



UNIVERSIDADE AGOSTINHO NETO
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE CIÊNCIAS DA COMPUTAÇÃO

BERNARDO ISSENGUEL
ESTIMA EPALANGA
HILQUIAS CHITAZO
MAER POLENHAS NETO

**SISTEMA DE ARQUIVAMENTO PARA A WEB ANGOLANA
(AWA)**

Luanda
2020

BERNARDO ISSENGUEL
ESTIMA EPALANGA
HILQUIAS CHITAZO
MAER POLENHAS NETO

**SISTEMA DE ARQUIVAMENTO PARA A WEB ANGOLANA
(AWA)**

Trabalho de Fim de Curso de Ciências da Computação
apresentado à Faculdade de Ciências da Universidade
Agostinho Neto, como requisito parcial para obtenção
do título de licenciado em Ciências da Computação.

Orientador: Prof. Mateus Padoca Calado, Ph.D.

Luanda
2020

Departamento de Ciências da Computação

Faculdade de Ciências da Universidade Agostinho Neto
Campus da Universidade Agostinho Neto | Rua Direita da Camama e Rua do Estádio 11 de Novembro
Município de Belas, Luanda - Sul
Luanda | Angola
e-mail: dcc@fc.uan.ao

(Esta página foi intencionalmente deixada em branco)

DEDICATÓRIA

À minha família por todo apoio, em especial a minha querida mãe por ter mostrado confiança no meu percurso desde o inicio.

“Bernardo Issenguel”

À todos os docentes que participaram no processo da minha formação, e a todos que de alguma maneira foram pilares fortes durante esse meu percurso académico.

“Hilquias Gerson da Conceição Chitazo”

À minha família, em especial aos meus pais pelo esforço despendido para que hoje a minha formação fosse concluída com êxito.

“Estima Epalanga C. Sahuma”

Aos meus pais, pela força, pela capacidade de acreditar e investir na minha formação; à minha família e amigos, pelo suporte dado para a elaboração deste trabalho.

“Maer Luckeny Polenhas Neto”

AGRADECIMENTOS

Agradecemos primeiramente à Jeová Deus, pela vida, pela força e por nos proporcionar saúde durante a realização deste trabalho e não só.

À nossa família, pelo apoio incondicional dado durante todo o processo de formação.

Ao Professor Doutor Mateus Padoca Calado pelo apoio prestado na elaboração deste trabalho como orientador.

Os nossos agradecimentos aos docentes do Departamento de Ciências da Computação da Faculdade de Ciências da Universidade Agostinho Neto, Dikiefu Fabiano, Vicente Lopes João, Lufialuiso Sampaio Velho, Darlines Munõz, Dongala Soares, Hélio Santana, Kissemma Eduardo Rafael, Bongo Cahisso e Amândio Almada.

Agradecimentos especiais ao colega Paulino Jonas pelos subsídios emprestados durante a elaboração do projecto nas diversas temáticas, servindo como coorientador.

Igualmente agradecemos à toda Direcção da Faculdade de Ciências da Universidade Agostinho Neto, por nos proporcionar ensino de qualidade e condições necessária e suficientes, que já é marca da unidade orgânica, durante esses anos todos da nossa formação acadêmica no ensino superior.

A elaboração deste trabalho não teria sido possível sem a colaboração, estímulo e empenho de diversas pessoas. Gostaríamos, por este facto, de expressar toda a nossa gratidão e apreço a todos aqueles que, directa ou indirectamente, contribuíram para que esta tarefa se tornasse uma realidade. A todos queremos manifestar os nossos sinceros agradecimentos.

Em primeiro lugar a Deus o todo Poderoso Jeová, aos meus pais, todos os meus docentes e colegas que directa ou indirectamente contribuíram para a minha formação e por último, mas não menos importante a minha família e amigos.

“Maer L. Polenhas Neto”

Primeiramente, rendo graças à Deus pelo seu imenso amor e continua graça e misericórdia em minha vida, aos meus pais especialmente por tudo que representam para mim, aos docentes e colegas do Departamento de Ciências da Computação.

“Bernardo Issenguel”

A Ele que guia todos os meus passos, que ilumina o meu caminho, que garante a minha força e motivação diária para cada dia de trabalho árduo, aqui fica o meu muito obrigado. A minha família e amigos, as suas motivações e incentivos foram os impulsos que ajudaram para a conclusão deste percurso de minha vida.

“Estima Espalaga”

Pela vida e pelos méritos alcançados e outros ainda almejados, pela força que Ele me dá a cada dia, para que eu possa levantar e batalhar arduamente pelos meus sonhos, agradeço a Deus Pai todo poderoso primeiramente, a minha família, amigos, em especial ao meu pai, a minha mãe pelo apoio, incentivo e coragem e a todos que directamente ou indirectamente contribuíram para a minha formação os meus sinceros agradecimentos.

“Hilquias Chitazo”

RESUMO

A Web é a maior fonte de informação alguma vez construída. A tendência verificada nos últimos anos indica que a popularidade da Web vai continuar a aumentar no futuro, assim como a quantidade de informação que nela é exclusivamente publicada. No entanto, a informação publicada na Web está disponível durante um período de tempo muito curto, findo o qual, por regra se perde para sempre. Surge assim o interesse na criação de arquivos da Web que permitam preservar esta informação para gerações vindouras.

O arquivamento na Web é o processo de coletar partes da World Wide Web para garantir que as informações sejam preservadas em um arquivo para futuros pesquisadores, historiadores e o público. Os arquivistas da Web geralmente utilizam rastreadores da Web para captura automatizada devido ao enorme tamanho e quantidade de informações na Web. A maior organização de arquivamento da Web baseada em uma abordagem de rastreamento em massa é o Internet Archive, que se esforça para manter um arquivo da Web inteira.

Palavras-chave: Extração, filtração, ordenação por assunto, pesquisa.

ABSTRACT

The Web is the largest source of information ever built. The trend in recent years indicates that the popularity of the Web will continue to grow in the future, as well as the amount of information solely published on it. However, the information published on the Web is available for a very short period of time, after which, as a rule, is lost forever. This motivates the creation of web archives that allow the preservation of this information for future generations.

Web archiving is the process of collecting information from the World Wide Web to ensure that information is preserved in the future for researchers, historians, and the public. Web archivists, for the most part, use Web crawlers for automated data capture and huge amounts of information on the Web. The largest Web archive organization based on a mass-tracking approach is the Internet Archive, which strives to maintain an entire web archive.

Keywords: Extraction, filtration, ordering by subject, research.

LISTA DE ABREVIATURAS E SIGLAS

XP	<i>eXtreme Programming</i>
RF	<i>Requisito Funcional</i>
RNF	<i>Requisito Não Funcional</i>
UML	<i>Unified Modeling Language</i>
HTML	<i>Hypertext Markup Language</i>
PHP	<i>Hypertext Preprocessor</i>
SQL	<i>Structured Query Language</i>
AWA	<i>Arquivo Web Angolano</i>

ÍNDICE DE TABELAS

Tabela 1: Sites angolanos com domínios e respectivas percentagens	2
Tabela 2: Sites angolanos com domínios e respectivas percentagens - 2.....	3
Tabela 3: Backlog do projecto.....	19
Tabela 4: Requisitos Funcionais.....	20
Tabela 5: Requisitos não funcionais.....	22
Tabela 6: Descrição dos casos de uso.....	26
Tabela 7: Descrição do caso de uso pesquisar.....	26
Tabela 8: Descrição do caso de uso crawler.....	27
Tabela 9: Descrição do caso de uso cadastrar	27
Tabela 10: Descrição do caso de uso Arquivar	27
Tabela 11: Actores.....	28
Tabela 13: Cronograma do projecto	69

