

Inteligência artificial na digitalização de documentos

Iris Viturino Guedes¹, Daniela Carvalho²

¹Graduando do curso de Sistema de Informação da Pontifícia Universidade Católica de Minas Gerais (PUC MINAS) - Uberlândia – MG – Brazil.

²Orientadora e professora do curso de Sistema de Informação da Pontifícia Universidade Católica de Minas Gerais (PUC MINAS) - Uberlândia – MG – Brazil.

irisguedes@hotmail.cm, danicarvalho@gmail.com

***Abstract.** Innovation and entrepreneurship are concepts already used for survival in increasingly competitive markets, and artificial intelligence (AI) is already being implemented in several segments, in partnership with IT and different business platforms, fostering new protocols and processes, as well as as quality and specialization, with numerous advantages and benefits. Thus, the objective of this work was to discuss the need to implement the digitization of documents using artificial intelligence.*

***Resumo.** Inovação e empreendedorismo são conceitos já utilizados para a sobrevivência em mercados cada vez mais competitivos, sendo que a inteligência artificial (IA) já está sendo implantada em diversos segmentos, em parceria com TI e plataformas diversas de negócio, fomentando novos protocolos e processos, bem como qualidade e especialização, com inúmeras vantagens e benefícios. Desta forma, o objetivo deste trabalho foi discorrer sobre as necessidades de implementação da digitalização de documentos com uso de inteligência artificial.*

Introdução

A inteligência artificial (IA) é uma ferramenta que sugere a substituição da inteligência humana na análise de grande escala de dados ao nível complexo, porém, possui vantagens e desvantagens (NUNES, 2019).

A digitalização de documentos e a gestão dessa implementação recorre a inteligência artificial, fomentando novas estratégias nos negócios e inovação, tão necessárias à sobrevivência e vantagens competitivas no que compete ao mercado (ANDRADE, 2020).

Plataformas e softwares são implementados anualmente, visando-se eficiência nesta atividade de digitalização de diversos documentos, tais como fotos, imagens, textos, livros, formulários e registros, antes preenchidos, armazenados e gerenciados manualmente, dentro de inúmeros processos, os quais estão sendo renovados. Assim, há uma interação com a inteligência artificial em background, em prol da inovação dos negócios, e uma gestão mais eficiente (CEDEFOP, 2019).

Conforme o exposto acima, o objetivo deste trabalho foi avaliar o emprego de inteligência artificial na digitalização de documentos, independente do segmento ou negócio.

Desenvolvimento

A inteligência artificial (IA) se refere aos sistemas tecnológicos polivalentes, que substituem a inteligência humana ao nível de percepção visual, reconhecimento de voz, e tomada de decisões, indicando um surgimento de novas necessidades no ramo empregatício, quanto à qualidade. Cada vez mais as empresas estão implementando processos que envolvam a prototipagem e sistemas mais efetivos de comercialização de produtos, com ênfase em automatização, com expansão de mercados neste nicho, visando à inovação e eficiência (NUNES, 2019).

Alan Turing aplicou a inteligência artificial já em 1956, ano que a mesma foi considerada Ciência. A digitalização permite a exploração econômica de propriedades de informações de diferentes setores, indicando inovação, em prol de reestruturar processos e leva-los a uma maior eficiência. Há indicações que haverá uma inteligência colaborativa, com transformações nos relacionamentos entre funcionários e máquinas, e entre máquinas e clientes, aumentando a ordem de magnitude nos negócios. Observa-se, neste contexto, treinamento e gestão de ferramentas de inteligência artificial, sendo que as máquinas melhoram as capacidades de processamento e análise de altas quantidades de dados de inúmeras fontes em tempo real. Assim, percebe-se uma necessidade de fundamentação de nova mentalidade e visão; empreendimento; liderança; dados e habilidades complementares (RUSSEL; NORVIG, 2004).

A partir da IA, amplifica-se dados em tempo real, cria-se interfaces avançadas para as interações, e há compartilhamento de trabalho com robôs. Os professores de máquinas auxiliarão os processadores de linguagem natural e os tradutores de idiomas, ensinando algoritmos de IA a imitar comportamentos humanos; os tradutores explicarão os complexos algoritmos aos profissionais não técnicos; e a sustentação do sistema será realizada por ferramentas que facilitarão o funcionamento. A IA já supera em execução de tarefas nas empresas de diversas áreas, ajudando a diminuir a complexidade de um problema, identificando relações causais, e auxiliando na tomada de ações, que passa a ser objetiva e imparcial, sem gerar conflitos. Porém, existe risco das máquinas serem tendenciosas e discriminatórias. O novo Regulamento Geral sobre a Proteção de Dados (GDPR) na UE contém direitos de proteção aos usuários na tomada de decisões automatizada (KOENIGKAM et al., 2019).

O mercado de softwares corporativos com recursos de inteligência artificial movimentava cerca de 11 bilhões de dólares, com taxa de crescimento anual de 56%. É um crescimento em 44 vezes do volume do *Big Data* atual, e, desta forma, o gerenciamento inteligente nesse crescente volume de dados será cada vez mais primordial. A digitalização de documentos incorre aos fundamentos de Inteligência Artificial (IA), *Big Data* e Análise de Dados, cujas ferramentas garantem diferenciação no mercado. Assim, a digitalização de documentos envolve um estabelecimento digital descentralizado, ágil, colaborativo e flexível inerente à transformação dos negócios com base em inovação, sendo que a inteligência artificial já é destaque. A parceria entre especialistas em TI (tecnologia de informação) e em negócios conduzem à melhor tomada de decisões, principalmente com base em plataformas com modelos flexíveis que estruturam as ideias inovadoras. Alguns exemplos dessa nova realidade empreendedora são já conhecidos, tais como Airbnb e Uber. Assim, a transformação digital é baseada nos pilares TI, software e plataforma de produtos e/ou serviços. Além disto, a possibilidade de erro em digitalizações é extremamente reduzida a partir do reconhecimento de imagens com base em inteligência artificial, ficando em torno de 3% (ANDRADE, 2020).

