A brief history of Optical Character Recognition (OCR)

Explore the past, present and future applications of the unsung hero of the business world.

By Martin Edmonds,
Pitney Bowes Business Manager, United Kingdom & Ireland

As you probably know, OCR is the process by which text images are converted into machine-encoded text. Digitising text means it can be easily presented, edited, stored and searched, optimising key administrative tasks such as invoicing and sales processing.

It is often used by business and finance teams to manage invoicing, but its detractors say that it slows down the process. At Pitney Bowes we're committed to making the whole invoicing process frictionless—allowing you to move from a state of *Argh* to feeling blissfully *Ahhh*.

But before we explore that, let's look at how the OCR was first developed, and whether it remains relevant.

The genius of Emanuel Goldberg

OCR traces its roots back to telegraphy. On the eve of the First World War, physicist Emanuel Goldberg invented a machine that could read characters and convert them into telegraph code. In the 1920s, he went a step further and created the first electronic document retrieval system.

At this time, businesses were microfilming financial records—great in principle, but quickly retrieving specific records from spools of film was nearly impossible. To overcome this, Goldberg used a photoelectric cell to do pattern recognition with the help of a movie projector. By repurposing existing technologies, he took the first steps toward the automation of record keeping. The U.S. patent for his "Statistical Machine" was later acquired by IBM.

Since then, OCR technology has proliferated, with businesses all over the world relying on it to help reduce overheads when it comes to extracting data from paper documents.

The challenges with OCR

Early versions of OCR had to be trained with images of each character and were limited to recognising one font at a time. In the 1970s, inventor Ray Kurzweil commercialised 'omni-font OCR', which could process text printed in almost any font. As technology advanced, so did the potential for document reading. By the early 2000s, OCR became available online as a cloud-based service, accessible via desktop and mobile applications.

Today, there are a host of OCR service providers offering technology (often accessible via APIs) capable of recognising most characters and fonts to a high level of accuracy. Even though the technology continues to improve, there is always a chance for errors, which requires costly human intervention to validate information. Not to mention, the continued use of paper comes with its own economic and environmental cost.

What has changed?

For decades, OCR was the only way to turn printed paper documents into machine-readable text documents that could be processed by computers. It remains the tool of choice (outside of EDI and invoice portals) for converting paper invoices into extractable data that can be integrated into finance systems.

Thankfully, e-document submission now offers businesses a far superior approach to areas such as invoicing and sales processing. Automation comes with a variety of benefits, including lowered costs and the ability to free up staff to focus on higher value tasks. Going forward, we expect advances in AI and Machine Learning to further accelerate the demise of data extraction.

From Argh to Ahhh.

The future of OCR

OCR will continue to be a valuable tool for filling in gaps when an application-generated electronic document cannot be created. Ultimately, the truly paperless business doesn't (yet) exist and data extraction is still a useful tool that can augment e-document processing.

That's why Pitney Bowes has developed a pain-free digital invoice processing platform to help clients who are grappling with paper. We use an analytical capture programming language that allows us to extract and process data from any form of digital invoice, removing the need for unreliable OCR.

