

FORM 2

THE PATENTS ACT, 1970

(39 of 1970)

&

The Patent Rules, 2003

COMPLETE SPECIFICATION

(See section 10 and rule 13)

TITLE OF THE INVENTION

“Enhancing accuracy and speed through designing at the early stages Robotic
Process Automation using Machine Learning”

We, applicant(s)

NAME	NATIONALITY	ADDRESS
1. Dr. Veena N	Indian	Associate Professor, Department of ISE, BMS Institute of Technology and Management, Avallahalli, Bangalore - 560064, Karnataka
2. Dr.Sandeep Bhat	Indian	Associate Professor, Department of Electronics and Communication Engineering, Sahyadri college of Engineering and Management, Mangaluru,575007
3. Mr.Pankaj Savita	Indian	Assistant Professor, Department of Information Technology, Lakshmi Narain college of Technology, Bhopal, Madhya Pradesh, India

4. Dr. Chandra Prakash	Indian	Professor, Department of Electronics and Communication Engineering, Rajshree Institute of Management and Technology, Bareilly, Uttar Pradesh 243122
5. Dr.M.K.Kirubakaran	Indian	Associate Professor, Department of Information Technology, St.joseph's institute of technology Old Mahabalipuram Rd, Kamaraj Nagar, Semmancheri, Chennai, Tamil Nadu 600119
6. Dr. Vikas Tripathi	Indian	Associate Professor, Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India, 248002
7. Dr.B.Gomathi	Indian	Associate Professor, Department of Computer Science, Shri Shankarlal Sundarbai Shasun Jain College For Women, Chennai-17

The following specification particularly describes the nature of the invention and the manner in which it is performed:

FIELD OF INVENTION:

In robotic process automation, the following description pertains to a product, technique, and system (RPA).

Background of the invention:

5 Robotic process automation, often known as RPA, is a technology that allows the automation of operations that require a lot of human labor and are performed repeatedly. In robotic process automation (RPA), a computer system or robot may imitate the activities of a human person to carry out a work based on a computer. In other words, RPA may be used to communicate with application software (or application, for short) via its user interface in the same way that a
10 person would engage with it. Because of this, it is not essential to connect RPA with the current applications on a programming level. This removes one of the challenges inherently associated with integration, which is the process of bringing together various components.

Blue Prism is one example of a program that may be used for RPA. Blue Prism offers a variety of connectors, sometimes known as libraries, that are designed to facilitate communication with
15 certain varieties of application user interfaces. For example, Blue Prism has an HTML connector, making it possible to deal with HTML, JavaScript, and other web interface components that are often used. Blue Prism can acquire a web page's source code and use it in various ways. For instance, it may use the code to identify and recognize various web page components, such as the search button. Blue Prism comprises various modules, each of which
20 is tailored to the user interface requirements of a particular application.

Therefore, there is a need for universal RPA, which refers to RPA in which a robot is capable of interacting with a range of user interfaces without the requirement of application-specific modules. Particularly, there is a need for RPA to be environment independent. This means it should be able to automate the execution of operations and/or commands regardless of the technical operation of a program and/or program module. One example of this would be using a remote desktop connection technology that sends an image stream without knowing the applications running in the background.

Summary of the present invention:

According to one of the aspects, a computer system is said to be given. The computer system may include a memory and a processor that are programmed to perform the following operations: acquire information relating to images (such as screenshots) or a stream of images shown on display; analyze the information relating to the images or the stream of images to detect activities on display; record activity information about the detected activities on display, and generate output data comprising a combination of the information relating to images or the stream of images shown on display.

In the context of RPA, this element of the invention relates to recording a process while a user is carrying it out. Interacting with an application's user interface is one way a user may successfully carry out a job. A user interface may be shown on the screen display of a computer,

and the processor may gather information related to what is displayed on the screen. What is shown on the screen may be seen in pictures that are separate, individual entities or in the form of a stream of images, such as a series of merged images or a video stream? Either style of presentation can be monitored. The word "images" may be used to refer to both individual
5 pictures and one or more of the images included in the stream of images in the following paragraphs.

One possible configuration for the computer system used to carry out the information collection is for it to include the display of the computer screen as an integral component or component of the system. In a further example, the screen might be a component of a distinct computer
10 system that is accessible in one of many ways, including the use of a remote desktop connection technology and/or a virtual machine (which may be running on the same computer system or a different computer system).

