



УКРАЇНА

(19) UA (11) 48350 (13) U
(51) МПК (2009)
G06K 9/62МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІОПИС
ДО ПАТЕНТУ
НА КОРИСНУ МОДЕЛЬвидається під
відповідальність
власника
патенту

(54) СПОСІБ ПЕРЕВЕДЕННЯ ДАНИХ З ПАПЕРОВИХ НОСІЇВ У ЕЛЕКТРОННИЙ ВИГЛЯД

1

(21) u200910690

(22) 22.10.2009

(24) 10.03.2010

(46) 10.03.2010, Бюл.№ 5, 2010 р.

(72) ЧУХЛІБ ЄВГЕН ОЛЕКСАНДРОВИЧ

(73) ДЕРЖАВНЕ ПІДПРИЄМСТВО "ДОКУМЕНТ"

(57) 1. Спосіб переведення даних з паперових носіїв у електронний вигляд, що включає сканування даних з паперових носіїв, який відрізняється тим, що перед скануванням здійснюють перебір карто- теки і наклейку на паперові носії штрихових кодів, сканування підготовлених паперових носіїв здійс- нюють за допомогою високошвидкісних сканерів, використовуючи програмне забезпечення ABBY, з наступним автоматичним розпізнаванням графіч- ного образу тексту документа, скановані дані пе- реносяться на змінні HDD-носії, які передають до центру обробки інформації, де дані зі змінних HDD-носіїв вносять до серверної частини компле- ксу в підсистему розпізнавання/обробки сканова- ного зображення, внесені дані автоматично обро- бляють за заданим алгоритмом, підготовлюють і передають для обробки у клієнтську частину комп- лексу - підсистему верифікації та обробки, за результатами обробки дані повертають до серве- рної частини центру обробки інформації, де їх до- датково перевіряють, зберігають, а оброблені дані

2

за вимогою користувача переносять на змінні HDD-носії у файли заданого типу та визначеної структури для подальшого збереження у центра- льній базі даних.

2. Спосіб за п. 1, який відрізняється тим, що у клієнтській частині комплексу здійснюють автома- тичну послідовну подачу у відповідне вікно ділянки зображення документа верифікатору для вводу інформації у відповідне поле.

3. Спосіб за п. 1, який відрізняється тим, що у клієнтській частині комплексу здійснюють автома- тичну перевірку введених даних документа, що складається з багатьох полів, у відповідності з попередньо визначеними правилами, із застосу- ванням: регулярних виразів, звірки по базах даних та словниках, крос-перевірку зв'язаних полів.

4. Спосіб за п. 1, який відрізняється тим, що у клієнтській частині комплексу здійснюють аналіз місць розміщення фото з наступним занесенням в базу.

5. Спосіб за п. 1, який відрізняється тим, що у клієнтській частині за допомогою клавіатури вно- сять значення рукописного тексту, що підсвічуєть- ся, на графічному зображенні картки в задане по- ле введення, при цьому здійснюють автоматичну перевірку по різноманітних довідниках та прив'язку до цих даних фотографій з форм.

Корисна модель належить до способів зчиту- вання та розпізнавання печатних чи рукописних текстів або графічних зображень з паперових носіїв та переведення їх у електронний вигляд, і може бути використана при зчитуванні заяв на отриман- ня внутрішнього паспорту, талонів реєстрації і зняття з реєстрації місця проживання з присутніми на них графічними зображеннями тощо.

Протягом багатьох століть основними носіями інформації, що використовувалися як для підгото- вки різних документів, так і для їх обробки та збе- рігання, був папір.

У наш час в багатьох технологіях паперові но- сії використовуються для підготовки вихідних до- кументів, їх обробки, переведення в електронний вигляд, тобто формування електронних версій даних.

Відомо спосіб переносу символічних даних з графічних образів на носій інформації комп'ютера цифровим кодом, який включає зорове зчитування символічних даних та перенесення їх в цифровому коді на носій інформації комп'ютера за допомогою літерно-цифрової клавіатури. Графічні образи сим- вольних даних, попередньо підготовлені у вигляді графічних файлів, відображають у першому вікні двовіконного інтерфейсу, зорове зчитування сим- вольних даних здійснюють з першого вікна двові- конного інтерфейсу, а перенесення літерно- цифрових позначень символічних даних на носій інформації комп'ютера здійснюють через друге вікно двовіконного інтерфейсу [UA № 58428, G06K 1/00 G06K 9/62 15.07.2003].