O gerenciamento eletrônico de documentos (GED) é um recurso que possibilita organizar o acervo, evitando-se acúmulo de papel, e incorporando as tecnologias atuais, incluindo a viabilidade de manutenção e backup para evitar perdas. Os softwares de gerenciamento devem ser flexíveis, customizados e funcionais, além de compatíveis e possibilitar a integração com outros sistemas. O treinamento de pessoas se torna primordial (SEBRAE, 2019). Para a implementação de uma gestão do processo, o modelo registrado na Figura 1 propõe também a necessidade de software para controle das informações e documentos digitalizados (SETTI, 2008).

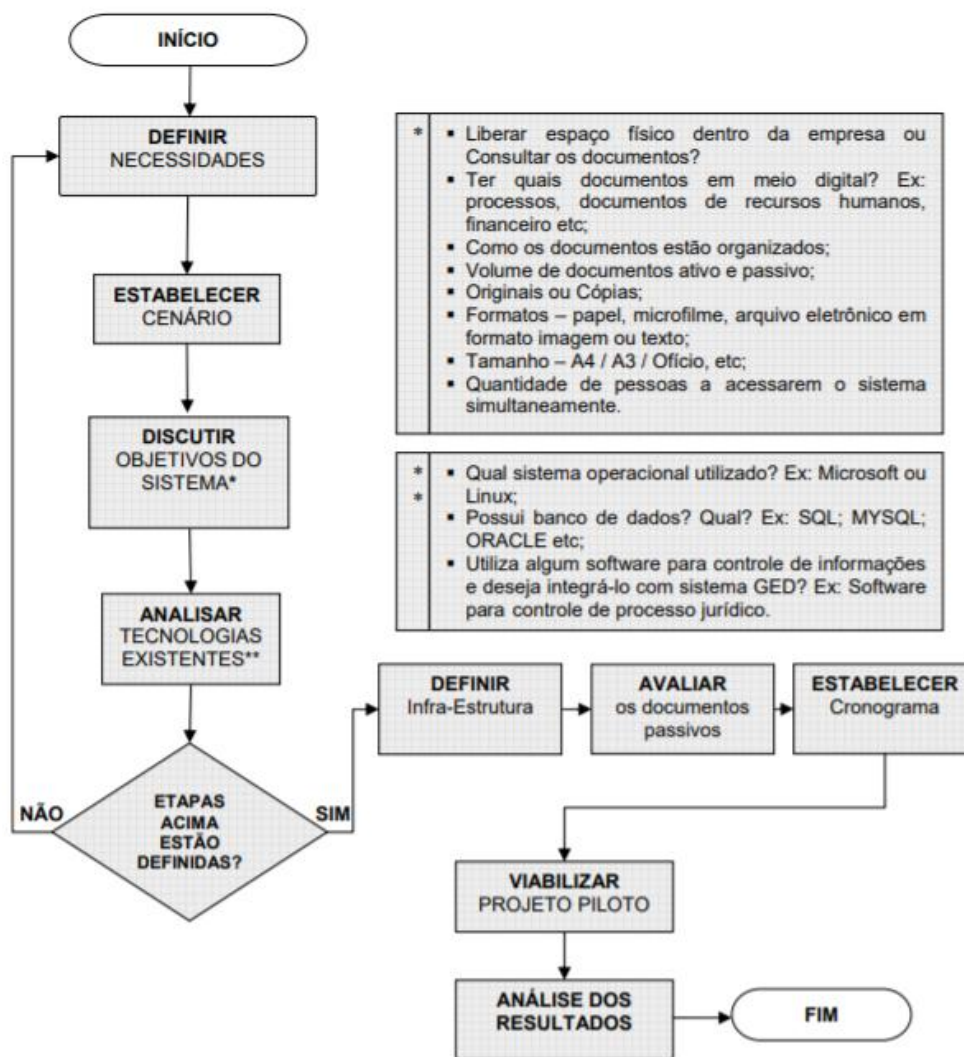


Figura 1 – Modelo para implantação de GED – gerenciamento eletrônico de documentos em ambientes corporativos. Fonte: (SETTI, 2008).

Segundo a CEDEFOP (2019), a mudança tecnológica indica obsolescência de competências, indicando necessidade de especialização. As novas demandas serão as competências técnicas, envolvendo conhecimentos em programação de informática, marketing, otimização de sistemas de informação e pesquisa; as competências linguísticas; as competências organizativas, tais como gestão de projetos, tempo e demais recursos; as competências comunicacionais, devido à interação com a diversidade; a obtenção de licenças comerciais; e a obtenção do trabalho em plataformas, com reformulação de preços. Serão

necessários alguns atributos pessoais nesta mudança, tais como flexibilidade, empatia, criatividade, disciplina, tolerância ao risco, pontualidade, independência e confiança.

Já existem no mercado ferramentas de processamento de PDF, tais como os que permitem a conversão de documentos em formatos PDF e PDF/A pesquisáveis. Na importação destes arquivos com extensão PDF, a camada de texto ausente pode ser inserida enquanto as propriedades são preservadas. Os dados XML podem ser extraídos de arquivos PDF/A-3, bem como inseridos ao salvar neste formato. Além disto, a inteligência artificial combinada com tecnologia ADRT ou tecnologia adaptável de reconhecimento de documentos da ABBYY permitem eliminação de estrutura lógica de um documento e identificar seus elementos de formatação, com aumento de precisão de reconhecimento. O FineReader Engine® analisará o layout do documento e avaliará individualmente as possibilidades (ABBYY, 2020). A forma que isto acontece está disponível na Figura 2.