The photographs may provide an exemplary representation of the user interface and other components depicting the user's interaction with the user interface. The acquisition of the
15 information relating to the images can be accomplished in a couple of different ways: first, by directly capturing the images as they appear on display (for example, by taking screenshots), and second, by collecting information about the images on display from sources other than the display itself.

Information may be especially acquired by computer vision while capturing images. This may include tracking mouse movements, the mouse pointer turning into a character prompt blinking in a text field, and/or characters appearing on the screen, or buttons showing visual alteration when pressed.

- 5 During the process of carrying out the assignment, the photographs can be taken in an (at least partially) continuous fashion, or they may only be taken at certain points in time. For instance, the acquisition may occur at predetermined intervals or after a triggering event. It may also take place both of these ways. Specifically, it is possible to acquire images each time an activity is detected on display. This activity could be the result of the user acting, such as clicking on a
- 10 button on the user interface, or it could be the result of an application acting, such as a web browser loading a web page or a dialogue box appearing. One illustration shows how the capture of the photographs may be done on the device's own, namely by sending a stream of images to the display (e.g., accessing the feed of output images from the graphics card to the display). The capture might also be done by an external picture acquisition device, such as a
- 15 camera focused on the display, as shown in a different illustration.

- In addition to this, or as an alternative, the CPU might be programmed to get information about the pictures being shown on display from other sources. For instance, the extra parts shown in the photographs can include a pointer attached to a pointing device (e.g., a mouse). The underlying operating system or the drivers managing the peripheral input devices, including
- 20 the pointing device, may provide the processor with information on extra elements like the

pointer. Alternatively, the CPU might get this information directly. For instance, the information that has been gathered may include the position of the pointer and/or an action carried out by the pointer, such as choosing a portion of the picture being shown on the screen.

The various stages of the operation may match the various activities that are shown on the

5 screen of the computer. For instance, the commencement of the procedure may coincide with the beginning of the process of launching a new program, which indicates the beginning of a new process that is operating. Each of these additional applications, which together make up a process, may execute simultaneously with or after the first one. As a result, a process change can occur whenever a new program is launched or while moving from one application to

10 another. A user may provide input to the user interface, yet another illustration of the technique.

This may be done, for instance, by using a peripheral input device such as a keyboard or a mouse. The user input may therefore include typing, which refers to the process of entering text by pressing keys on a keyboard or using a pointer of a pointing device (such as a mouse) for performing actions such as clicking, dragging, and/or scrolling. Typing refers to entering

15 text by pressing keys on a keyboard. In yet another illustration, engaging with the user interface may result in a scene change. This refers to a shift in the pictures shown on the screen and might happen, for instance, when a user clicks on a link on one web page to go to another. You may have heard someone refer to a scene change as a screen change.

It is possible to analyze the gathered information about the pictures to identify activities taking

20 place on display. For instance, the data that has been saved in the form of screenshots may be

evaluated to identify whether or not a scene change has taken place, that is, whether or not there is a difference between two screenshots that have been taken in succession. The use of computer vision methodologies to carry out this analysis is recommended. The term "computer vision" refers to a subfield of information technology that focuses on the processing, interpreting, and analyzing visual data by a computer. Specifically, computer vision may replicate human vision's capabilities by digitally seeing and comprehending a picture. The use of computer vision might thus make it possible to examine the data associated with the pictures to identify actions taking place on display.

In addition to this, or as an alternative, the information about the pictures on the display collected from sources that are not the display itself may be studied to determine the activities that are taking place on display. If this information includes an indication of an action done by something like a mouse pointer, for instance, then the activity of the "user input" kind may be identified.

The processor could record the information that has been obtained. For instance, the information pertaining to the actions, also known as activity information, as well as the information pertaining to the photos, may be saved in a log file located in the memory.

After the activity information has been captured, the processor might be programmed to provide output data consisting of a combination of the information about the pictures and the activity information. This would be done after the activity information has been recorded.

According to one illustrative scenario, a screenshot may either be placed onto or surround the activity information in a format that is legible by humans (e.g., natural text language). For example, a screenshot that displays the mouse pointer over a button could be edited to include a written instruction that details the action being carried out at the time the screenshot was being taken. This instruction could say something like "Left Click" to indicate that the button was activated by clicking the left mouse button. Another illustration shows that a set of computer instructions to duplicate an activity automatically may be contained with a screenshot or overlaid on it. This allows the activity to be reproduced in an exact manner every time.

