



**ФЕДЕРАЛЬНАЯ СЛУЖБА  
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**

**(12) ЗАЯВКА НА ИЗОБРЕТЕНИЕ**

(21)(22) Заявка: 2014112237/08, 31.03.2014

Приоритет(ы):

(22) Дата подачи заявки: 31.03.2014

(43) Дата публикации заявки: 10.10.2015 Бюл. № 28

Адрес для переписки:

119019, Москва, Гоголевский б-р, 11, этаж 3,  
"Гоулингз Интернэшнл Инк.", Соболеву  
Александрю Юрьевичу

(71) Заявитель(и):

Общество с ограниченной ответственностью  
"Аби Девелопмент" (RU)

(72) Автор(ы):

Панферов Василий Владимирович (RU),  
Исаев Андрей Анатольевич (RU)

**(54) ВВОД ДАННЫХ С ИЗОБРАЖЕНИЙ ДОКУМЕНТОВ С ФИКСИРОВАННОЙ СТРУКТУРОЙ****(57) Формула изобретения**

1. Способ ввода данных с изображения документа с фиксированной структурой с помощью процессорного устройства, способ содержит операции:

получение электронного изображения документа с фиксированной структурой;

распознавание ключевых слов на изображении документа;

идентификацию, по меньшей мере, одного опорного объекта на основе распознанных ключевых слов, где, по меньшей мере, один опорный объект указывает, по меньшей мере, на одну область изображения документа с фиксированной структурой, откуда следует выполнить ввод данных;

наложение нескольких шаблонов из множества шаблонов на документ с фиксированной структурой с помощью идентифицированного, по меньшей мере, одного опорного объекта;

выбор шаблона из множества шаблонов с помощью качества одного или более распознанных ключевых слов; и

извлечение данных из изображения документа с помощью выбранного шаблона.

2. Способ по п. 1, дополнительно содержащий выполнение коррекции искажений изображения документа с фиксированной структурой.

3. Способ по п. 2, в котором выполняется коррекция искажений, содержащая, по меньшей мере, одно из следующих:

выравнивание линий, коррекция перекоса, коррекция геометрии документа, цветовая коррекция, восстановление смазанных и нефокусированных областей и удаление шума.

4. Способ по п. 2, в котором выполняется коррекция искажений, содержащая обнаружение границ на изображении документа с фиксированной структурой.

5. Способ по п. 4, дополнительно содержащий обрезку изображения вдоль обнаруженных границ.

6. Способ по п. 1, дополнительно содержащий применение, по меньшей мере, одного фильтра к изображению документа с фиксированной структурой.

7. Способ по п. 1, дополнительно содержащий определение типа документа на основе выбранного шаблона.

8. Способ по п. 7, в котором выбор шаблона основан на показателе качества наложения шаблона.

9. Способ по п. 1, в котором опорным объектом является регулярное выражение.

10. Способ по п. 1, в котором распознавание изображения документа основано на дополнительной информации о распознанных ключевых словах.

11. Способ по п. 1, дополнительно содержащий:

вычисление для каждого шаблона из множества шаблонов показателя качества наложения шаблона на изображение документа с фиксированной структурой и идентификацию шаблонов из множества шаблонов, качество наложения которых выше предварительно заданного порога;

сохранение идентифицированных шаблонов.

12. Способ по п. 1, дополнительно содержащий вычисление показателя качества распознанного текста в распознанном ключевом слове.

13. Способ по п. 12, дополнительно содержащий экспорт распознанного текста, если качество распознанного текста превышает пороговое значение.

14. Способ по п. 1, в котором множество шаблонов содержит, по меньшей мере, один предварительно созданный шаблон.

15. Способ по п. 1, дополнительно содержащий создание одного или более шаблонов на основе, по меньшей мере, одного идентифицированного опорного объекта.

16. Способ по п. 1, дополнительно содержащий распознавание изображения документа с помощью выбранного шаблона.

17. Система ввода данных с изображения документа с фиксированной структурой с помощью процессорного устройства, причем система содержит:

процессорное устройство, причем процессорное устройство выполняет:

получение электронного изображения документа с фиксированной структурой;

распознавание ключевых слов на изображении документа;

идентифицирование, по меньшей мере, одного опорного объекта на основе распознанных ключевых слов, где, по меньшей мере, один опорный объект указывает, по меньшей мере, на одну область изображения документа с фиксированной структурой, откуда следует выполнить ввод данных;

наложение нескольких шаблонов из множества шаблонов документа с фиксированной структурой с помощью идентифицированного, по меньшей мере, одного опорного объекта;

выбор шаблона из множества шаблонов с помощью качества распознанных ключевых слов; и

извлечение данных из изображения документа с помощью выбранного шаблона.

18. Система по п. 17, дополнительно содержащая выполнение коррекции искажений изображения документа с фиксированной структурой.

19. Система по п. 18, в которой выполняется коррекция искажений, содержащая, по меньшей мере, одно из: выравнивание линий, коррекция перекоса, коррекция геометрии документа, коррекция цветов, восстановление смазанных и нефокусированных областей и удаление шума.

