

## PROPOSTA DE DESENVOLVIMENTO DE UM SISTEMA PARA EXTRACÇÃO DOS DADOS DO BILHETE DE IDENTIDADE ANGOLANO BASEADO EM INTELIGÊNCIA ARTIFICIAL

### PROPOSAL FOR THE DEVELOPMENT OF A SYSTEM FOR EXTRACTION OF DATA FROM THE ANGOLAN IDENTITY TICKET BASED ON ARTIFICIAL INTELLIGENCE

Delphin Kabey Mwinken\*  
Stefan João Samukuanha Rodrigues

E-mail: [delphinsrc@gmail.com](mailto:delphinsrc@gmail.com) e [stefanrodrigues2015@gmail.com](mailto:stefanrodrigues2015@gmail.com)

#### RESUMO

O presente trabalho faz uma análise da situação problemática do processo de inscrição no Instituto Superior Politécnico do Huambo (ISPH). Este processo tem de ser rápido e íntegro na captação e processamento dos dados dos candidatos, mas actualmente a execução das actividades deste processo não satisfazem integralmente os requisitos ora mencionados, o que faz com que sejam utilizados muitos recursos humanos e materiais. Neste trabalho foi desenvolvido um sistema informático para dar solvência as fraquezas do mesmo, por meio do uso das Tecnologias de Informação e Comunicação (TIC). A metodologia de desenvolvimento de *software eXtreme Programming* (XP) foi utilizada para guiar o desenvolvimento do sistema proposto. A linguagem de programação *Python* agregada ao *framework Django* possibilitou o desenvolvimento de um produto de *software* de acordo as histórias de usuário da metodologia XP. Para persistência dos dados se utilizou uma base de dados relacional. Com as técnicas e metodologias usadas no presente trabalho foi possível desenvolver um produto de *software* que carrega as imagens de frente e verso de um Bilhete de Identidade (BI), analisa e edita tais imagens e extrai os dados presentes no BI de uma forma rápida e económica. O sistema desenvolvido no presente trabalho vai auxiliar todos os envolvidos na realização e execução do processo de inscrição do ISPH, pois dinamiza e automatiza os várias actividades decorrentes nele.

**Palavras-chave:** Análise, inscrição, processo, persistência, *python*.

#### ABSTRACT

*This paper analyzes the problematic situation of the registration process at the Instituto Superior Politécnico do Huambo (ISPH). This process has to be fast and accurate in capturing and processing candidate data, but currently the execution of the activities of this process does not fully meet the requirements mentioned above, which means that a lot of human and material resources are used. In this work, a computer system was developed to solve its weaknesses through the use of Information and Communication Technologies (ICT). The eXtreme Programming (XP) software development methodology was used to guide the development of the proposed system. The Python programming language added to the Django framework made it possible to develop a software product according to XP methodology user stories. For data persistence a relational database was used. With the techniques and methodologies used in the present work it was possible to develop a software product that uploads the two-sided images of an Identity Card (IC), analyzes and edits such images and extracts the data present in BI in a fast and easy way economic. The system developed in the present work will assist all involved in the implementation and execution of the ISPH application process, as it streamlines and automates the various activities arising from it.*

**Keywords:** Analysis, inscription, process, persistence, *python*

## Introdução

Tecnologia da informação e comunicação , é a área que utiliza ferramentas tecnológicas com o objetivo de facilitar a comunicação e o alcance de um alvo comum. Além de beneficiar a produção industrial de um determinado bem, as TICs também servem para potenciar os processos de comunicação. As tecnologias de informação e comunicação (TIC), podem ser definidas como o conjunto de todas as actividades e soluções providas por recursos de computação, usando todos os meios técnicos para tratar a informação e auxiliar na comunicação. (REYNOLDS, 2006)

As TIC são utilizadas das mais diversas formas, na indústria (no processo de automação), no comércio (no gerenciamento e nas diversas formas de publicidade), no sector de investimentos (informação simultânea e comunicação imediata) e na educação (no processo de ensino-aprendizagem e na gestão universitária), representando um avanço na educação a distância. Na actualidade vivemos na chamada Sociedade da informação onde é fundamental estar informado e comunicar por telefone ou através da *internet*, as novas tecnologias (TIC) têm uma grande importância na vida das pessoas(coletiva e individual). Graças às novas tecnologias foi também possível criar novos empregos através do aparecimento da *internet*. As tecnologias têm um grande potencial inovador que se tem reflectido em todos os aspectos da nossa sociedade: mudaram as formas de aprender e ensinar, e também as formas de trabalhar, de comunicar e lazer (B.Escolas, 2016).

