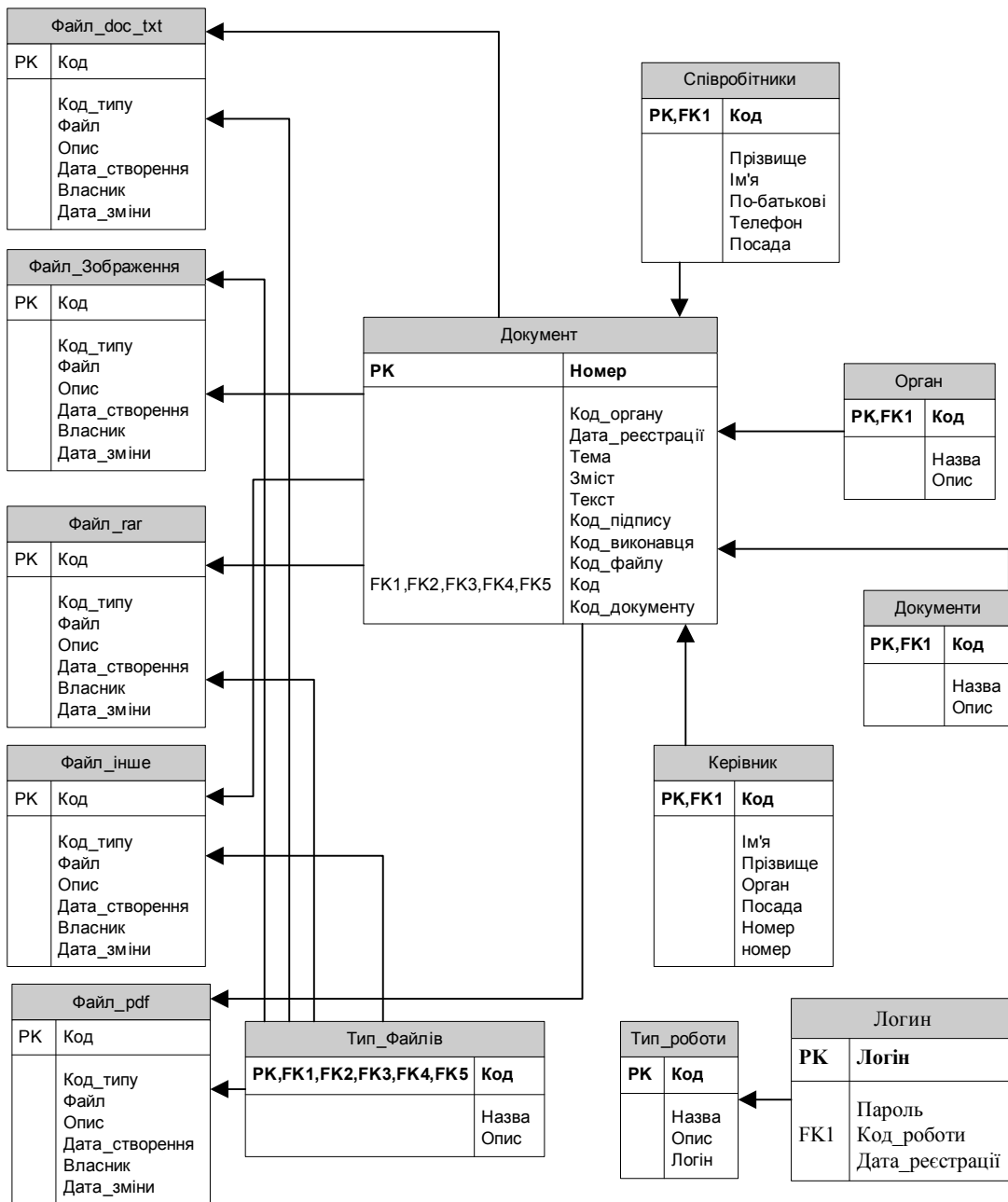


Рис. 5. Структура бази даних

Програмний комплекс був впроваджений як компонент КІАС АТ «Брокбізнесбанк». Він дозволив створити електронний архів нормативних документів банку, які вже існували у форматі електронних документів. На цей час програмний комплекс успішно використовується в роботі управління ризик-менеджменту АТ «Брокбізнесбанк», що значно спрощує пошук документів і дає змогу співробітникам

кам управління ризик-менеджменту АТ «Брокбізнесбанк» оперативно знаходити необхідні документи, таким чином заощаджуючи робочий час працівників.

## **Висновки**

Запропоновано удосконалений метод потокового введення документів, який вирішує задачу інтеграції до КІАС документів, графічні образи яких існують лише у форматі електронного документа. На базі цього методу розроблено алгоритм автоматизованої обробки зазначеного електронного документа та збереження метаданих і графічного образу документа в інформаційному сховищі КІАС. Документи розпізнаються та зберігаються як файли різних форматів. Було розроблено програмний комплекс, який реалізує зазначений алгоритм і використовується як компонент відповідної КІАС.

1. *Нестеренко О.В.* Основи побудови автоматизованих інформаційно-аналітичних систем органів державної влади / О.В. Нестеренко. — К.: Наук. думка, 2005. — 628 с.
2. *Электронные документы в корпоративных сетях: второе пришествие Гутенберга* / [Клименко С.В., Крохин И.В., Куц В.М., Гагутин Ю.Л.]. — Москва: Анкей: Экотрендз, 1999. — 271 с. — (Технологии электронных коммуникаций).
3. *Пашков Д.В.* Обзор современных систем автоматизации делопроизводства и документооборота / Д.В. Пашков // Документооборот и Делопроизводство. — 2004. — № 5. — С. 10–19.
4. *Хашковский В.В.* Разработка и исследование методов организации параллельно-конвейерных систем массового ввода / В.В. Хашковский. — Таганрог: ООО «Антон», 2002. — 166 с.
5. *Юмин И.Ф.* Электронные документы и архивы: теория и практика / И.Ф. Юмин // Отеч. арх. — 1999. — № 3. — С. 109–112 .
6. *Англо-український тлумачний словник з обчислювальної техніки, Інтернету і програмування.* [авт.-уклад. Е.М. Пройдаков, Л.А. Теплицкий]. — К.: Видавничий дім «СофтПрес», 2006. — 824 с.

Надійшла до редакції 17.06.2009