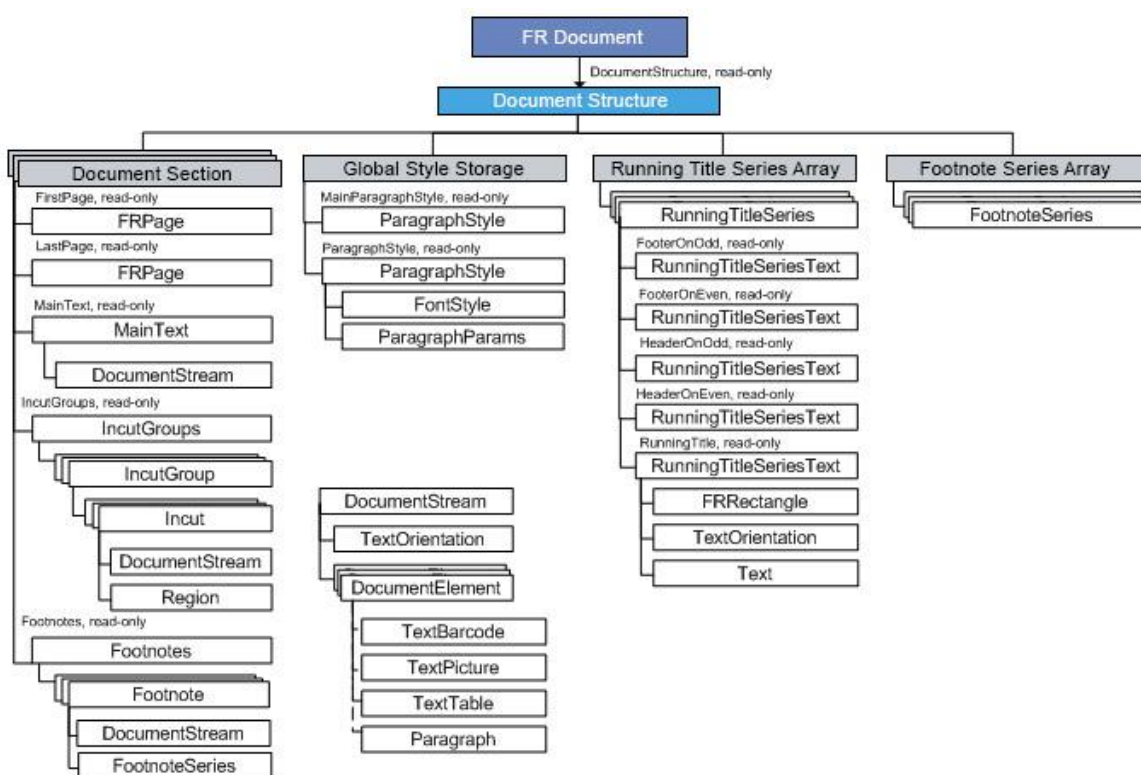


Figura 2. Estrutura do FineReader Engine®. Fonte: ABBYY (2020).

O reconhecimento ótico de caracteres (OCR) permite também transformar um documento PDF digitalizado em texto e imagens editáveis, com suporte a muitos idiomas. A conversão pode ser realizada em vários formatos, tais como DOCX, XLSX, PDF/A, EPUB, HTML, RTF, TXT, JPG, etc.; e também criar PDF a partir de qualquer tipo de documento. Porém, não é gratuito, sendo cobrado um valor anual (ADOBE, 2020).

O que se averigua na intenção de comercialização de produtos, é o oferecimento de customização e configuração multiníveis, suportes diversos, arquitetura WEB e integração XML e WEB-serviços, gestão de projeto, integração de software e materiais, gerenciamento da implantação, dentre outros. O reconhecimento ótico de caracteres (OCR) e reconhecimento inteligente de caracteres (ICR), além do reconhecimento de códigos de barras (OBR) e marcas de verificação (OMR) podem ser utilizados neste processo (CEDEFOP, 2019).

A classificação destas tecnologias mencionadas estão descritas na Tabela 1.

Tabela 1 – Classificação das tecnologias para conversão de diferentes tipos de documentos

Tecnologia	Classificação
OCR (Reconhecimento Ótico de Caracteres)	Reconhecimento e conversão de textos ou imagens em dados editáveis e pesquisáveis
ICR (reconhecimento inteligente de caracteres)	Reconhece caracteres impressos à mão em campos e enquadramentos, em diferentes idiomas
OBR (reconhecimento de código de barras)	Reconhece código de barras automaticamente em documentos
OMR (reconhecimento das marcas de verificação)	Reconhece marcas de seleção simples e agrupadas

Fonte: Adaptado de ABBYY (2020).

Para estas tecnologias, se a precisão é de 99%, apenas um caractere permanece incerto, e a tomada de ação final necessita de um usuário. Documentos de alta qualidade possuem maior precisão. Textos com scripts em árabe, malaio ou chinês, por exemplo, reduzem a precisão, ainda que seja alta quando comparada com outras opções tecnológicas existentes (ABBYY, 2020).

A resolução do escaneamento, em DPI ou pontos por polegada, cuja resolução mínima é de 300 dpi para alta precisão; e a qualidade do documento original, passível de reconhecimento pelo olho humano para que seja também para o reconhecimento, são itens necessários que influenciam um bom reconhecimento pelo software. Não há reconhecimento de texto com letra cursiva ou escrito à mão. Assim, o OCR reconhece específicos tipos de caractere após treiná-lo a partir de textos impressos na fonte, nos níveis de caractere e de palavra (LE ROY, 2017).