ÍNDICE DE FIGURAS

Figura 1: Arquitectura geral de um arquivo web.....	4
Figura 2: Web archive workflow.....	4
Figura 3: Página inicial do arquivo.pt	7
Figura 4: Resultado de uma pesquisa no arquivo.pt.....	7
Figura 5: Resultado de uma pesquisa no arquivo.pt.....	8
Figura 6: Página inicial do Internet Archive	9
Figura 7: Resultado de uma pesquisa no Internet Archive	10
Figura 8: Site da UAN em 22/06/2004, visualizado no Internet Archive	11
Figura 9: Padrão MVC	17
Figura 10: Diagrama de Casos de Uso do Sistema	24
Figura 11: Caso de uso gestão de arquivo	25
Figura 12: Caso de uso gestão de arquivo	25
Figura 13: Diagrama de sequência do caso de uso autenticação do utilizador	29
Figura 14: Diagrama de sequência do caso de uso arquivar	30
Figura 15: Diagrama de sequência do caso de uso pesquisar.....	31
Figura 16: Diagrama de estado – autenticação.....	32
Figura 17: Diagrama de classe	33
Figura 18: Modelo de classe persistente.....	34
Figura 19: Diagrama de objecto	35
Figura 20: Arquitectura	36
Figura 21:Protótipo da página inicial	37
Figura 22: Protótipo da página resultado da pesquisa	38
Figura 23: Protótipo do resultado de uma pesquisa	39
Figura 24: Protótipo da tela de login	39
Figura 25: Protótipo da tela inicial do BackOffice.....	40
Figura 26: Tela da página de arquivamento	41
Figura 27: Tela de erro ao arquivar	42
Figura 28: Tela da página de estatística	43
Figura 29: Tela da opção de exportar relatório	44
Figura 30: Tela com relatório	44
Figura 31: Tela da página de mensagem	45
Figura 32: Tela de cadastro de utilizador	46
Figura 33: Tela com a listagem dos utilizadores do sistema	47
Figura 34: Tela no motor de busca do Arquivo Web Angolano	48
Figura 35: Resultado de uma pesquisa no AWA	49
Figura 36: Visualização de uma pesquisa no AWA.....	50
Figura 37: Tela de login para o backoffice do AWA	51
Figura 38: Painel de dashboard do AWA.....	52
Figura 39: Tela de arquivamento do AWA	53
Figura 40: Erro de domínio	54
Figura 41: Registo dos utilizadores no AWA	55
Figura 42: Listagem dos utilizadores no AWA	56
Figura 43: Tela de estatísticas do AWA	57
Figura 44: Imprimir relatório no AWA	58
Figura 45: Relatório de uma pesquisa no AWA	59
Figura 46: Tela inicial do AWA Mobile	60

Figura 47: Resultado de uma pesquisa no AWA Mobile 61

ÍNDICE

DEDICATÓRIA.....	V
AGRADECIMENTOS	VI
RESUMO	VIII
ABSTRACT	IX
LISTA DE ABREVIATURAS E SIGLAS	X
ÍNDICE DE TABELAS	XI
ÍNDICE DE FIGURAS	XII
ÍNDICE	XIV
1 INTRODUÇÃO.....	1
1.1 JUSTIFICATIVA.....	2
1.2 HIPÓTESE	2
1.3 SITUAÇÕES PROBLEMÁTICAS.....	2
1.4 OBJECTIVO GERAL	3
1.5 OBJECTIVOS ESPECÍFICOS	3
1.6 MOTIVAÇÃO	3
1.7 ESTRUTURA DO DOCUMENTO	5
2 FUNDAMENTAÇÃO TEÓRICA DE ARQUIVOS WEB	1
2.1 DEFINIÇÃO DE WEB ANGOLANA	1
2.2 PERFIL DE AMOSTRA	2
2.3 ARQUITECTURA DE ARQUIVOS WEB	3
2.4 WORKFLOW	4
2.5 SISTEMAS EXISTENTES PARA ARQUIVAÇÃO DA WEB.....	6
2.5.1 <i>O Arquivo da Web Portuguesa (AWP)</i>	6
2.5.2 <i>Internet Archive (Wayback Machine)</i>	8
3 METODOLOGIAS	12
3.1. METODOLOGIA DE INVESTIGAÇÃO CIENTÍFICA.....	12
3.2. METODOLOGIA DE DESENVOLVIMENTO DE SOFTWARE	13
3.3 METODOLOGIA APLICADA AO PROJECTO	13
3.4 PLANEAMENTO DE QUALIFICAÇÃO DO PROJECTO	14
4 OPÇÕES TECNOLOGICAS	15
4.1 DESENVOLVIMENTO Web.....	15
4.2 JAVASCRIPT	15
4.3 PHP.....	15
4.4 KOTLIN.....	15
4.5 CODEIGNITER	16
4.6 BOOTSTRAP	16
4.7 MYSQL.....	16
4.8 UML	16
4.9 ASTAH.....	16
4.10 VISUAL STUDIO CODE	17

4.11	ANDROID STUDIO	17
4.12	XAMPP	17
4.13	PADRÃO DE DESENVOLVIMENTO E DE ARQUITECTURA DO SOFTWARE	17
5	CONSTRUÇÃO DO ARQUIVO WEB ANGOLANO	18
5.1.1	ÂMBITO DO PROJECTO.....	18
5.3	PRODUCT BACKLOG	18
5.4	ENGENHARIA DE REQUISITOS	19
5.5	REQUISITOS FUNCIONAIS.....	20
5.8.1	<i>Especificação de Casos de Uso.</i>	26
5.9	DIAGRAMA DE SEQUÊNCIA	28
5.9.1	<i>Mensagens</i>	28
5.9.2	<i>Conceitos</i>	28
5.10	DIAGRAMA DE ESTADOS	31
5.11	MODELO DE CLASSES DE DESENHO	32
5.12	MODELO DE CLASSES PERSISTENTE	34
5.13	DIAGRAMA DE OBJECTOS	35
5.14	ARQUITECTURA.....	36
5.15	PROTÓTIPOS DE BAIXA FIDELIDADE.....	37
5.15.1	<i>Página Inicial</i>	37
5.15.2	<i>Página Resultado da pesquisa</i>	37
5.15.3	<i>Pagina de Resultado da pesquisa final</i>	38
5.15.4	<i>Página de Autenticação - Login</i>	39
5.15.5	<i>Tela de Inical da Administração do Sistemas</i>	40
5.15.6	<i>Tela da página de Arquivamento</i>	40
5.15.7	<i>Erro no domínio</i>	41
5.15.8	<i>Página de Estatística</i>	42
5.15.9	<i>Gerar Relatório</i>	43
5.15.10	<i>Página de Mensagem</i>	45
5.15.11	<i>Página dos Utilizadores</i>	46
6	RESULTADOS E DISCUSSÃO	48
6.8	ACESSO A APLICAÇÃO.....	48
6.2	ADMINISTRADOR DO SISTEMA.....	50
6.10	SOLUÇÕES DE SEGURANÇA	64
6.11	<i>CONTROLO DE ACESSO</i>	64
6.12	<i>NEGAÇÃO DE SERVIÇO (DoS)</i>	64
6.13	<i>CSRF</i>	64
6.14	<i>CRYPTOGRAFIA</i>	64
6.15	<i>SQL INJECTION</i>	65
6.16	<i>PHP INJECTION E XSS</i>	65
7	CONCLUSÃO.....	66
7.1	RECOMENDAÇÕES E TRABALHO FUTURO	66
7.2	IMPACTO SOCIAL DO ARQUIVO WEB ANGOLANO.....	67
7.3	PRINCIPAIS DIFICULDADES	67

1 INTRODUÇÃO

Durante séculos, os historiadores analisaram a mídia impressa publicada no passado, como livros ou cartas, para pesquisar e escrever história. Portanto, as informações comuns de hoje serão o recurso de amanhã para a pesquisa histórica.

A web vem substituindo os meios impressos e grande parte das informações que caracterizam os nossos dias são publicadas exclusivamente online. Por exemplo, sites estão substituindo jornais e livros, blogs estão substituindo diários e galerias de fotos estão substituindo álbuns de fotos. Assim, a web reflete os dias atuais e é um recurso fundamental para apoiar a pesquisa em Humanidades.

A Web possibilita que cada um de nós disponibilize informação acessível a todos de uma forma rápida e económica. Diariamente, são publicados milhões de conteúdos na Web como textos, fotografias ou vídeos. A quantidade de informação que é publicada exclusivamente na Web tem vindo a aumentar rapidamente nos últimos anos. No entanto, passado relativamente pouco tempo, a grande maioria desta informação deixa de estar acessível e perde-se irremediavelmente.

Acontecimentos de grande importância para a História do Mundo, como por exemplo o evento UEFA Euro 2004 ou o 11 de Setembro, originaram ações de arquivo extraordinárias por parte do Arquivo.pt e do Internet Archive. É difícil para uma única organização fazer um arquivo exaustivo de todos os conteúdos publicados na Web, porque esta está em permanente mutação e muitos conteúdos desaparecem antes de poderem ser arquivados.

No entanto, a documentação de acontecimentos históricos de relevância nacional para Angola não é prioritária para o Arquivo.pt ou o Internet Archive e grande parte da informação publicada na Web angolana perde-se irremediavelmente.

Este problema é sentido igualmente por outras comunidades nacionais e pelo menos 46 países já iniciaram as suas próprias iniciativas de arquivo da Web.

Surge assim o interesse na criação de sistemas nacionais que contribuam para garantir a preservação dos conteúdos arquivados, em caso de falha do armazém central de conteúdos dos arquivos da Web internacionais.

1.1 Justificativa

O Arquivo Web Angolano é um sistema sem fins lucrativos que recolhe e arquiva conteúdos da Web à escala nacional. É difícil para um único sistema fazer um arquivo exaustivo de todos os conteúdos publicados porque a Web está em permanente mutação e muita informação desaparece antes de poder ser arquivada.