Um exemplo da importância das TICs na educação, é a possibilidade de uma melhor aprendizagem por parte dos alunos e um melhor ensino dos professores quando se recorre por exemplo ao uso do computador, *internet* ou quadros interactivos. As TIC são também uma fonte de rendimento de vários países e bem como nas demais organizações (PORTER, 1986).

A sociedade angolana actualmente também tem usufruído dessas tecnologias em áreas como na administração pública central e local, na estrutura empresarial da mesma e na educação. A difusão das TICs, em Angola, contribuiu para simplificar processos administrativos e proporcionar a redução dos custos que lhe estão associados. Agilizando também o relacionamento entre as instituições e os cidadãos (Evandroamaral, 2018).

Contudo o Instituto Superior Politécnico do Huambo (ISPHbo), visto ser uma instituição de ensino tecnológico, tem acompanhado as tendências modernas no que tange as TIC, usando-as de acordo a sua estrutura organizacional e funcional.

Tendo em vista a estrutura organizacional e funcional do ISPHbo, já foram desenvolvidos vários artefactos de *software* de gestão por estudantes finalistas, desde a data em que se deram início às actividades no mesmo, para as mais diversas áreas funcionais do Instituto, como: Biblioteca, Sistema de Gestão de Segurança de Informação para o ISPHbo , Secretaria Geral, dentre outros; com o intuito de

dinamizar e automatizar as actividades decorrentes nestes departamentos. Mas nem todos estes sistemas têm sido usados após serem concluídos, as vezes por não satisfazerem a real necessidade da Instituição, o que leva ao desinteresse da mesma em usar tais sistemas.

## **1.2 Situação Problemática**

Actualmente no processo de inscrição do ISPHbo não se faz o uso automático das inserções dos dados do bilhete de identidade dos aspirantes no computador.

O sistema Luvili apresenta também debilidades como a não leitura dos caracteres especiais, e nem com isso se eliminaram os erros humanos que se cometiam na realização das inscrições e a perca de tempo nas extrações dos dados do bilhete de identidade dos respectivos aspirantes, ainda se notam enchentes no acto das inscrições, isso devido as demoras consequentes do processamento manual de algumas actividades no processo de inscrição dos aspirantes. Pretende-se tornar com o presente trabalho mais ágil o processo de inscrição ao ISPHbo.

### **1.2.1 Metodologia de investigação**

Este trabalho se enquadra na classificação de um projecto de desenvolvimento tecnológico, por que o resultado esperado é um sistema informático baseado em inteligência artificial, que ajudará nas abstrações dos dados do BI durante o processo de inscrições no ISPHbo.

### **1.2.2 Métodos teóricos**

Análise-síntese: Estão presentes ao longo de todo o processo investigativo. Por exemplo, a informação dos documentos selecionados durante a revisão sistemática da literatura primeiro se submeterá a uma análise a partir das perguntas da investigação e os critérios de qualidade assumidos, extraem-se as respostas às perguntas, e logo se realiza a síntese da informação mediante diversas formas (texto, tabelas e gráficos).

Indução-dedução: Estão presentes ao longo de todo o processo inquiridor. Por exemplo, realizou-se o diagnóstico inicial a partir de casos particulares e logo se generalizou para formular o problema (procedimento indutivo). Mediante um procedimento dedutivo, por exemplo, tomam-se como premissas principais as experiências positivas que se reportam na literatura sobre o desenvolvimento de softwares de inteligência artificial (IA), para inferir que se este trabalho busca também desenvolver um software de IA para o caso específico do processo de inscrição do ISPHbo, então se chega à conclusão de que é possível aplicar algumas dessas experiências positivas em caso específico.