No processo de digitalização, as etapas de preparação (retirada de clipes e grampos, por exemplo); digitalização (disponibilizar os documentos para que sejam escaneados); indexação (possibilitar que o acesso ao documento seja preciso e rápido); controle de qualidade (revisar a digitalização); e disponibilização (armazenar o trabalho feito); também são partes integrantes do processo de gestão. Esta gestão permite que documentos sejam disponibilizados em um sistema inteligente que permita o acesso e busca dos mesmos por parâmetros, tais como tipo e nome do documento ou mesmo pelo seu conteúdo. As vantagens desta gestão de documentos são a permissão da diminuição do tempo na procura e do custo com espaço físico para o armazenamento, além de alta segurança (CAPELA, 2008).

Conforme o SEBRAE (2019), os consumidores estão cada vez mais procurando produtos e serviços que viabilizem sustentabilidade. Para se produzir 1 tonelada de papel são necessárias de 2 a 3 toneladas de madeira, além de uma grande quantidade de água e energia. O branqueamento da celulose utiliza químicos que colocam em risco a saúde humana, e podem contaminar solos e mananciais de água. O Brasil recicla 37% do papel que produz, e destes, 2% é destinado à produção de papel para impressão, 80% para a elaboração de embalagens, e 18% para papéis sanitários. Assim, o serviço deve ser diferenciado, averiguando-se desafios e soluções dentro de curto prazo.

Método

O trabalho se refere à uma adequação da implementação de digitalização de documentos em interface com inteligência artificial, de forma a adaptar um modelo específico, que contemple gestão, a partir de uma revisão de literatura. Há poucos artigos sobre o assunto on-line.

A revisão de literatura foi desenvolvida a partir de artigos científicos com o conteúdo proposto (GIL, 2008). A partir desta pesquisa, houve uma elucidação sobre o tema, mas de forma sistematizada, com fundamentos e procedimentos preconizados pela metodologia científica (GARCIA, 2016).

Foi proposto um objetivo e, assim, a busca em artigos on-line e livros com informações sobre o assunto proporcionaram a elaboração deste trabalho, e, cujas análises elucidaram sobre o assunto, e ainda foram propostas alternativas viáveis de implementação destas na área em questão (GARCIA, 2016).

Resultados e discussão

Verificaram-se muitas empresas especializadas em fornecer recursos, tais como softwares para captura, tratamento e gestão de conteúdo de imagens de grande volumetria, com plataformas modulares e seguras, organizadas por gestor de fluxos, com a inteligência artificial em prol de automatizar e gerenciar esse processo desde a captura do documento até o seu arquivamento eletrônico. Foram observados também serviços de reconhecimento e classificação automática, bem como acompanhamento das atividades e relatórios gerenciais.

Percebe-se a necessidade do estudo de definição de softwares e materiais e, posteriormente, configuração e personalização, além de implementação em si e teste. A formação de usuários e início das atividades também devem ser viabilizadas, além de assistência técnica e suporte permanentes.

A digitalização de documentos pretende minimizar o tempo de intervenção, assegurando a reprodução padronizada das operações realizadas (DOCSYSNET, 2020). Assim, as empresas devem agregar a inteligência artificial com confiança, a partir de uma plataforma de hardware ou servidores de nuvem inteligente pública, que também necessitam de placas e software compatíveis, em prol de eficiência, pois, a partir da IA, a automatização dos processos permitirá a organização e classificação de informações, reduzindo o tempo de processamento, evitando-se erros humanos e atrasos, além de possibilitar mais lucro.

Os documentos e formulários serão categorizados em tempo real durante o processo de digitalização, com indexação automática de itens, indicando uma gestão de documentos. São os processos de inteligência artificial que identificam, categorizam e interpretam os dados tratados, tornando possível aos softwares de gestão amenizar complexidades e padrões críticos.

Percebe-se que a definição das necessidades e o estabelecimento do cenário são, a priori, a tecnologia e mão-de-obra necessária para a execução da digitalização, que são realizados posteriormente à discussão dos objetivos do negócio, que se refere à maior eficiência de gestão do processo. A comparação das tecnologias existentes e dos resultados frente à implementação, viabilizarão a adequação do projeto e a análise dos resultados, para

posterior otimização do processo em uma empresa de qualquer segmento, que pretenda automatizar e melhorar a eficiência de atividades inerentes ao negócio.

Desta forma, independente do segmento, a gestão englobará a separação dos documentos a serem escaneados, as tecnologias e hardwares necessários, a forma de armazenamento e organização, e a mão-de-obra para possíveis correções e viabilização desta operação. Testes de eficácia serão úteis para mensurar a eficiência (tempo *versus* recursos) e o número de funcionários exigidos para tal atividade.

Por meio das câmeras dos gadgets é possível realizar escaneamentos, sendo que a maioria dos sistemas de gestão de documentos oferecem um aplicativo mobile com essa função, além de recursos para escaneamento e recuperação de imagens danificadas, sendo a novidade a inserção da ferramenta de inteligência artificial neste processo (ADOBE, 2020). Também é possível, além desta digitalização de documentos, a extração e coleta de dados contidos no arquivo gerado, bem como a classificação e a validação desses dados, de maneira automatizada e em tempo real.

Muito além da tecnologia de reconhecimento óptico de caracteres (OCR), a ferramenta de inteligência artificial capacita os sistemas de aprendizagem baseando-se em conteúdo e estrutura dos documentos, e as ferramentas atuais são capazes de identificar estes dados e corrigi-los, independente de formatação.