10

Brief description of the invention:

In the text to follow, a comprehensive explanation of several cases will be provided concerning the pictures. It is important to remember that the examples shown might be altered in various ways. The components of one example may be mixed with those of another example and utilized to generate brand new instances unless it is specifically stated otherwise.

15

By the present invention, the example architecture for RPA shown in Figure 1 is as follows: A monitoring agent, an executing agent, a display, a memory, and a processor are all possible components of the RPA computer system (not shown). For example, the monitoring and

executing agents may be two separate entities in this scenario. Another example is if the monitoring agent and the executing agent are the same people.

The monitoring agent may be programmed to acquire information related to images and/or a stream of images shown on the display. At the same time, a user performs operations in the computer system, provided that the computing operations have a direct counterpart on display.

This information may be gathered while the user performs operations in the computer system.

It is important to realize that the information obtained from the display may be based on pictures that are separate, independent entities or on a stream of images that may be continuous or discrete. One example of the latter would be a series of merged images or a video stream. In

the following, the word "images" may be used to refer not just to the individual photographs but also to one or more of the images included in the stream of images. The monitoring agent can examine the information that is related to the pictures to identify the activity on display.

The monitoring agent may then record activity information about the observed activities on display and create output data consisting of a combination of the information about the pictures and the activity information. This output data may then be displayed.

When the executing agent interacts with the display, the data from the output may be utilized by the executing agencies to repeat the activities that the user has already completed in an autonomous and automated manner. To put it another way, the executing agent may read the output data and, based on the output data, may recreate the activities on the display that corresponds to the steps that must be taken to accomplish a job.

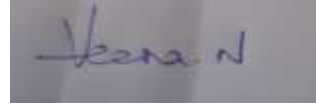
The monitoring agent could keep an eye on the display of a computer while a user is carrying out a job on the device, like sending an email, for instance. In particular, the monitoring agent may gather information about the photos by capturing the images transmitted to the display, such as from a stream of images supplied therein and/or in the form of screenshots. In either case, the agent may take the images as it sends them. It is also possible for the monitoring agent to record the visuals on the screen using a camera (for example, a CCD camera). This may be done instead of or in addition to the previous option. To be more specific, the display may show the steps performed by a user sitting in front of the display to carry out the task, for example, by operating a pointing device (such as a mouse) and/or inputting commands and/or information via an input device such as a keyboard, a gesture recognition device, a touchpad, a scanner, an eye tracker, or something similar. For example, the display may show the steps performed by a user sitting in front of the display to carry out the; In particular, the camera may be an integral component of or be integrated into smart glasses (such as Google glasses). For instance, sending an email may cause one or more of the following actions to be triggered or entered, for example, by a user and, as a result, shown or presented on display.

We Claim:

1. A computer system that includes a memory and a processor and is programmed to:
acquire information relating to images or a stream of images that are shown on display;
analyze the information relating to the images or the stream of images to detect
activities on display; record activity information about the activities that were detected
on display, and generate output data that includes a combination of the information
relating to the images or the stream of images and the activity information.
2. The computer system, according to claim 1, wherein the output data is generated in
such a way that the combination of the information relating to the images or the stream
of images and the activity information comprises at least one image or a sequence of
images with superimposed and/or enclosed activity information in a format that is
readable by humans, particularly one that includes natural text language.
3. The display is an additional computer system component that fulfills the requirements
of any previous claims.
4. The computer system, according to any one of the claims that came before this one,
wherein the activities that take place on display include at least one of the following: a
process change; a scene change; and a user input; and wherein the user input, in
particular, includes at least one of typing and using a pointer from a pointing device to
perform actions.

Dated this 21st day of June 2022

Signature:

A rectangular box containing a handwritten signature in blue ink. The signature appears to be 'Veena N'.

Applicant(s)

Dr. Veena N et. al.

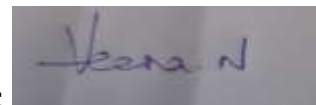
ABSTRACT

Enhancing accuracy and speed through designing at the early stages Robotic Process Automation using Machine Learning

A memory and a CPU are included in the computer system offered here. The computer
5 system is set up to perform the following tasks: acquire information relating to images that are
displayed on display; analyze the information relating to the images to identify activities that
are occurring on display; record activity information about the activities that have been
identified as occurring on display; and generate output data that is comprised of a combination
of the information relating to the images and the activity information. In addition, a technique
10 and a computer program product that correspond to the invention are disclosed.

Dated this 21st day of June 2022

Signature:



Applicant(s)

Dr. Veena N et. al.