Histórico-lógico: Analisaram-se antecedentes, boas práticas e lições aprendidas sobre o desenvolvimento de software de IA, com ênfase naquelas que se relacionam com o tipo de software que se seleccionou para a disciplina.

Modelação: No processo de desenho do sistema de IA se construirão modelos tendo em conta as características das metodologias ágeis que se utilizem.

## **Desenvolvimento**

### **2. Processamento de imagens**

A área da computação que estuda a melhoria de informações visuais para a interpretação humana, processamento de dados de cenas para percepção automática através de máquina. Processar uma imagem significa modificar as informações contidas nessa imagem sob vários aspectos, de modo que o resultado seja uma imagem ou informações extraídas dela. Processamento de imagens é a ciência que manipula imagens. É composta por uma grande gama de métodos, que se encontram presentes nas mais diversas aplicações. (Santos, Saito, & Peluzio, 2010)

Estas técnicas de processamento podem realçar ou distorcer uma imagem, destacar certas características importantes e restaurar imagens degradadas pela ação do tempo ou pela forma de aquisição. Armazenamento, transmissão, representação, extração de informações, visão computacional. Há várias finalidades para o processamento de imagens, como por exemplo converter sinais a partir de um sensor de imagem para imagens digitais. Profissionais de diversas áreas – fotógrafos por exemplo – dependem de softwares que executem processamento de imagens para melhorar a exposição, remover ruídos, otimizar imagens e prepará-las para impressão. (Santos D. , 2010)

#### **2.1. Antecedentes do trabalho**

O Sistema Luvili actualmente é utilizado para o processo de inscrição dos candidatos ao ISPHbo, no entanto não permite a inserção dos dados dos candidatos de forma automática, o que tem levado às vezes a inserção de informações dos candidatos não condizentes com a realidade dos documentos por eles apresentados. Devido a facto emanado anteriormente, tem surgido várias reclamações no que tange aos dados dos candidatos durante a fase de inscrição o que tem levado a muitos alguns constrangimentos. Portanto neste trabalho não se desenvolverá um sistema de inscrição completo, mas sim se focará no aspecto da abstração dos dados dos candidatos do BI como descrito no objecto de estudo na introdução deste trabalho, para dirimir às reclamações e aumentar e acelerar a inscrição de um candidato.

#### **2.2 Descrição do processo a ser automatizado**

O processo de inscrição no ISPHbo inicia à partir do momento em que os candidatos se dirigem ao mesmo onde lhes é solicitado os documentos (cópia do BI, certificado de habilitação literárias do ensino médio, comprovativo de pagamento da inscrição no banco, fotografia tipo passe e para os candidatos do sexo masculino talão de recenseamento militar) para averiguação da autenticidade e validade dos

mesmos. Feita a análise documental, é então permitida a entrada dos candidatos a sala de inscrições para o cadastramento do candidato no sistema de inscrições denominado Luvili. Assim sendo, no sistema Luvili é onde é associado os dados do candidato ao curso ou cursos em que o mesmo deseja inscrever-se para realização do teste de admissão. Feita a inscrição do candidato na base dados, é então impresso um recibo que confirma a inscrição do candidato ao teste de admissão.

### 2.3 Tecnologias utilizadas

Para realização deste projecto foi utilizado as seguintes tecnologia

#### a) Tesseract

Tesseract é um software de reconhecimento ótico de caracteres de código aberto, originalmente desenvolvido pela Hewlett-Packard e foi por um tempo mantido pelo Google, atualmente o projeto está hospedado no GitHub. Se aplica a imagens em formato [tiff](#) com texto puro em uma única coluna, convertendo a saída em um arquivo [txt](#). Não possui mecanismos para reconhecimento de layout, desta forma não é recomendável para textos que possuam fórmulas ou mais de uma coluna. (Github, 2016)

#### b) OpenCV

A visão computacional procura modelar e replicar a visão humana usando software e hardware. O processamento de imagens tem uma imagem como parâmetro de entrada e um conjunto de valores numéricos como saída (objetos complexos, vetores, matrizes).