A lógica de programação destas ferramentas interpreta e extrai informações de maneira precisa e confiável. A inteligência artificial também atua no fluxo de comunicação via e-mail, chats, mídias sociais, voz, vídeos gerenciando processos de trabalho e conectando as etapas (DOCSYSNET, 2020).

No trabalho de Pohlmann et al. (1998), foi realizada a digitalização de documentos envolvendo o reconhecimento óptico dos caracteres e sua posterior conversão para HTML e PDF (Portable Document Format). Os passos necessários para tal trabalho se viabilizou de digitalização de páginas de livro de História da PUCRS, e conversão em texto mediante reconhecimento óptico de caracteres (OCR), além da criação de imagens no formato JPG, devido à alta taxa de compressão e qualidade. Os recursos utilizados foram um scanner HP Scanjet II, software de OCR (Reconhecimento Óptico de Caracteres) OmniPage Pro 5.0, editor de textos MS-Word 6.0 e o conjunto de softwares do Netscape Communicator 4.0. O processo está registrado na Figura 3.

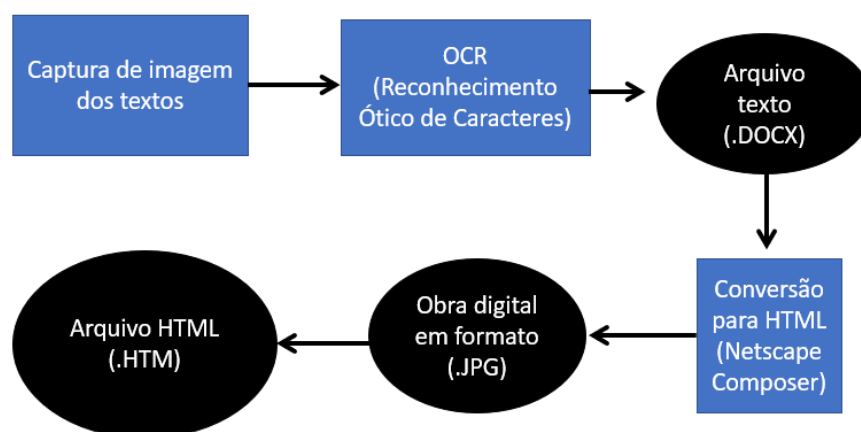


Figura 3. Reconhecimento óptico dos caracteres e sua posterior conversão para HTML.
 Fonte: POHLMANN et al. (1998).

Raabe; Pohlmann Filho (1998) ressaltam, em suas análises comparativas, que a conversão para texto possibilita edição e manipulação dos textos digitalizados, pesquisas full-text, mas que o processo para conversão em texto pode ser lento e trabalhoso. Já a conversão para imagem não possibilita edição e nem pesquisas full-text, porém, o processo de digitalização é rápido, mas pode ocupar um espaço para armazenagem considerável, em torno de 20 vezes daquele ocupado por textos. A velocidade de transmissão de dados deve atingir a expectativa almejada de transferência de arquivos em imagem.

Blanke; Bryant; Hedges (2012) reitera alguns pontos, tais como a necessidade de reformatação, disponibilização do texto (full-text) e codificação no formato HTML, além dos recursos disponíveis, que retratam tanto o tamanho dos arquivos, a análise de armazenagem e transmissão.

Se o objetivo é a conversão de documentos para formato digital e publicação em HTML ou em PDF, será, portanto, obrigatória a leitura ótica dos documentos, reconhecimento do texto (OCR), correção e/ou revisão do texto em editor de texto (Microsoft Word®), e finalização da conversão para formato HTML ou PDF.

Com o software OCR Omni Page Pro 8.0®, houve um tempo de conversão alto, de cerca de 510 minutos ou 8,5 horas para um número de 50 páginas e 12 figuras, devido à necessidade de revisão de erros gerados a partir do processo de reconhecimento óptico de caracteres (OCR). Para a captura das imagens e reconhecimento de caracteres, incluindo a criação do arquivo texto, com duas pessoas realizando o trabalho, o tempo médio foi de 65 minutos; para a revisão e correção do texto, 400 minutos; a conversão dos arquivos texto para HTML, 15 minutos; a captura de imagens e criação de arquivos BMP, 20 minutos; a inserção de imagens no texto e finalização, mais 10 minutos. O scanner utilizado possuía o recurso ADF (Automatic Document Feeder), para melhorar a performance (POHLMANN et al., 1998).

Há possibilidade de definir zonas de texto manualmente, com fidelidade ao layout original, porém, como necessita da interação do usuário, o processo se torna muito lento. Se o processo for automatizado, perde-se a qualidade quanto à intenção.

O total de palavras escaneado foi de cerca de 4785, indicando uma taxa de reconhecimento de 96,5%, com 168 palavras erradas. A taxa de reconhecimento eficiente deve ser superior a 98%, caso contrário, recomenda-se a redigitalização. Esta taxa é mensurada a partir do número de edições necessárias, que competem a inserções, substituições e outros, e o número total de caracteres, considerando o trabalho de localização dos erros no texto. Quando há muita linguagem técnica, torna-se necessária a averiguação manual, que dispenderá tempo para tal conferência (CAERE, 2019).

Assim, a taxa de eficiência de softwares de reconhecimento deve ser de 100% para não onerar o processo em tempo e pessoas para revisão. Outro ponto a ser considerado é a necessidade de um dicionário que abranja os termos técnicos da área de implementação deste recurso de conversão.

Raabe; Pohlmann Filho (1998) indicam que o processo se torna mais eficiente quando realizado em larga escala e com mínima interação do usuário.

Segundo a CAERE Corporation (2019), o software Omni Page Pro 8.0® atinge um reconhecimento superior a 99% em fontes padrão (Arial), quando há alta qualidade de impressão dos documentos.