O presente projecto justifica-se pelo facto de não existir em Angola um sistema de *arquivamento* local voltado somente a web angolana podendo assim fornecer dados estatísticos confiáveis. Por isso, surgiu a necessidade de se desenhar e desenvolver um sistema de *arquivamento*, uma vez que a *Internet* é o melhor canal para se divulgar ou publicitar qualquer produto, devido ao amplo alcance e a facilidade de acesso que possui.

1.2 Hipótese

De forma a concordar com o já exposto nos parágrafos acima, trabalharemos a hipótese de que com o desenvolvimento e aplicação de uma solução computacional adequada à realidade da web angolana, para manter as informações que são necessárias às para criar um arquivo web angolano.

1.3 Situações Problemáticas

Dentre muitas questões, as que mais impulsionaram a escolha deste tema foram as seguintes:

- Dificuldades que os investigadores e utilizadores da web enfrentam na medição da web angolana;
- Dificuldade ao realizar-se uma colheita de informação para produção de estatísticas;
- Falta de controlo e instabilidade dos sites;

1.4 Objectivo Geral

Para solucionar os problemas da web angolana como foi exposto acima, determinou-se como o objectivo geral deste trabalho. Desenhar e Implementar de um sistema de Arquivamento para a Web Angolana especificamente para arquivamento e gestão do histórico da web angolana onde pretendemos cadastrar toda web angolana, desde o cadastro inicial ate durante as suas mudanças mediante o tempo, permitindo assim ter um histórico ou seja um arquivo que descreve a web angolana.

O objectivo do projecto do Arquivo da Web Angolana (AWA) é de desenvolver um sistema que terá como missão recolher, armazenar e preservar a informação publicada na Web angolana, proporcionando uma cobertura mais exaustiva da informação relacionada com Angola.

Podemos assim afirmar que para conseguirmos cumprir com os principais objectivos deste trabalho devemos mergulhar no conceito da web angola e assim tentar igualizar com o conceito internacional para não ferir com critérios internacionais.

1.5 Objectivos específicos

Os objectivos específicos deste trabalho são:

- Fornecer medições da evolução da Web angolana;
- Estudar os modelos de arquivamento da web já existentes;
- Descrever um plano de projecto para o desenvolvimento da ferramenta informática;
- Desenvolver e documentar a ferramenta informática.
- Criação de um aplicativo móvel;

1.6 Motivação

Após 1 ano, apenas 20% de um conjunto de endereços web se mantêm válidos. Ou seja, 8 em um universo de 10 websites se perderam. Diariamente, são publicados milhões de páginas na Web e a quantidade de informação que é exclusivamente publicada neste meio de comunicação tem vindo a aumentar rapidamente nos últimos anos. No entanto, passado relativamente pouco tempo, a grande maioria desta informação deixa de estar disponível e perde-se irremediavelmente. Surge assim o interesse no arquivo e preservação da informação publicada na Web angolana para que o conhecimento nela contida esteja acessível às nossas gerações futuras.

As razões que serviram como base para realizar este trabalho partem do facto de não existir um mecanismo para solucionar problemas como a falta de uma forma segura de arquivar e recuperar as informações da web angolana, falta de ferramentas eficientes, eficazes, uteis e adequadas à legislação angolana com custos reduzidos, por intermédio de uma ferramenta informática adequada à realidade angolana, adaptável às

características da web angolana, de forma a suprir as dificuldades que as mesmas enfrentam no desempenhar das suas atividades.

Ninguém sabe o tamanho real da rede mundial de computadores. De acordo com o Google, em 2008 a web tinha mais de um trilhão de páginas exclusivas. O rápido desenvolvimento das TIC (Tecnologia da Informação e Comunicação) teve um grande impacto neste crescimento. Na última década, a população mundial com acesso à Internet cresceu mais de 1.000% em algumas regiões. Dispositivos baseados em computador e telefones celulares com conectividade com a Internet são agora cerca de 5 bilhões, muitos dos quais equipados com tecnologia que capacita as pessoas a criar dados facilmente. Além disso, ferramentas como redes sociais, blogs e CMSs (Content Management Systems) tornaram mais fácil para as pessoas publicar e compartilhar dados. Essa combinação de fatores resultou na maior fonte de informação já criada.

A web tem um caráter democrático, onde todos podem publicar todo tipo de informação. Notícias, blogs, wikis, enciclopédias, entrevistas e opiniões públicas são apenas alguns exemplos dessa enorme lista. Parte dessas informações é única e historicamente valiosa. Porém, como a web é muito dinâmica, uma grande quantidade de informações é perdida todos os dias. Ntoulas et al. descobriram que 80% das páginas da web não estão disponíveis após um ano [Ntoulas et al., 2004]. Em alguns anos, eles provavelmente desaparecerão, criando uma lacuna de conhecimento para as gerações futuras. Muito do que foi escrito hoje não vai persistir e, como afirma a UNESCO, isso constitui um empobrecimento do patrimônio de todas as nações.

Os arquivos atuais da web são construídos com base na tecnologia de mecanismo de pesquisa da web. Essa parece ser a solução lógica, já que a web é o foco principal de ambos os sistemas. Eles coletam e exploram a web para criar índices especiais para pesquisá-la.

No entanto, os arquivos da web reproduzem as versões do documento armazenado o mais próximo possível do original em momentos distintos. Os mecanismos de pesquisa da web redirecionam apenas os usuários para servidores da web que hospedam os documentos.

Os arquivos da web permitem a pesquisa de vários instantâneos da web do passado, enquanto os mecanismos de pesquisa da web só permitem a pesquisa de um instantâneo do presente próximo. A missão dos arquivos da web é preservar a web, enquanto a única preocupação dos mecanismos de pesquisa na web é a pesquisa.

Dadas as diferenças acima entre os dois sistemas, formulamos a hipótese de que os usuários de mecanismos de pesquisa e arquivos da web têm diferentes necessidades de informação que deveriam ser tratadas de maneira diferente. O objetivo desta tese é entender essas diferenças e dar um passo à frente no desenvolvimento de abordagens de IR (recuperação de informações) para arquivos da web, que melhor atendam às necessidades de informação de seus usuários.

Para além do exposto acima, é crucial evidenciar, que de um modo geral, a web angolana não faz recurso de métodos e mecanismos arcaicos ou soluções não contextualizadas a realidade angolana, para preservar sua informação. Em consequência nota-se que não temos de pesquisa de dados, na extracção de dados estatísticos concisos para definição de estratégias e políticas para melhoramento do funcionamento da web angolana e auxílio no processo de tomada de decisão, morosidade na apresentação de relatórios dos processos da web. Estes foram alguns dos aspectos que deram propulsão na escolha deste trabalho.

1.7 Estrutura do documento

Este documento está estruturado da seguinte forma:

- O primeiro capítulo, a Introdução, visa contextualizar e apresentar ao leitor o desafio proposto para o trabalho.
- O segundo capítulo, apresenta um estudo realizado sobre a arquivação da web internacionalmente, permitindo uma visão geral do estado actual da implantação de sistemas em Angola.
- O terceiro capítulo, apresenta uma revisão bibliográfica sobre os conceitos e tecnologias pertinentes para a concretização do trabalho.
- O quarto capítulo, faz um panorama da construção da solução proposta, apresentando como foi o seu processo de desenvolvimento.
- O quinto capítulo, apresenta a solução desenvolvida, ilustrando as principais funcionalidades implementadas.
- O sexto capítulo, apresenta o processo de implantação do ARQUIVO WEB ANGOLANO.
- O sétimo capítulo, apresenta as conclusões do trabalho, referenciando também as contribuições e os trabalhos futuros.
- No final são apresentadas as referências bibliográfica.

2 FUNDAMENTAÇÃO TEÓRICA DE ARQUIVOS WEB

Este capítulo visa abordar o estudo realizado sobre o nível conceptual de arquivos da Web. O principal objectivo deste estudo foi a recolha de conceitos sobre os Sistemas de Arquivamento da Web existentes nas a nível global.

Os arquivos da Web são entidades que têm como objectivo preservar e disponibilizar a informação publicada na Web. A primeira iniciativa de arquivo da Web foi levada a cabo pelo Internet Archive.

Os arquivos da Web permitem que os pesquisadores pesquisem em alguns segundos milhões de documentos escritos de diferentes perspectivas. Como os documentos da web podem ser publicados por qualquer pessoa, eles fornecem testemunhos heterogêneos e originais em primeira pessoa sobre eventos históricos.

Acreditamos que os arquivos da web podem fornecer informações que contribuem para melhorar a pesquisa histórica em humanidades digitais. Por sua vez, os pesquisadores das ciências humanas podem fornecer contribuições valiosas, aplicando seu conhecimento de campo para selecionar e organizar informações da web de interesse manifesto a serem preservadas.

Os pesquisadores podem identificar sites que contêm informações valiosas a serem preservadas e sugerir para preservação. Por exemplo, o Web Archive of Catalonia oferece um formulário para receber sugestões de sites relacionados a esta comunidade autônoma da Espanha. Além disso, os pesquisadores podem contribuir gerando metadados adicionais e organizando informações que já estão arquivadas. Por exemplo, um pesquisador pode criar uma página da web que documenta um evento histórico e citar documentos arquivados na web como fontes de informação. Essa ação aumentaria a visibilidade, a disseminação e a reutilização de documentos arquivados da web.