Our pain-free digital invoice processing solution provides 100% accuracy, guaranteed. There's no need for data input, no time wasted and no costly mistakes. **By switching to our solution, you can remove manual data entry almost entirely and maximise your inbound processing, creating savings in excess of 80%.**[*]
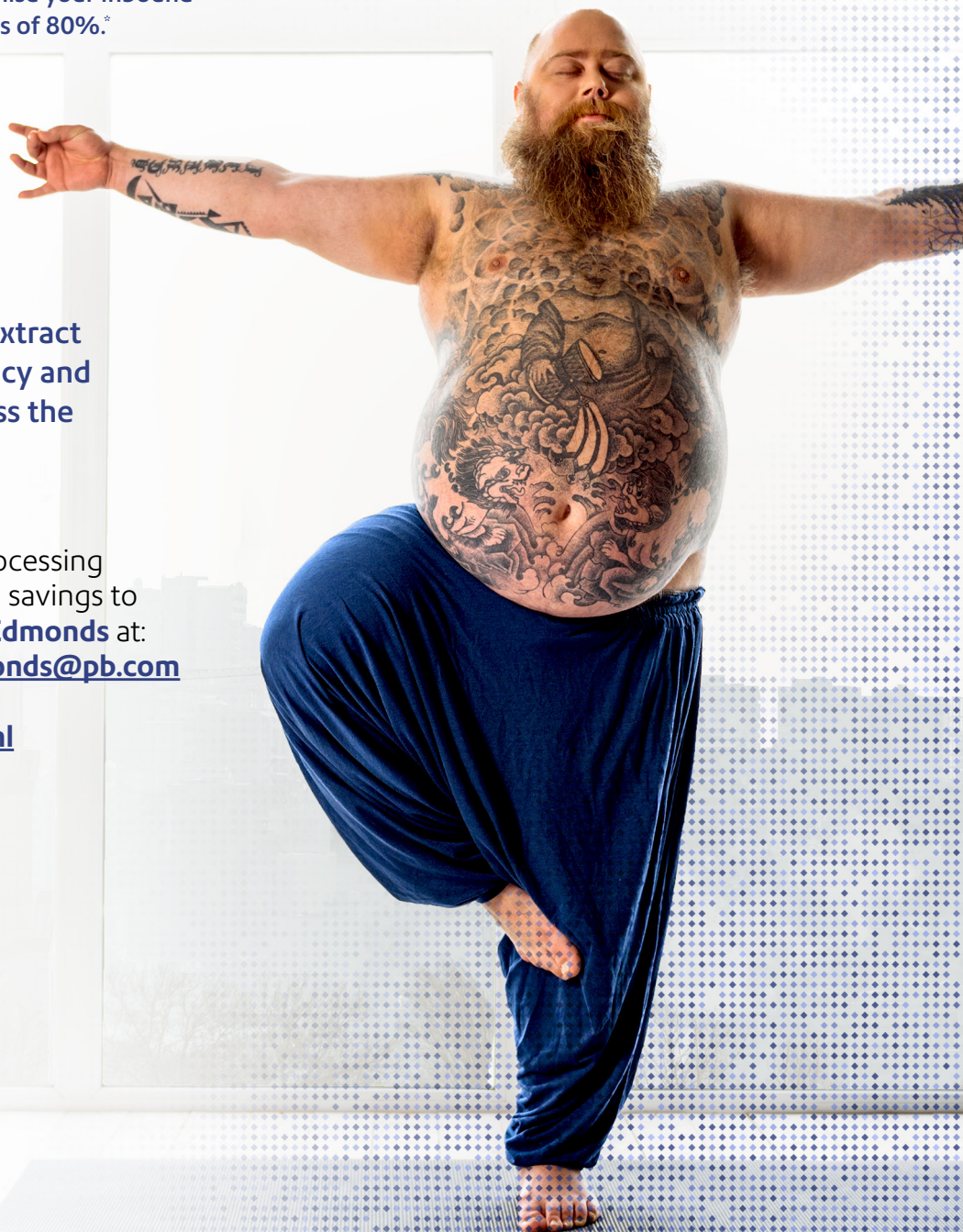
[*]Source to be provided

We will manage our technology on your behalf, around the clock, taking full responsibility for the accuracy of the processed data, so you can be certain the exact information you need is extracted from your invoices. Our software picks out errors, populates data into your system and offers you reporting and analytical information.

When you leverage the capabilities of automated data capture, you can make decisions more quickly, increase overall agility, and free up your finance team from repetitive manual tasks. The advantages you realise will benefit your business for years to come.

Our solution enables you to extract invoice data with total accuracy and makes it easy to share it across the entire organisation.

For more information on how Pitney Bowes Digital Invoice Processing can bring more efficiencies and savings to your business, contact **Martin Edmonds** at: **07712549318** or **martin.edmonds@pb.com** or visit: **pitneybowes.com/uk/ digital-invoice-processing.html**

pitney bowes