O Acrobat PR DC® permite criação e publicação de documentos on-line, mantendo o layout original, em formato de arquivo PDF (Portable Document Format) e permite conversão de documentos com uma necessidade menor de interação (ADOBE, 2020).

Raabe; Pohlmann Filho (1998) digitalizaram a mesma obra da PUCRS em scanner HP Network Scanner 5 com bandeja para entrada automática de papel ADF (Automatic Document Feeder), que permitia apenas digitalização em tons de cinza, e gerava uma página em PDF para cada imagem escaneada. O texto foi convertido pelo software Adobe Acrobat Exchange®, mesmo que o mesmo não possuísse dicionário adequado para a língua portuguesa, inviabilizando o reconhecimento, porém, a leitura do texto foi possível. O tamanho dos arquivos se encontrou reduzido em até quatro vezes do original quando aplicado o OCR, mas as páginas foram mantidas como imagens. Para a realização de pesquisas full-text foi necessária a conversão de documentos para o formato PDF, evitando-se o passo de reconhecimento ótico dos caracteres (RAABE; POHLMANN FILHO, 1998).

Pohlmann; Raabe (1999) converteram arquivos em Microsoft Word® (DOC) e do Microsoft Excel® (XLS) e os imprimiram pelo Adobe PDF Writer®, que realizou a conversão dos arquivos e os agruparam em um. A captura das imagens dos textos no scanner decorreu em tempo médio de 16 minutos, com separação dos arquivos em arquivos em aproximadamente 30 minutos; o tempo médio para digitalização em PDF foi médio de 137 minutos ou 2 horas e 17 minutos, considerado eficiente, cujo processo está descrito na Figura 4.

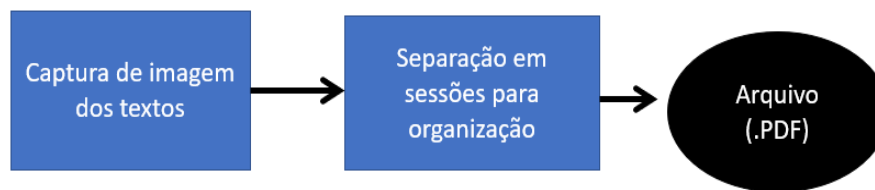


Figura 4. Digitalização e conversão para PDF. Fonte: POHLMANN et al. (1998).

A partir destes resultados, a sistemática em PDF viabilizou eficiência, em menor tempo e de forma confiável. Porém, deve-se considerar os recursos para tal execução, tais como hardware, software, instalações, operacionalização e treinamento. Porém, apresentou necessidade de maior espaço para armazenamento e tempo para transmissão em rede, além de não possibilitar pesquisas full-text e manipulação no caso de captura.

Raabe; Pohlmann Filho (1998) indicam a necessidade de um computador com um bom processador; um scanner com bandeja ADF; licenças de softwares utilizados, tais como Windows, Caere Omni Page, etc.; ar condicionado, mesas e cadeiras, luminárias, pontos de rede; e remuneração e encargos sociais. Para este item, recursos humanos, a alternativa em PDF pode viabilizar menor custo, devido à menor necessidade de interação.

Vários softwares são lançados periodicamente em prol da digitalização de documentos. A Adobe® lançou este ano um aplicativo gratuito que permite que a foto de qualquer documento seja convertida em PDF, permitindo inclusive pesquisa. Esta é uma tecnologia avançada de processamento de imagem, alimentada pela inteligência artificial Adobe Sensei®, em plataforma de aprendizagem de máquina da companhia em questão, com vantagem de se conectar a serviços já preexistentes. Este aplicativo possui recursos tais como captura, detecção de limites, correção de perspectiva, limpeza de imagem e recorte automático (ADOBE, 2020).

Pelo Microsoft Word®, é possível modificar um documento em PDF em smartphones. Seleciona-se o arquivo em PDF e, na sequência, o ícone da caneta no canto superior esquerdo da tela, logo abaixo do nome do arquivo. O aplicativo converterá o documento em editável. Pelo teste de eficiência realizado, observou-se que o processo levou 22 segundos, no total, com 100% de eficiência, sem necessidade de correções adicionais (Figura 5). O smartphone utilizado foi um com Android, S10 Samsung, com processador de até 12 GB de memória RAM.

Já para a digitalização de uma foto para PDF com o Adobe Scan®, pelo smartphone, é possível tirar a foto do arquivo, selecionar a área de texto e continuar, para a finalização e conversão do arquivo em PDF. O processo levou, ao todo, 16 segundos (Figura 6). O arquivo fica salvo em nuvem, gerenciada pela própria Adobe®.