OpenCV (*Open Source Computer Vision*) é uma biblioteca de programação, de código aberto e inicialmente desenvolvida pela Intel com o objetivo de tornar a visão computacional mais acessível a desenvolvedores e hobistas. Atualmente possui mais de 500 funções, pode ser utilizada em diversas linguagens de programação (*C++, Python, Ruby, Java*) e é usada para diversos tipos de análise em imagens e vídeos, como detecção, tracking e reconhecimento facial, edição de fotos e vídeos, detecção e análise de textos. (Rocha, 2018).

### 2.4 Tecnologias do lado cliente

Desenvolvimento ***web front-end*** é a prática de converter dados em uma interface gráfica, através do uso de HTML, CSS e JavaScript, para que os usuários possam visualizar e interagir com esses dados. (MrOllie, 2019). Existem várias ferramentas e plataformas (*wordpress, magento etc.*) disponíveis que podem ser usadas para desenvolver o ***front end*** de um sítio, e entender quais ferramentas são mais adequadas para tarefas específicas que auxiliará o desenvolvimento de um sítio bem projectado e escalável que são:

#### a) Hyper Text Markup Language (HTML)

A Hyper Text Markup Language (HTML) é a espinha dorsal de qualquer processo de desenvolvimento de sites, sem o qual não existe uma página da web. Hipertexto significa que o texto possui links, denominados *hiperlinks*, incorporados. Quando um usuário clica em uma palavra ou frase que possui

um *hiperlink*, ele trará outra página da web. Uma linguagem de marcação indica que o texto pode ser transformado em imagens, tabelas, links e outras representações. É o código HTML que fornece uma estrutura geral da aparência do site. HTML foi desenvolvido por *Tim Berners-Lee* (MrOllie, 2019) .

**b) CSS (*Cascading Style Sheets*)**

O CSS (*Cascading Style Sheets*) controla o aspecto da apresentação do site e permite que ele tenha uma aparência única. O CSS é uma outra linguagem, separada do HTML, com objectivo único de cuidar da estilização da página. A vantagem é que o CSS é bem mais robusto que o HTML para estilização. Pois, escrever formatação visual misturado com conteúdo de texto no HTML se mostrou algo bem impraticável. O CSS resolve isso separando as coisas; regras de estilo não aparecem mais no HTML, apenas no CSS. (Book, 2016)

**c) JavaScript**

JavaScript é uma linguagem de programação imperativa baseada em eventos (em oposição ao modelo de linguagem declarativa do HTML ) usada para transformar uma página HTML estática em uma interface dinâmica. O código JavaScript pode usar o DOM (Document Object Model), fornecido pelo padrão HTML, para manipular uma página da Web em resposta a eventos, como a entrada do usuário. (Gonçalves, 2005)

### **3.Resultado esperado e Discussão**

A utilização de uma metodologia de desenvolvimento de *software* é de fundamental importância para que um produto de *software* cumpra de forma íntegra aquilo que se pretende que realize. Em consequência disso, neste trabalho utilizou a metodologia de desenvolvimento XP para que o sistema através dele desenvolvido cumprisse com às requisições a ele desejadas que dão solução as arduidades identificadas na execução do processo de abstração dos dados dos candidatos na altura de inscrição ao ISPH. Assim sendo, neste capítulo se demonstram as fases de desenvolvimento do Sistema, através da utilização dos artefactos (HU, TA e TE) da metodologia XP para o desenvolvimento ágil de *software*.

#### **3.1 Requisitos do sistema**

Requisitos são solicitações, desejos, necessidades. Um requisito é a propriedade que um *software* exibe para solucionar problemas reais, é a conjuntura indispensável para satisfazer um objeto. (Codificar, 2107)

Requisito Funcional (RF) refere-se à requisição de uma função que um *software* deverá atender/realizar, ou seja, exigência, solicitação, desejo, necessidade, que um software deverá materializar. (Ventura, 2018)

### **Conclusões**

O desenvolvimento do presente trabalho possibilitou demonstrar como um sistema informático de gestão pode aumentar a celeridade na execução de tarefas e redução de custos em recursos humanos dos diferentes processos que são realizados em instituições de ensino superior, especificamente no processo de inscrição do ISPHbo. Realmente os objectivos específicos, propostos na introdução foram alcançados, tendo em conta que, foi desenvolvida a base de dados para persistência dos dados dos candidatos. Também as ferramentas selecionadas permitiram desenvolver o algoritmo de edição e abstração dos dados dados de um candidato. A metodologia de desenvolvimento de software *XP*, possibilitou descrever o processo objecto de estudo, bem como permitir que o sistema esteja bem documentado. Quanto ao objectivo geral proposto no presente trabalho, tal foi alcançado porque, os testes de aceitação realizados da metodologia *XP* comprovam que o sistema desenvolvido é funcional.