Um pesquisador poderia identificar um fenômeno artístico do passado documentado na web e criar uma coleção temática de páginas arquivadas sobre ele. Essa ação enriqueceria e facilitaria o acesso às informações arquivadas na web.

2.1 Definição de Web angolana

Intuitivamente entende-se como Web angolana o conjunto de conteúdos publicados na Web de interesse para a comunidade de Angola. Esta definição é subjectiva e difícil de definir como um critério automático de selecção. No entanto, um sítio web referenciado por um nome sob o domínio .AO está por definição relacionado com Portugal. Assim sendo, é assumido que um conteúdo pertence à Web angolana se o domínio de topo do nome do seu sítio web respeita uma das seguintes condições:

1. Está sob a hierarquia .AO ou .CO.AO;
2. Não está sob a hierarquia .AO mas o conteúdo está embebido numa página alojada sob .AO. O objectivo é recolher a informação necessária para que as páginas angolanas arquivadas possam vir a ser apresentadas de forma completa.
3. Não está sob a hierarquia .AO mas existe um redireccionamento a partir de um nome sob .AO. Por exemplo, uma firma multinacional que regista o domínio .AO com a sua marca mas que o coloca a apontar para o sítio web principal da empresa que está sob a hierarquia .COM.

2.2 Perfil de Amostra

Como foi dito acima, estamos diante de uma problemática em que, é progressiva a perca das informações da web mundial em particular a angolana, para tal procuramos números próximos sobre as informações da web angolana.

A tabela 1, nos mostra os possíveis numero de sites angolanos e a sua percentagen na quantidade de site angolano, segundo o *hostingweb.com*, em que para tal os únicos dados sólidos sobre a web angola foi a de 2017:

Tabela 1: Sites angolanos com domínios e respectivas percentagens

Domínio	No. de Sites	Percentagem
.ao	72	7,33%
.co.ao	98	9,93%
.com	641	65,04%
.org	133	13,43%
.net	42	4,27%
Total	986	100%

Fonte: Autor

Também a nossa fonte mostrou-nos um dados super interessantes sobre a web angolana, que consiste que, as informações da web tem se perdido de forma exponencial e que vários sites têm saído do ar e com elas também se vão muita informação numa faixa simestral, a *tabela 2* mostra-nos que após 6 meses a web angolana é totalmente diferentes ou seja a maioria dos sites e informações são alteradas.

Tabela 2: Sites angolanos com domínios e respectivas percentagens - 2

Domínio	No. de Sites	Percentagem
.ao	72	7,33%
.co.ao	98	9,93%
.com	641	65,04%
.org	133	13,43%
.net	42	4,27%
Total	986	100%

Fonte: Autor

2.3 Arquitectura de arquivos web

Os arquivos da Web efectuam recolhas periódicas dos conteúdos publicados na Web. A recolha dos conteúdos é realizada por um Batedor. A actividade de um batedor consiste num processo cíclico.

Os conteúdos recolhidos são armazenados no Armazém de conteúdos. No fim da recolha, é realizado o processo de indexação dos conteúdos. Este processo consiste em gerar estruturas de dados que permitam realizar pesquisas rápidas e eficazes. A indexação é executada pelo Indexador, que pode criar índices de termos e URL. Os índices de termos permitem pesquisa por termos-chave, os índices de URL permitem a pesquisa por localização de um recurso na Internet.

Terminado o processo de indexação, os utilizadores podem pesquisar na informação recolhida, utilizando uma Interface de pesquisa. Esta interface permite que os utilizadores possam pesquisar na informação arquivada através dos índices. A Figura 1 ilustra a arquitectura geral de um arquivo da Web.

Fonte: arquivo.pt

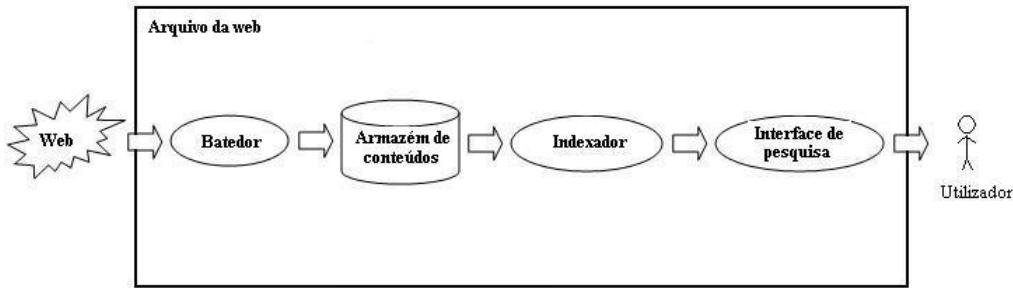


Figura 1: Arquitectura geral de um arquivo web

2.4 WorkFlow

Os dados da web passam por várias fases onde são transformados em pipeline até serem apresentados ao usuário. A Figura 2 ilustra o seguinte fluxo de trabalho:

Fonte: Google.com



Figura 2: Web archive workflow.

Aquisição: os dados da web podem ser adquiridos por vários caminhos, como de uma entidade que os arquivou anteriormente ou da digitalização de publicações em papel.

No entanto, o caminho mais comum é rastrear partes da web. Rastreamento é o processo de busca e coleta de dados. Ele começa com o download de um conjunto de URLs, que são então analisados para extrair os URLs aos quais se vinculam. Este processo é repetido continuamente para os URLs extraídos que ainda não foram baixados, até que uma condição de parada seja atendida. A decisão do que arquivar é difícil, pois não há espaço de armazenamento suficiente para guardar tudo e a web está em constante crescimento.

Armazenamento: os dados da web são armazenados de forma persistente na memória secundária. Normalmente, os arquivos da web concatenam sequências de documentos da web compactados em arquivos longos perto de 100 MB, onde cada documento é precedido por um pequeno cabeçalho.

Este formato é denominado ARC e foi originalmente desenvolvido pelo Internet Archive. Oferece uma maneira mais fácil de gerenciar e agilizar o acesso aos documentos, uma vez que os sistemas de arquivos têm dificuldade em lidar com bilhões de arquivos. Os documentos da web e seus sites podem passar por

vários processos durante ou após o armazenamento. Por exemplo, eles podem ser enriquecidos com metadados descritivos ou sua qualidade pode ser avaliada com uma medida de integridade.

Indexação: os dados da web são lidos do armazenamento, descompactados, divididos em palavras (tokenizados) e analisados sintaticamente (analisados). A análise é necessária para distinguir o texto dos metadados e analisar a que parte do documento o texto pertence. É um desafio porque existem centenas de formatos de arquivo que devem ser tratados e continuar a evoluir, como HTML, PDF ou novos formatos. Outros processos podem ser aplicados, como a redução morfológica de palavras à sua forma de raiz (radical) ou a eliminação de palavras de alta frequência chamadas de palavras irrelevantes, que não têm impacto na discriminação do texto (por exemplo, o e de). Em seguida, estruturas de índice sobre as palavras e os metadados são criadas para uma pesquisa eficiente.

Pesquisa: as estruturas de índice são usadas para pesquisar os documentos que correspondem a uma consulta recebida. Essa correspondência depende do modelo de recuperação implementado. Normalmente, para coleções em grande escala, como a web, um modelo é escolhido onde todos os termos de consulta devem ocorrer nos documentos correspondentes.

Os documentos são então classificados por suas pontuações de relevância que medem o quanto bem eles atendem às necessidades de informações de um usuário. Essa necessidade é formalmente representada por uma consulta. Visto que milhões de documentos podem corresponder a uma consulta, classificar os documentos por relevância é essencial para encontrar as informações desejadas com eficácia. Essa classificação é calculada com um conjunto de heurísticas, com base em vários recursos, como a proximidade dos termos ou o número de links que um documento recebe.

Apresentação: os resultados são formatados e exibidos em listas classificadas para consumo do usuário final. Normalmente, cada resultado é aumentado com metadados, como título, URL e carimbo de data / hora de quando foi arquivado. Os resultados também podem ser agrupados por tempo para uma percepção mais fácil de sua distribuição temporal ou exibidos ao longo de uma linha do tempo para apoiar as tarefas de exploração. Quando um documento arquivado é mostrado, todos os seus hiperlinks são alterados para que as referências apontem para o arquivo em vez da web ao vivo. Isso permite que os usuários naveguem interativamente na web como antigamente.

Preservação: é um processo paralelo neste workflow, para garantir que os documentos web fiquem acessíveis a longo prazo. Portanto, os dados devem ser replicados dentro do data center e entre data centers espalhados por diferentes localizações geográficas, para evitar todos os tipos de falhas. Os dados também devem ser armazenados de maneira à prova de falsificação para evitar que alguém reescreva o histórico. Pessoas mal-intencionadas podem tentar tirar proveito desse fato em seu próprio benefício. O monitoramento da obsolescência potencial em formatos de arquivo e tecnologia deve ser constante para uma migração oportuna dos dados antes que eles não sejam mais acessíveis ou utilizáveis.