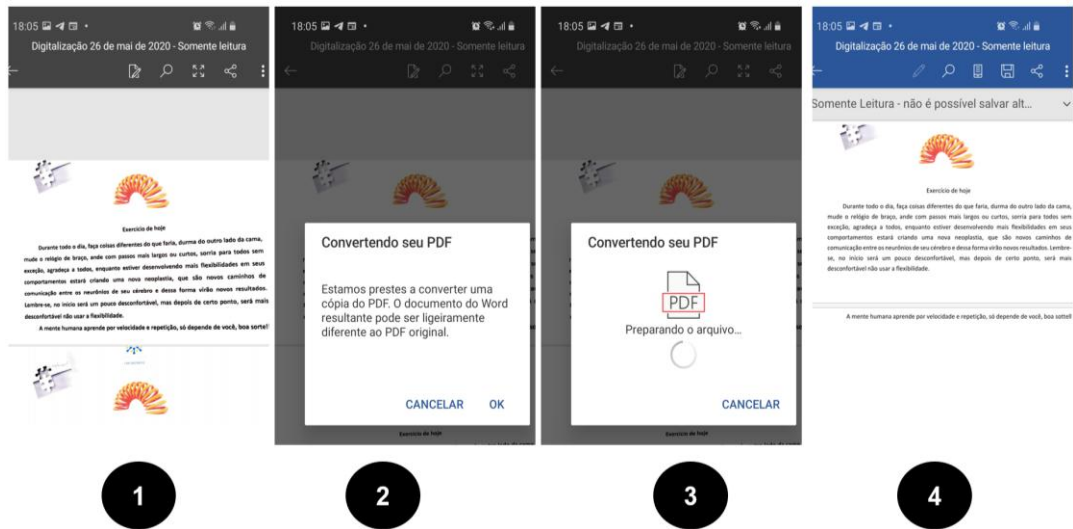


Figura 5 – Etapas de conversão de PDF para arquivo editável no Microsoft Word®, sendo 1: Abrir o arquivo; 2: Converter o PDF através do botão com a caneta do lado esquerdo superior da tela, loco abaixo do nome do arquivo; 3: aguardar o processo, que levou em torno de 10 segundos; 4: edição do arquivo no Microsoft Word®. Fonte: própria.

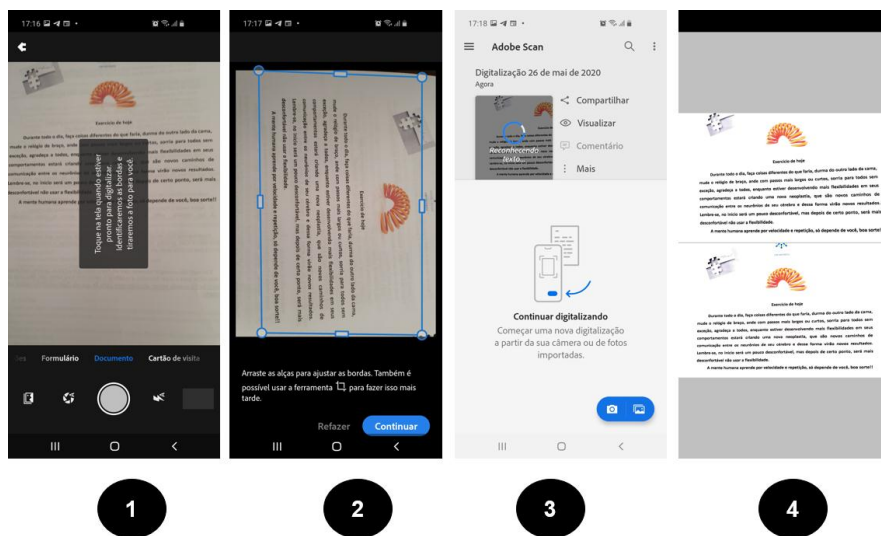


Figura 6 – Etapas de digitalização de documentos para PDF, sendo 1: Capturar a imagem; 2: Selecionar a área do documento; 3: Continuar e aguardar o processo, que levou em torno de 9 segundos; 4: Acessar o documento digitalizado, em formato PDF. Fonte: própria.

Assim, averiguou-se que é possível o processo de digitalização de documentos e conversão para DOCX, para edição do mesmo, e posterior armazenamento em PDF, que levará em torno de 38 segundos, viabilizando eficiência no processo, através destes softwares gratuitos e de fácil operação.

Verifica-se a importância para o desenvolvimento de ferramentas para o devido tratamento ótico nos arquivos PDF, para aprimorar a aparência do arquivo digitalizado final.

Se necessário, deve ser escolhido um software com interface para consultas on-line, com técnicas avançadas de pesquisa.

Desta forma, parametrizando-se a implementação de digitalização de documentos em qualquer área, deve-se averiguar as necessidades: scanner, computador, mesas e cadeiras, e demais recursos humanos e materiais; licenças de softwares; gestão em nuvem; além de gerenciamento destas atividades, o que otimizará os processos e os tornarão mais eficientes, com redução de custos e inovação que manterá a empresa no *roll* de negócios avançados e competitivos.

Considerações finais

Concluiu-se que o emprego de inteligência artificial na digitalização de documentos, independente do segmento ou negócio, torna-se fundamental, pois permite a implementação uma gestão eficiente, que viabilize redução de tempo e acessibilidade.

A inteligência artificial está atribulada aos softwares disponíveis para a digitalização de documentos, facilitando a implementação e a gestão. A aplicabilidade depende dos dados analisados, as implicações do sistema, os erros e tomadas de decisões, e demais desafios da implementação. Contudo, é um recurso disponível e necessário para implementar a eficiência do negócio e processos inerentes ao mesmo.

Foram encontrados softwares viáveis para captura, tratamento e gestão de conteúdo, além de serviços de reconhecimento, tais como o reconhecimento ótico de caracteres (OCR) e reconhecimento inteligente de caracteres (ICR), além do reconhecimento de códigos de barras (OBR) e marcas de verificação (OMR).

A digitalização de documentos minimiza o tempo de intervenção, assegurando a reprodução padronizada das operações, e, na gestão de documentos, os processos de inteligência artificial identificam, categorizam e interpretam os dados tratados, tornando possível aos softwares de gestão amenizar complexidades e padrões críticos. A inteligência artificial também atua no fluxo de comunicação.

A digitalização de documentos envolvendo o reconhecimento ótico dos caracteres e sua posterior conversão para PDF possui mais vantagens, pois viabiliza maior eficiência, com menor tempo de conversão. Porém, os demais recursos, tais como hardware, software, instalações, operacionalização e treinamento, devem ser considerados. O armazenamento e acesso à pesquisa rápida e confiável aos documentos digitalizados também devem ser considerados. Deve ser escolhido um software com interface para consultas on-line, com técnicas avançadas de pesquisa. Observa-se a importância de fácil manutenção e backup do acervo para evitar possíveis perdas.