Найближчим до корисної моделі, що заявля- ється, є спосіб автоматичного введення даних у

(19) UA (11) 48350 (13) U

пам'ять комп'ютера з паперових документів, що включає сортування паперових документів і розділення їх на пачки, сканування та виділення полів зчитування відповідно до попередньо розробленого шаблону. У технології Imagewriter використовують візуальне розпізнавання та автоматичне введення для чорно-білих документів, заповнених суцільним текстом, в технології автоматичного розпізнавання - автоматичне розпізнавання та автоматичне введення друкованих документів [Мучник М.М. Технология ввода данных с бумажных документов, МСУ, компания "Сканпрог", Украина, [http://dSPACE.nbuv.gov.ua:8080/dSPACE/handle/123456789/1684].

Зазначений спосіб, як і попередній аналог, має досить вузькі функціональні можливості, довгий час переведення даних з паперових носіїв у електронний вигляд та не дає гарантії від помилок.

В основу корисної моделі поставлено задачу створити спосіб переведення даних з паперових носіїв у електронний вигляд, який би мав широкі функціональні можливості, прискорив би переведення даних з паперових носіїв у електронний вигляд та зменшив ймовірність помилок.

Поставлену задачу вирішують тим, що у спосіб переведення даних з паперових носіїв у електронний вигляд, який включає сканування даних з паперових носіїв, згідно з корисною моделлю, перед скануванням здійснюють перебір картотеки і наклейку на паперові носії штрихових кодів, сканування підготовлених паперових носіїв здійснюють за допомогою високошвидкісних сканерів, використовуючи програмне забезпечення ABBY, з наступним автоматичним розпізнаванням графічного образу тексту документа, скановані дані переносять на змінні HDD-носії, які передають до центру обробки інформації, де дані зі змінних HDD-носіїв вносять до серверної частини комплексу в підсистему розпізнавання/обробки сканованого зображення, внесені дані автоматично обробляють за заданим алгоритмом, підготовлюють і передають для обробки у клієнтську частину комплексу - підсистему верифікації та обробки, за результатами обробки дані повертають до серверної частини центру обробки інформації, де їх додатково перевіряють, зберігають, а оброблені дані за вимогою користувача переносять на змінні HDD-носії у файли заданого типу та визначеної структури для подальшого збереження у центральній базі даних.

У клієнтській частині комплексу здійснюють автоматичну послідовну подачу у відповідне вікно ділянки зображення документа верифікатору для вводу інформації у відповідне поле.

У клієнтській частині комплексу здійснюють автоматичну перевірку введених даних документа, що складається з багатьох полів, у відповідності з попередньо визначеними правилами, із застосуванням: регулярних виразів, зв'язки по базам даних та словникам, кросс-перевірку зв'язаних полів.

У клієнтській частині комплексу можуть здійснювати аналіз місць розміщення фото з наступним занесенням в базу.

У клієнтській частині за допомогою клавіатури можуть вносити значення рукописного тексту, що

підсвічується, на графічному зображенні картки в задане поле введення, при цьому здійснюють автоматичну перевірку по різноманітних довідниках та прив'язку до цих даних фотографій з форм.

Спосіб, що заявляється, має широкі функціональні можливості - дозволяє розпізнавати тип паперового носія, печатні символи, графічні зображення, виділяти зону рукописного тексту за заданими критеріями, обробляти рукописний текст за спеціальним алгоритмом, перевіряти результати обробки, зберігати дані у визначеному форматі, переносити дані за визначеними критеріями, вести статистичну звітність за результатами роботи системи та діями користувачів.

Спосіб дозволяє прискорити переведення даних з паперових носіїв у електронний вигляд, а також зменшує ймовірність помилок.

Корисна модель пояснюється схемою, де зображено систему переведення даних з паперових носіїв у електронний вигляд.

Система складається з мобільних комплексів сканування, які можуть працювати окремо у різних місцях, де зберігаються картотеки, та центрального комплексу обробки отриманих графічних зображень і збереження результатів обробки в заданому вигляді.

Система містить паперові носії 1 із штриховими кодами, високошвидкісні сканери 2, змінні HDD-носії 3, центр 4 обробки інформації і змінні HDD-носії 5.

Спосіб переведення даних з паперових носіїв, наприклад, паперових карток, які зберігаються у паспортних столах (Форма 1), талонів реєстрації та зняття з реєстрації (Адресно-довідкове бюро), в електронний вигляд здійснюють наступним чином.