Este trabalho enfoca principalmente os processos de indexação e busca, apesar de todos os demais processos do workflow terem influência no resultado final. Esses dois processos geralmente envolvem dois sistemas trabalhando em conjunto. Um é o sistema de busca, cujo objetivo é criar índices sobre os dados armazenados e usá-los para acelerar a correspondência de documentos que atendem a uma consulta. O sistema de classificação então usa os dados indexados dos documentos correspondentes para estimar a relevância da consulta. Os documentos que correspondem à consulta são classificados em ordem decrescente por sua pontuação de relevância, o que permite aos usuários encontrar informações de maneira eficaz e eficiente.

2.5 SISTEMAS EXISTENTES PARA ARQUIVAÇÃO DA WEB

2.5.1 O Arquivo da Web Portuguesa (AWP)

O projecto de Arquivo da Web Portuguesa foi iniciado em 2007 na Fundação para a Computação Científica Nacional. Desde 2008 que o AWP faz recolhas exaustivas da Web portuguesa 4 vezes por ano, incluindo ficheiros de todos os tipos (ex. HTML, imagens, PDF).

A recolha é feita com restrições para garantir o bom funcionamento do Arquivo e não prejudicar o normal funcionamento dos sítios Web visitados. Por exemplo, é respeitada uma pausa entre pedidos a um mesmo sítio da Web para não o sobrecarregar. Apenas é arquivada informação pública e não são preenchidos formulários. Como tal, todas as páginas protegidas por palavra-passe ou outros mecanismos de restrição de acesso não são recolhidas. Cerca de 90% dos ficheiros da Web portuguesa são recolhidos ao fim de 7 dias. No entanto, a recolha continua para os sítios mais lentos ou com maior número de conteúdos.

O AWP adoptou três sistemas pertencentes ao projecto Archive-Access: o sistema de recolha Heritrix, o sistema de pesquisa NutchWAX e o sistema de acesso a conteúdos arquivados Wayback. O sistema de recolha foi adoptado sem alterações significativas. Já os sistemas de pesquisa e acesso tiveram de modificados para que os objectivos do AWP fossem alcançados e satisfeitas as necessidades da comunidade portuguesa:

- A interface de utilização foi traduzida para português;
- A usabilidade e grafismo das páginas foram melhorados;
- Os índices foram reestruturados para permitirem respostas mais rápidas;
- A ordenação dos resultados foi melhorada para permitir fornecer resultados mais relevantes.

Abaixo se segue algumas telas do Arquivo.pt:

Fonte: arquivo.pt

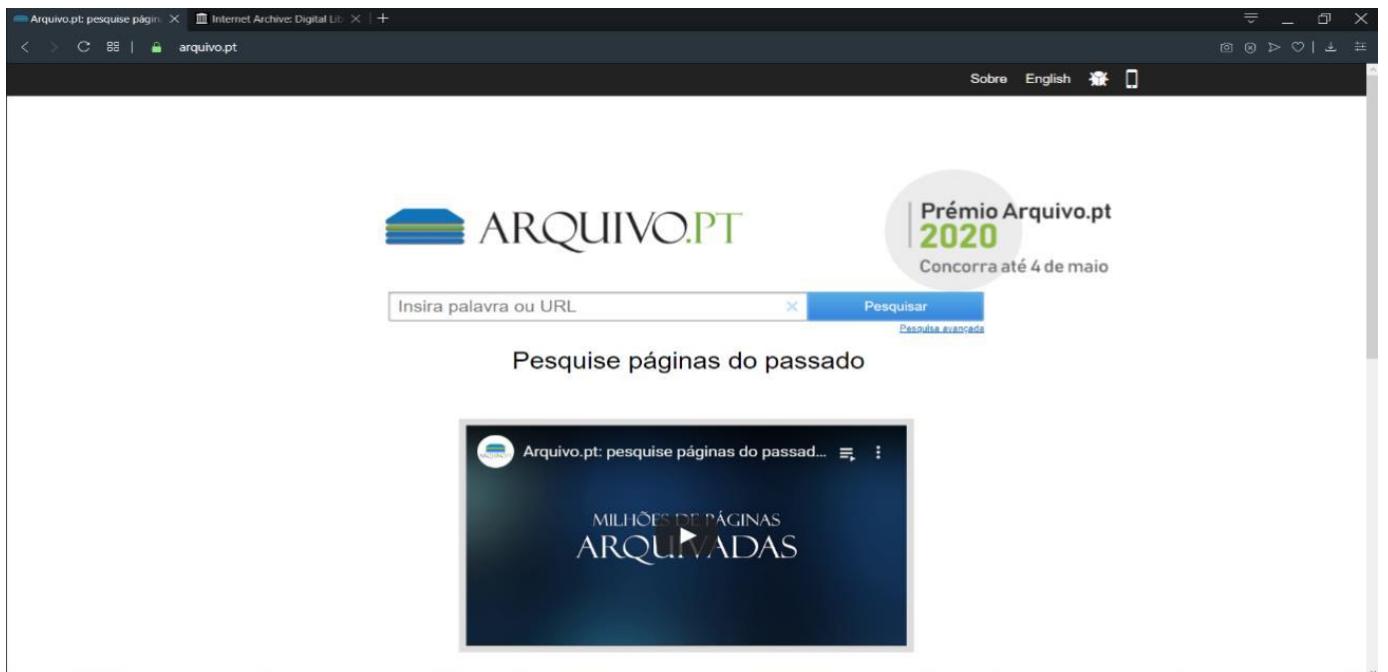


Figura 3: Página inicial do arquivo.pt

Esta é a sua tela de pesquisa, com o seus respectivos resultados:

Fonte: arquivo.pt

 A screenshot of a web browser window showing search results for the query "angola" on arquivo.pt. The search bar shows "angola" and date filters "entre 01/01/1996 e 31/12/2019". The results are listed under "Páginas" and include:

- 301 Moved Permanently**
18 Dezembro, 2009 - Listar versões
301 Moved Permanently Moved Permanently The document has moved here ...
<http://blog.comunidades.net/angchimbarr71>
- 301 Moved Permanently**
3 Agosto, 2010 - Listar versões
301 Moved Permanently Moved Permanently The document has moved here ...
<http://blog.comunidades.net/angchimbarr71>
- PNUD em Angola**
6 Dezembro, 2017 - Listar versões
PNUD em Angola PNUD no mundo **Angola** Objetivos de Desenvolvimento Sustentável Sobre o PNUD Mais Notícias
Após a resposta nacional ao VHNSIDA em **Angola** recebe nova subvenção do Fundo Global para o ... Charlotte Kristiansson, salientou que a nova subvenção do Fundo Global disponibilizada à **Angola**, tem ...
<http://www.ao.undp.org/>
- Jobs in Angola - Search Angola Jobs - XpatJobs**
24 Novembro, 2014 - Listar versões
Jobs in **Angola** - Search **Angola** Jobs - XpatJobs Follow @xpatjobs xpat jobs on Xpat Jobs Xpat Rentals Xpat Sales Xpat Pages Connecting you to 23,050 jobs in **Angola** What looking for multilingual jobs ... Andorra **Angola** Argentina Australia Austria Bahrain Bangladesh Belarus Belgium Belize Bolivia Bosnia ...
<http://angola.xpatjobs.com/>
- Angola – Wikipédia, a encyclopédia livre**
8 Abril, 2015 - Listar versões
Angola – Wikipédia, a encyclopédia livre **Angola** Origem: Wikipédia, a encyclopédia livre. Ir para: navegação , pesquisa
Nota: Para outros significados, veja **Angola** (desambiguação). Coordenadas: 11° 54' S 17° 12' E **Angola** República da

Figura 4: Resultado de uma pesquisa no arquivo.pt

Resultado final de uma pesquisa feita:

Fonte: arquivo.pt

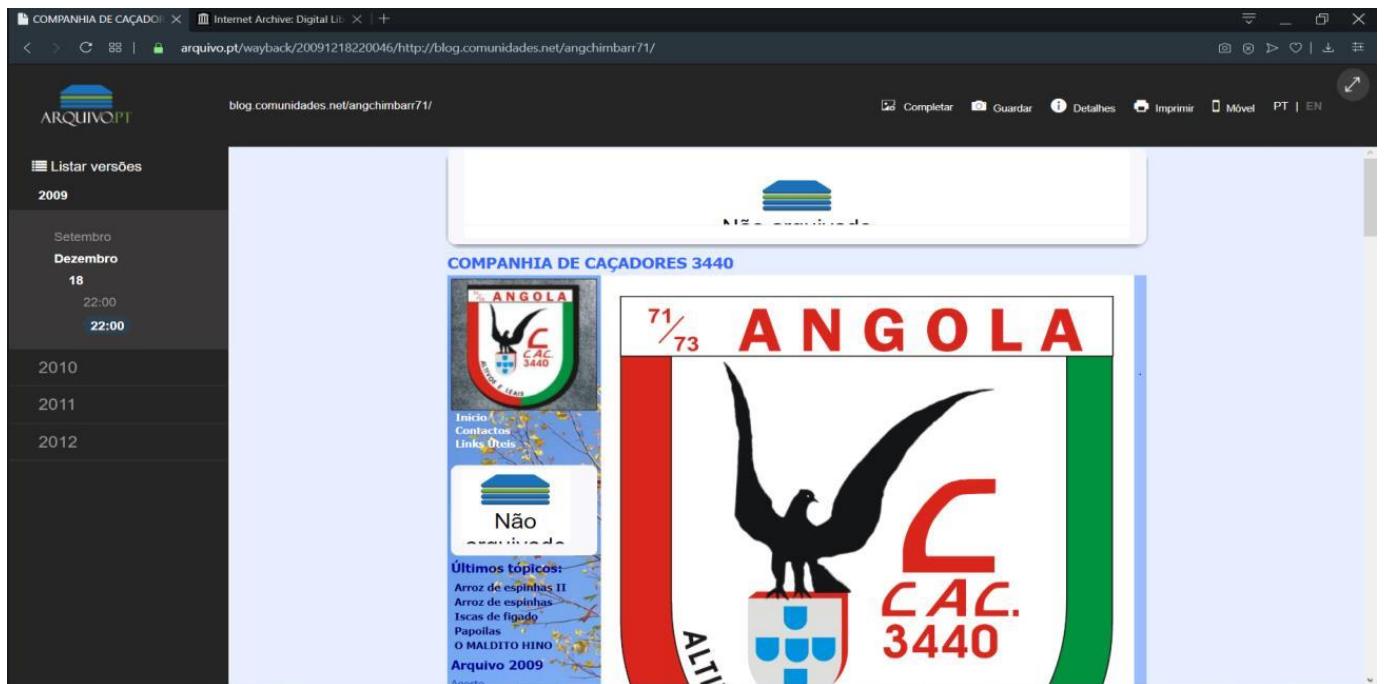


Figura 5: Resultado de uma pesquisa no arquivo.pt

2.5.2 Internet Archive (Wayback Machine)

O Internet Archive foi criado em 1996 com o objetivo de manter um registro histórico da World Wide Web. Seu objetivo é preservar o conhecimento e a cultura humana criando uma biblioteca na Internet para pesquisadores, historiadores e acadêmicos. Em 2003, sua coleção total era de cerca de 100 terabytes, crescendo 12 terabytes por mês; hoje, inclui mais de 40 bilhões de páginas da web e ocupa um petabyte de armazenamento, crescendo 20 terabytes por mês.

O Internet Archive usa o nome "Wayback Machine" para o seu serviço que permite que arquivos da World Wide Web possam ser pesquisados e acessados. Este serviço permite aos usuários visualizarem páginas da web arquivadas. O Wayback Machine foi criado como um esforço conjunto entre a Alexa Internet e o Internet Archive quando um índice tridimensional foi construído para permitir a navegação no conteúdo de páginas da web arquivadas. Milhões de sites e seus dados associados (imagens, código fonte, documentos, etc) são salvos em um banco de dados gigantesco.

O serviço pode ser usado para ver as versões anteriores de sites usados para parecer, para pegar o código fonte original a partir de sites que podem não ser diretamente disponíveis, ou a visitar sites que já nem sequer existem.

Segue-se algumas telas do Internet Archive:

Fonte: archive.org

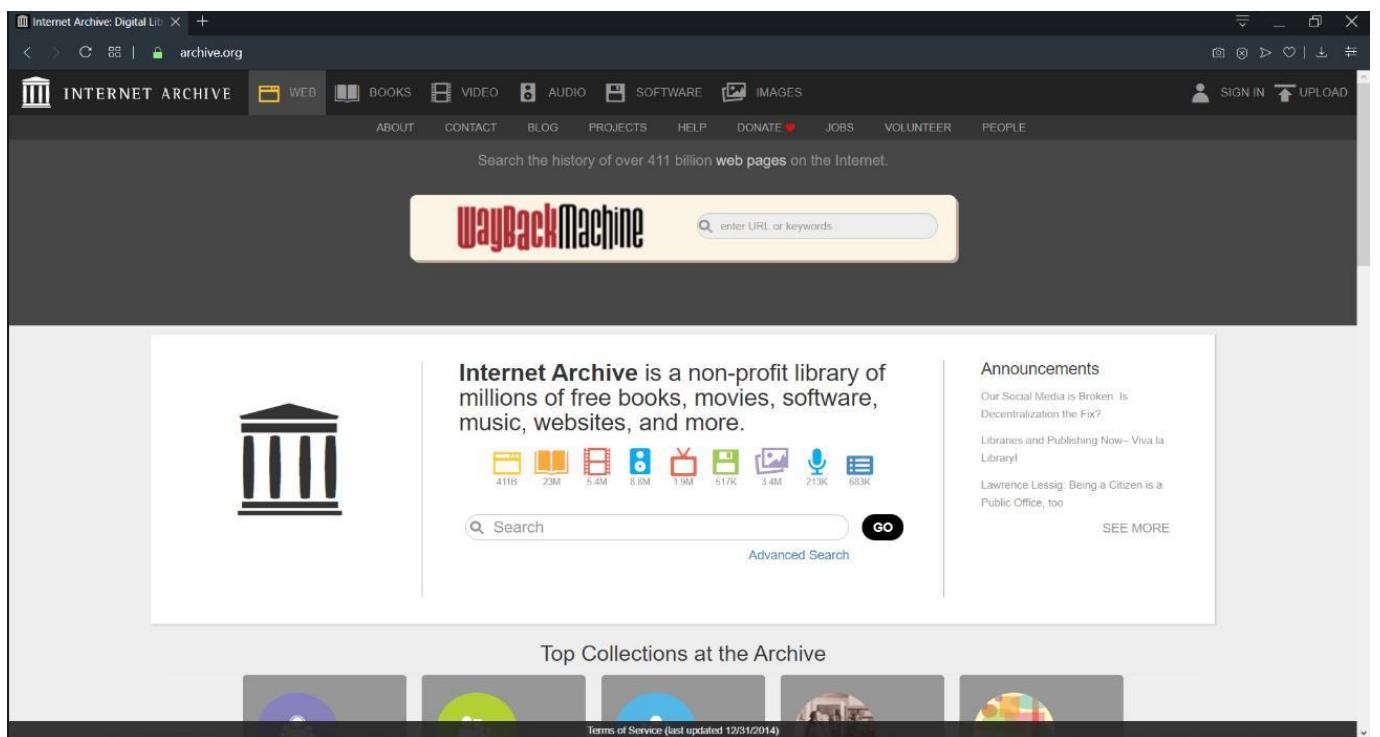


Figura 6: Página inicial do Internet Archive

Esta é a sua tela de pesquisa, com o seus respectivos resultados:

Fonte: archive.org

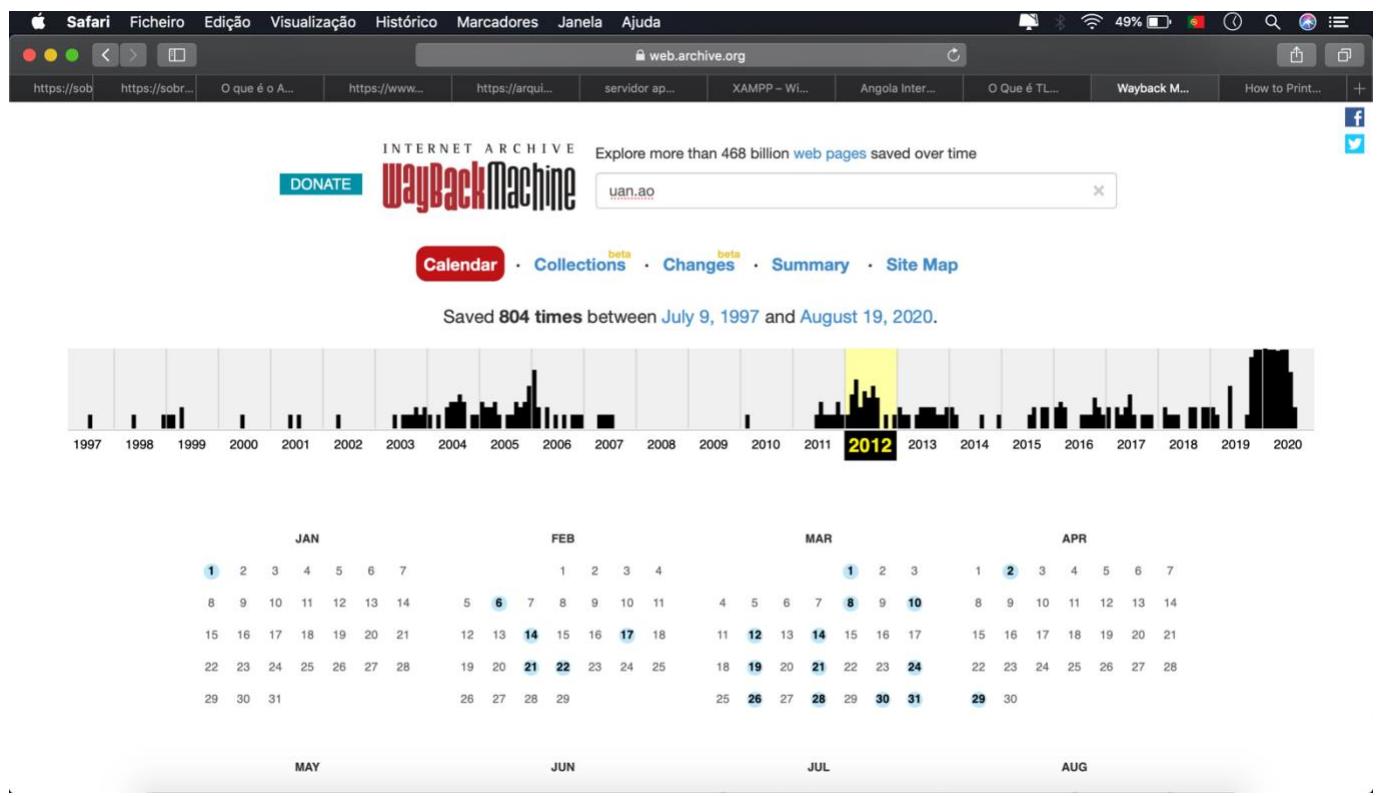


Figura 7: Resultado de uma pesquisa no Internet Archive

Esta é a sua tela de pesquisa, com o seus respectivos resultados:

Fonte: archive.org

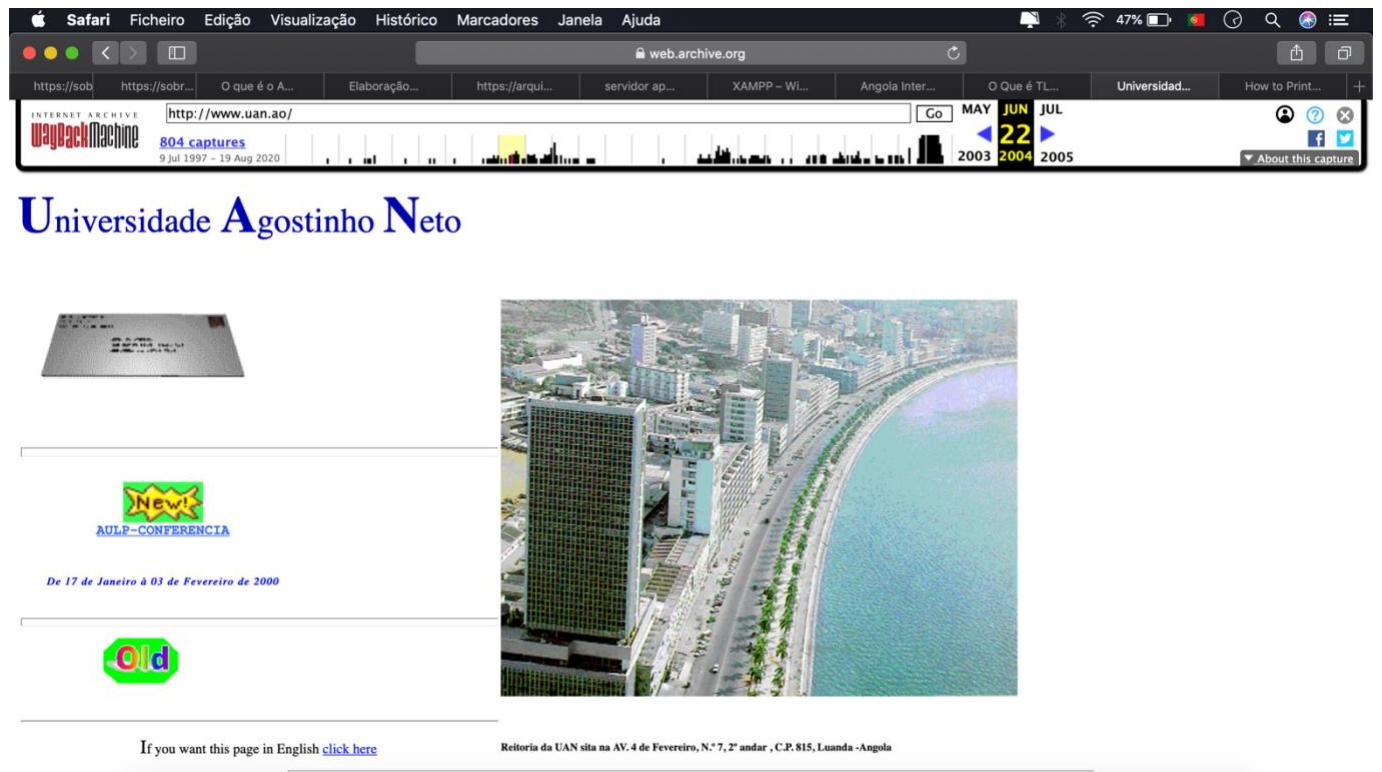


Figura 8: Site da UAN em 22/06/2004, visualizado no Internet Archive

3 METODOLOGIAS

3.1. Metodologia de Investigação Científica

Para a obtenção dos conhecimentos relacionados a arquivamento da web, recorreu-se a operações mentais e técnicas que possibilitessem a sua verificação. Um método pode ser definido como caminho para se chegar a um determinado fim. Cientificamente “é o conjunto de procedimentos intelectuais e técnicos adoptados para atingirmos o conhecimento” (Prodanov & Freitas, 2013). Já os procedimentos técnicos estão relacionados ‘à maneira como os dados para a elaboração da pesquisa são obtidos’ (Prodanov & Freitas, 2013).

Desta forma, para alcançar os objectivos traçados, foram utilizados os métodos e as técnicas seguintes no presente projecto:

Método dedutivo (que é um método de abordagem): fez-se uma abstração dos fenómenos em todos os sectores de arquivo da web de forma geral, e particularmente o arquivo.pt, para que se definisse os procedimentos lógicos a serem seguidos durante a investigação.

Método comparativo: permitiu confrontar as soluções existentes a partir das suas principais funcionalidades, de modo a verificar diferenças ou semelhanças com a aplicação proposta.

Quanto à colecta de dados, foram utilizadas as seguintes técnicas:

Pesquisa bibliográfica: fez-se a revisão de algumas bibliografias que compõem pareceres, informações e estudos sobre o arquivamento da web.

Pesquisa de campo: visita em sites e iniciativas de arquivamento para se verificar a correlação existente entre os diversos processos previamente estudados.

Observação: recolha e registro da realidade nas comunidades referidas na pesquisa de campo.

Testes: o modelo de testes é qualitativo, pois será importante verificar o nível de aceitabilidade da solução e também se esta realmente diminui as dificuldades que os potenciais utilizadores enfrentam.

Pesquisa de mercado: foram feitas pesquisas para se conhecer as tecnologias já implantadas voltadas ao arquivamento da web, com o fim de se levantar parte dos requisitos necessários, para uma melhor adequação do que realmente se precisa para a realidade nacional. Durante a realização do estudo, fez-se uma síntese baseando-se na análise dos pontos fortes e fracos apresentados pelas mais variadas tecnologias de informação e de comunicação já implantadas para o sector do arquivamento da web.

Recorreu-se ao Google Scholar, uma das mais abrangentes bases de dados, devido a variedade de artigos científicos sobre diversos temas, e ao Google Search, o maior e mais utilizado motor de busca na *Internet*. A pesquisa incluiu estudos e artigos sobre o agronegócio e os que o relacionam com a área de *softwares* e, também, aplicações já implantadas voltadas a este sector.

3.2. Metodologia de Desenvolvimento de Software

Ainda não há consenso sobre qual a melhor prática para o desenvolvimento de *softwares*, mas existe um esforço em encontrar soluções para reduzir as dificuldades da natureza do *software*, da sua complexidade, de necessidades de se cumprir com os seus objectivos e da rapidez com que sofre alterações. No entanto, a adoção da disciplina é primordial para lidar com essas dificuldades, com o fim de se reduzir ao máximo a influência delas no processo de desenvolvimento de sistemas.

E, de facto, o melhor processo de *software* é aquele cujas pessoas que realizam o trabalho se sintam próximas. Tal processo só será efectivo se for aberto as adaptações que atendam às necessidades da equipa do projecto, porque só se desenvolve um processo que melhor se adequé e atenda às necessidades mais amplas da equipa. Alternativamente, a própria equipa pode criar o seu próprio processo e atender às suas necessidades mais específicas. (Pressman, 2011).