A digitalização de documentos vem crescendo a cada dia no mercado, sendo que toda a comunidade ganha. Percebe-se rapidez para acesso aos documentos gerenciados, sem onerar as empresas em tempo e mão-de-obra, de maneira sustentável, visto que a produção de papel está entre as atividades industriais que mais causa danos ao meio ambiente.

Desta maneira, para a eficiência testada com softwares, averiguou-se que é possível o processo de digitalização de documentos e conversão para DOCX, para edição do mesmo, e

posterior armazenamento em PDF, que levou em torno de 38 segundos, viabilizando eficiência no processo, através destes softwares gratuitos e de fácil operação.

Referências bibliográficas

- ABBYY-SDK. **ABBYY**. 2020. Disponível em: <<https://www.abbyy.com/pt-br/ocr-sdk/key-features/adrt/>>. Acesso em 19 mai. 2020.
- ADOBE. **Acrobat PRO DC®**: Product Information. 2020. Disponível em: <<https://acrobat.adobe.com/br/pt/acrobat.html?promoid=C12Y324S&mv=other>>. Acesso em 19 mai. 2020.
- ANDRADE, M. D. Inteligência artificial para o rastreamento de ações com repercussão geral: o projeto Victor e a realização do princípio da razoável duração do processo. **Revista Eletrônica de Direito Processual – REDP**, v. 21, n. 1, p. 312-335, 2020.
- BLANKE, T.; BRYANT, M.; HEDGES, M. Open source optical character recognition for historical research. **Journal of Documentation**, v. 68, n. 5, p. 659-683. 2012.
- CAERE Corporation. **A Quantum Leap in Accuracy**. 2019. Disponível em: <<https://www.nuance.com/about-us/newsroom/press-releases/2019/Nuance-Earns-Highest-Rating-Opus-EIA-Report.html>>. Acesso em 19 mai. 2020.
- CAPELA, G. A. C. **Reconhecimento de símbolos musicais manuscritos na framework Gamera**. 98 f. Dissertação do MIEIC 2007/2008. Faculdade de Engenharia da Universidade do Porto. Mestrado Integrado em Engenharia, Informática e Computação. 2008. Disponível em: <<http://www.inescporto.pt/~jsc/students/2008ArturCapela/2008relatorioArturCapela.pdf>>. Acesso em 26 mai. 2020.
- CEDEFOP – Centro Europeu para o Desenvolvimento da Formação Profissional. **A digitalização e o futuro dos empregos e das competências: oportunidades e riscos**. NOTA INFORMATIVA. JUNHO 2019.
- DOCSYSNET. **Como funciona a inteligência artificial na gestão de documentos**. 2020. Disponível em <<http://www.docsysnet.com.br/veja-como-funciona-a-inteligencia-artificial-na-gestao-de-documentos/>>. Acesso em 10 mai. 2020.
- GARCIA, E. Pesquisa bibliográfica versus revisão bibliográfica: uma discussão necessária. **Revista Línguas & Letras**. v. 17, n. 35, 2016.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2008.
- KOENIGKAM, S.; FERREIRA JÚNIOR, J. R.; WADA, D. T.; TENÓRIO, A. P. M.; BARBOSA, M. H. N.; MARQUES, P. M. A. Inteligência artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica: avanços da imagem rumo à medicina de precisão. **Radiologia Brasileira**, v. 52, n. 1, p. 387–396, 2019.
- LE ROY, L. **O que é OCR (Reconhecimento Ótico de Caracteres) e como funciona**. SESIN – Soluções Integradas para documentos e processos. 2017. Disponível em: <<http://sesin.com.br/o-que-e-ocr-reconhecimento-otico-de-caracteres-e-como-funciona/>>. Acesso em 26 mai. 2020.

- NUNES, B. P. B. **Inteligência Artificial na GRH**: competências necessárias aos gestores para a criação de valor. 2019. 57 f. Dissertação de mestrado em Gestão de Recursos Humanos. Universidade de Lisboa (ISEG - Lisbon School of Economics & Management).
- POHLMANN, O. F.; CAMPOS, M. B.; RAABE, A. L.; JOHN, F.; VIERA, S. Em direção a criação de uma biblioteca digital na Pontifícia Universidade Católica do Rio Grande do Sul: uma experiência prática. In: **II Seminário Internacional de Bibliotecas associadas a UNESCO**, Cienfuegos – Cuba. 23 a 27 de maio de 1998.
- POHLMANN, O. F.; RAABE, A. L. Direito autoral no contexto de bibliotecas digitais. In: **III Congresso Internacional de Informática Educativa**, Santa Fé – Argentina. 14 a 17 de abril de 1999.
- RAABE, A.; POHLMANN FILHO, O. Estudo comparativo entre sistemáticas de digitalização de documentos: formatos HTML e PDF. **Ci. Inf.**, v. 27, n. 3, p. 300-310, 1998.
- RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. Campus: São Paulo, 2004. 1040 p.
- SEBRAE – Serviço de apoio às micro e pequenas empresas. **Ideias de Negócios**: digitalização de documentos. 2019. Disponível em:
<<https://respostas.sebrae.com.br/categorias/digitalizacao-de-documentos/>>. Acesso em: 26 mai. 2020.
- SETTI, R. J. **Estudo GED**: Gerenciamento Eletrônico de Documentos no âmbito organizacional. 62 f. Monografia. UNIVERSIDADE FEDERAL DO PARANÁ - DEPARTAMENTO DE CIÊNCIA E GESTÃO DA INFORMAÇÃO. 2008