Спочатку перебирають картотеку з паперовими носіями 1 та наклеюють на них штрихові коди.

Паперові носії 1 можуть бути різні за фактурою паперу (тонкий лист, картон) та їх зношеністю (наприклад, паперові носії, яким кілька десятків років). Причому інформація на паперових носіях 1 може знаходитися з обох сторін. А, наприклад, на заяві на отримання паспорта також з обох боків знаходяться фото громадян, зображення яких виділяють окремо для зберігання у електронній базі даних, для створення у подальшому пошукової системи за зображенням.

Після наклейки штрихових кодів на паперові носії 1 облікових картотек їх сканують за допомогою високошвидкісних сканерів 2 під керуванням відповідного програмного забезпечення ABBY з наступним автоматичним розпізнаванням графічного образу тексту документа.

Сформований PDF файл сканованого двостороннього зображення паперового носія 1 переносять у заданий каталог файлової системи комп'ютера. Ім'я файлу відповідає номеру наклеєного штрихового коду.

Скановані дані переносять на змінні HDD-носії 3.

Змінні HDD-носії 3 передають до центру 4 обробки інформації, де дані зі змінних HDD-носіїв 3 вносять до серверної частини комплексу в підсистему розпізнавання/обробки сканованого зображення. Внесені дані автоматично обробляють за

заданим алгоритмом, підготовлюють і передають для обробки у клієнтську частину комплексу - підсистему верифікації та обробки.

У клієнтській частині комплексу здійснюють наступні операції:

- автоматичну послідовну подачу у відповідне вікно ділянки зображення документа верифікатору для вводу інформації у відповідне поле;

- автоматичну перевірку введених даних документа, що складається з багатьох полів, у відповідності з попередньо визначеними правилами, із застосуванням: регулярних виразів, звірки по базах даних та словникам, кросс-перевірку зв'язаних полів;

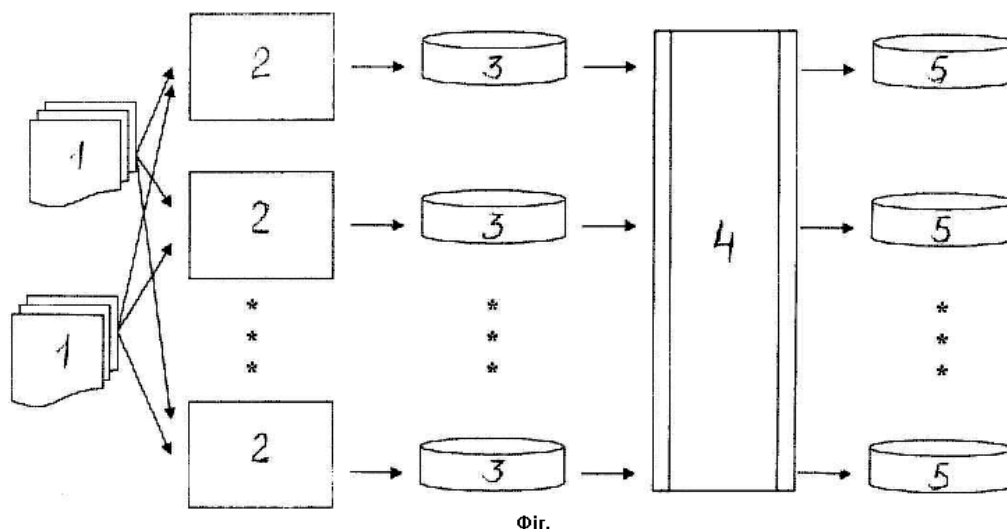
- при необхідності перевірку даних вручну в режимах: групової перевірки фотографій, контекстної перевірки полів, цілісної перевірки розпізнаного документа;

- у разі необхідності передачу документа в чергу виключень для обробки старшим верифікатором, якщо в документі за будь-яких причин є критичні помилки;

Оператор клієнтської частини за допомогою клавіатури вносить значення рукописного тексту, що підсвічується, на графічному зображенні картки в задане поле введення. При цьому відбувається автоматична перевірка по різноманітних довідниках та прив'язка, при потребі, до цих даних фотографій з форм.

За результатами обробки дані повертають до серверної частини центру 4 обробки інформації, де їх додатково перевіряють і зберігають.

Оброблені дані за вимогою користувача переносять на змінні HDD-носії 5 у файли заданого типу та визначеної структури для подальшого збереження у центральній базі даних (не показано).



Фіг.