## Referências Bibliográficas

1. Ailin Orjuela Duarte, MSc., & Mauricio Rojas C., MSc. (2008). *Las Metodologías de Desarrollo Ágil como una Oportunidad para la Ingeniería del Software Educativo*. Colombia: Universidad de Pamplona.
2. B.Escolas. (18 de março de 2016). *As Tecnologias de Informação e Comunicação*. rio de janeiro.
3. Book, F. P. (2016). *Note For Professionals*. Css Group Company.
4. Cethos. (1 de 11 de 2019). *Introdução às Linguagens de Programação/Python - Wikiversidade*. Obtido de [https://pt.m.wikiversity.org/wiki/P%C3%A1gina\\_principal](https://pt.m.wikiversity.org/wiki/P%C3%A1gina_principal)
5. Codificar. (14 de Agosto de 2107). *Requisitos funcionais e requisitos não funcionais, o que são?* Obtido em 02 de Julho de 2018, de <https://codificar.com.br/aplicativos/requisitos-funcionais-nao-funcionais/>
6. Crockford, D. (3 de Setembro de 2001). *JavaScript*. Obtido em 10 de 09 de 2019, de The world's most misunderstood programming language: <http://javascript.crockford.com/javascript.html>
7. Evandroamaral. (8 de julho de 2018). *sapoblogs*. Obtido em 11 de 2019 de 2019, de <http://www.evandroamaral.blogs.sapo.ao/a-utilizacao-das-tecnologias-de-10666>
8. Github. (04 de fevereiro de 2016). *ReadMe*. Obtido de Site da Github: <http://github.com/tesseract-ocr/tesseract/blob/master/ReadMe.md>
9. Gonçalves, L. R. (2005). *Curso Básico de JavaScript*.
10. Javascript, w. s. (16 de 01 de 2007). *w3 schools*. Obtido em 04 de 10 de 2019, de w3 schools: <http://www.w3schools.com/js/default.asp>

11. Jeffries, R. (16 de Março de 2011). *What is Extreme Programming?* Obtido em 2 de Julho de 2018, de <http://xprogramming.com/what-is-extreme-programming/>
12. Longo, H. E., & Silva, M. P. (1 de Julho de 2014). A Utilização de Histórias de Usuários no Levantamento de Requisitos Ágeis. *A Utilização de Histórias de Usuários no Levantamento de Requisitos Ágeis.*
13. Makesys. (20 de Novembro de 2013). *Definindo seu software: o que são requisitos funcionais e não funcionais?* Obtido em 2 de Julho de 2018, de <http://blog.makesys.com.br/definindo-seu-software-o-que-s>

## Síntese Curricular dos Autores

Autor <sup>1</sup> **Delphin Kabey Mwinken\***

Mestre em Engenharia Civil , Licenciado em Ciências Exactas Professor de Álgebra e Geometria Analitica,Matemática IV,Analise Numerica e Electronica Teorica , Chefe de Departamento de Arquitectura e Engenharia Civil ,Engenharia Hidráulica e Mecânica no Instituto Superior Politécnico do Huambo da Universidade José Eduardo dos Santos, [delphinsrc@gmail.com](mailto:delphinsrc@gmail.com) telemóvel:924935008, Huambo-Angola

Autor <sup>2</sup> **Stefan João Samukuanha Rodrigues**

Engenheiro **Stefan João Samukuanha Rodrigues** formado em Engenharia Informática e Computadores no Instituto Superior Politécnico do Huambo da Universidade José Eduardo dos Santos **Contactos** [stefanrodrigues2015@gmail.com](mailto:stefanrodrigues2015@gmail.com) d Telemóvel:931767800 Huambo-Angola