20. Система по п. 18, в которой выполняется коррекция искажений, содержащая обнаружение границ на изображении документа с фиксированной структурой.

21. Система по п. 20, дополнительно содержащая обрезку изображения вдоль обнаруженных границ.

22. Система по п. 17, дополнительно содержащая применение, по меньшей мере, одного фильтра к изображению документа с фиксированной структурой.

23. Система по п. 17, дополнительно содержащая определение типа документа на основе выбранного шаблона.

24. Система по п. 23, в которой выбор шаблона основан на показателе качества наложения шаблона.

25. Система по п. 17, в которой опорным объектом является регулярное выражение.

26. Система по п. 17, в которой распознавание изображения документа основано на дополнительной информации о распознанных ключевых словах.

27. Система по п. 17, дополнительно содержащая вычисление для каждого шаблона из множества шаблонов показателя качества наложения шаблона на изображение документа с фиксированной структурой и идентификацию шаблонов из множества шаблонов, качество наложения которых выше предварительно заданного порога;

сохранение идентифицированных шаблонов.

28. Система по п. 17, дополнительно содержащая вычисление показателя качества распознанного текста в распознанном ключевом слове.

29. Система по п. 28, дополнительно содержащая экспорт распознанного текста, если качество распознанного текста превышает пороговое значение.

30. Система по п. 17, в которой множество шаблонов содержит, по меньшей мере, один предварительно созданный шаблон.

31. Система по п. 17, дополнительно содержащая создание одного или более шаблонов на основе, по меньшей мере, одного идентифицированного опорного объекта.

32. Система по п. 17, дополнительно содержащая распознавание изображения документа с помощью выбранного шаблона.

33. Машиночитаемый носитель данных, имеющий хранящиеся на нем команды, исполнение которых процессором приводит к выполнению процессором операций, содержащих:

получение электронного изображения документа с фиксированной структурой;

распознавание ключевых слов в изображении документа;

идентификацию, по меньшей мере, одного опорного объекта на основе распознанных ключевых слов, где, по меньшей мере, один опорный объект указывает, по меньшей мере, на одну область изображения документа с фиксированной структурой, откуда следует выполнить ввод данных;

наложение нескольких шаблонов из множества шаблонов для документа с фиксированной структурой с помощью идентифицированного, по меньшей мере, одного опорного объекта;

выбор шаблона из множества шаблонов с помощью качества распознанных ключевых слов; и

извлечение данных из изображения документа с помощью выбранного шаблона.

34. Машиночитаемый носитель данных по п. 33, дополнительно содержащий выполнение коррекции искажений изображения документа с фиксированной структурой.

35. Машиночитаемый носитель данных по п. 34, в котором выполняется коррекция искажений, содержащая, по меньшей мере, одно из: выравнивание линий, коррекция перекоса, коррекция геометрии документа, цветовая коррекция, восстановление смазанных и несфокусированных областей и удаление шума.

36. Машиночитаемый носитель данных по п. 34, в котором выполняется коррекция искажений, содержащая обнаружение границ на изображении документа с фиксированной структурой.

37. Машиночитаемый носитель данных по п. 36, дополнительно содержащий обрезку изображения вдоль обнаруженных границ.

38. Машиночитаемый носитель данных по п. 33, дополнительно содержащий

применение, по меньшей мере, одного фильтра к изображению документа с фиксированной структурой.

39. Машиночитаемый носитель данных по п. 33, дополнительно содержащий определение типа документа на основе выбранного шаблона.

40. Машиночитаемый носитель данных по п. 39, в котором выбор шаблона основан на показателе качества наложения шаблона.

41. Машиночитаемый носитель данных по п. 33, в котором опорный объект является регулярным выражением.

42. Машиночитаемый носитель данных по п. 33, в котором распознавание изображения документа основано на дополнительной информации о распознанных ключевых словах.

43. Машиночитаемый носитель данных по п. 33, дополнительно содержащий вычисление для каждого шаблона из множества шаблонов качества наложения шаблона на изображение документа с фиксированной структурой и идентификацию шаблонов из множества шаблонов, качество наложения которых выше предварительно заданного порога;

сохранение идентифицированных шаблонов.

44. Машиночитаемый носитель данных по п. 33, дополнительно содержащий вычисление показателя качества распознанного текста в распознанных ключевых словах.

45. Машиночитаемый носитель данных по п. 44, дополнительно содержащий экспорт распознанного текста, если качество распознанного текста превышает пороговое значение.

46. Машиночитаемый носитель данных по п. 33, в котором множество шаблонов содержит, по меньшей мере, один предварительно созданный шаблон.

47. Машиночитаемый носитель данных по п. 33, дополнительно содержащий создание одного или более шаблонов на основе, по меньшей мере, одного идентифицированного опорного объекта.

48. Машиночитаемый носитель данных по п. 33, дополнительно содержащий распознавание изображения документа с помощью выбранного шаблона.

RU 2014112237 A

RU 2014112237 A