Watts Humphrey, citado por Pressman, afirma que é, sim, possível criar um processo de *software* pessoal e/ou de equipa; embora requeiram trabalho árduo, treinamento e coordenação, mas são alcançáveis.

Essa abordagem justifica a escolha que se fez das tarefas, escolhidas de acordo com as necessidades do presente projecto, para melhor controlo dos objectivos e das tarefas a realizar para que eles sejam atingidos. Pela natureza do projecto, não foi possível usar a metodologia que mais se adapta às nossas necessidades na íntegra.

3.3 Metodologia aplicada ao projecto

O XP é uma metodologia ágil de desenvolvimento de *softwares* focada no desenvolvimento de projectos com qualidade, que tem como princípios básicos a simplicidade e aceitar mudanças incrementais.

A seguir são listadas e apresentadas as fases e/ou actividades que pertinentes ao projecto para a realização das tarefas necessárias.

Planning Game (Jogos de Planeamento): planeamento incremental das funcionalidades do sistema e sua priorização.

Small Releases (Pequenas Versões): pequenas versões funcionais do projecto.

Metaphor (Metáforas): entendimento da realidade dos *stakeholders*.

Customer Tests (Testes de Aceitação): presença dos utilizadores no desenvolvimento para partilha de ideias sobre o funcionamento do sistema, para ver-se se está de acordo com as especificações. Para isso, recorrer-se-á também ao TDD.

Test Driven Development (Desenvolvimento Orientado a Testes): para se saber o que deve ser verdadeiro no sistema e o que não deve ser para que ele funcione correctamente, pensando primeiramente no problema e depois na solução. Vale salientar que será primordial o recurso aos testes de unidade e de aceitação, sem desprimor para outros da qual poderá se recorrer.

Refactoring (Refactoração): melhoria contínua da programação, com o mínimo possível de erros e reaproveitamento de código-fonte.

3.4 Planeamento de qualificação do projecto

Ao passo que os softwares se tornam cada vez mais integrados em todas as actividades das nossas vidas, clama-se por maior qualidade. Tem-se gastado muito dinheiro em *softwares* que não apresentam as características e funcionalidades prometidas e uma falha grave no sistema poderia tornar inutilizáveis importantes infraestruturas, aumentando o custo. Porém, a realização dos testes pode ajudar neste sentido.

Todas as pessoas podem realizar o teste, para que se verifique se atende às necessidades do cliente, execute de forma precisa e confiável e gere valor para todos os que utilizam o *software*. A realização dos testes pode garantir que o trabalho foi realizado correctamente, mas só se se verificar os erros antes da entrega e os defeitos durante a sua produção.

Segundo Garvin, pode-se medir a qualidade do *software* em oito dimensões, das quais destacamos:

- *Qualidade do desempenho*: o *software* fornece todo o conteúdo, funções e recursos que são especificados como parte do modelo de requisitos de forma a gerar valor ao utilizador final?
- *Qualidade dos recursos*: *software* fornece recursos que surpreendem e encantam utilizadores finais que os utilizam pela primeira vez?
- *Confiabilidade*: o *software* fornece todos os recursos e capacidades sem falhas? Está disponível quando necessário? Fornece a funcionalidade sem a ocorrência de erros?
- *Conformidade*: o software está de acordo com os padrões de *software* locais e externos relacionados com a aplicação? Segue as convenções de projecto e codificação de fato? Por exemplo, a interface com o utilizador está de acordo com as regras de projecto aceitas para a entrada de dados?
- *Durabilidade*. O *software* pode ser mantido (modificado) ou corrigido (depurado) sem a geração involuntária de efeitos colaterais indesejados? As correções farão com que a taxa de erros ou a confiabilidade diminuam com o passar do tempo?
- *Facilidade de manutenção*. O *software* pode ser mantido (modificado) ou corrigido (depurado) em um período de tempo aceitável e curto? A equipa de suporte pode obter todas as informações necessárias para realizar alterações ou corrigir defeitos?

Outro bom recurso é o teste caixa-preta (comportamental), que se foca nos requisitos funcionais do software. Ele tenta encontrar erros nas funções incorrectas ou que estejam em falta, erros de estruturas de dados ou acesso a base de dados externas e ainda erros de comportamento ou de desempenho. Este teste é essencial para se responder às seguintes questões bastante pertinentes:

- Como o comportamento e o desempenho do sistema é testado?
- O sistema é particularmente sensível a certos valores de entrada?
- Como as fronteiras de uma classe de dados é isolada?
- Que taxas e volumes de dados o sistema pode tolerar?
- Que efeito combinações específicas de dados terão sobre a operação do sistema?

4 OPÇÕES TECNOLOGICAS

Nota-se a enorme evolução que a área de Tecnologias da Informação (TI) tem sofrido, e em paralelo a produção de software. O presente capítulo vem apresentar uma visão geral acerca dos conceitos referentes as tecnologias, ferramentas e metodologias para o desenvolvimento de software, aspectos estes utilizados como base para o desenrolar da solução apresentada. Neste capítulo são apresentados breves conceitos e definições cruciais para a compreensão dos assuntos que são tratados ao longo do trabalho.

4.1 Desenvolvimento Web

Visando a escalabilidade e a facilidade de acesso, o ARQUIVO WEB ANGOLANO foi desenvolvido utilizando as tecnologias para desenvolvimento Web. O desenvolvimento web refere-se a um processo de construção e testes do software específico para a web, com a finalidade de se obter um conjunto de programas, que satisfaçam as funções pretendidas, quer em termos de usabilidade dos utilizadores ou compatibilidade com outros programas existentes, permitindo que os nossos utilizadores finais possam utilizar o sistema, em qualquer ponto do mundo.

Apresentamos em seguida as principais tecnologias utilizadas no desenvolvimento do sistema de Arquivo.pt.

4.2 JavaScript

JavaScript é uma linguagem de programação interpretada. Que foi implementada como parte dos navegadores web para que scripts pudessem ser executados do lado do cliente e interagissem com o utilizador sem a necessidade deste script passar pelo servidor, controlando o navegador, realizando comunicação assíncrona e alterando o conteúdo do documento exibido. Bem como para a realização de algumas validações a nível do cliente (Pereira & Poupa, 2011).

4.3 PHP

O PHP (um acrônimo recursivo para PHP: Hypertext Preprocessor) é uma linguagem de script open source de uso geral, muito utilizada, e especialmente adequada para o desenvolvimento web e que pode ser embutida dentro do HTML.

PHP é uma linguagem de programação utilizada por programadores e desenvolvedores para construir sites dinâmicos, extensões de integração de aplicações e agilizar no desenvolvimento de um sistema. Atualmente, o PHP está na versão 7.2

4.4 Kotlin

Kotlin foi criada pela JetBrains com o objetivo de ser uma linguagem de programação concisa e segura para o desenvolvimento web, mobile e desktop, com um amplo conjunto de ferramentas e IDEs fáceis de utilizar.

O Kotlin é uma linguagem estaticamente tipada, desenvolvida pela JetBrains, cuja sintaxe é mais expressiva e concisa do que a do Java. Com recursos como expressões lambda, sobrecarga de operadores, templates de strings e muito mais.

4.5 CodeIgniter

Para desenvolvimento de páginas web, existem várias tecnologias que aumentam a produtividade e a facilidade de implementação de várias necessidades.

Para este projecto utilizamos o codeigniter como framework por ser de fácil instalação, manutenção e implementação. A comunidade é maior o que facilita na resolução de problemas.

4.6 Bootstrap

O Bootstrap é um *framework frontend* com recurso a utilização do JavaScript e CSS para, por meio de classes pré-definidas, tornar páginas HTML responsivas, ou seja, adaptáveis as dimensões de tela de diferentes dispositivos electrônicos.

4.7 MYSQL

Os dados que trafegam na aplicação precisam de ser armazenados, utilizamos o Mysql como gestão da base de dados.

Foi escolhido o SGBD MySQL, devido a sua facilidade de uso e por possuir versões open source e principalmente pela sua compatibilidade e integração com as linguagens de plataforma web, especificamente o PHP que foi utilizado no desenvolvimento da aplicação.

4.8 UML

O UML (Unified Modeling Language) é uma linguagem para especificação, visualização, construção e documentação de componentes de software orientados por objectos. Por meio de seus diagramas é possível representar sistemas de software sob diversas perspectivas de visualização. Permite a comunicação de todas as pessoas envolvidas no processo de desenvolvimento de um sistema.

Existem actualmente no mercado várias ferramentas que auxiliam o desenvolvimento de software recorrendo ao UML. Neste projecto foi utilizada a ferramenta: Astah.

4.9 ASTAH

O Astah é uma ferramenta de design de para projectos de *software*. Ele suporta UML, e é utilizado para o desenvolvimento de diagramas de *software*, tal como fluxogramas, diagramas de casos de uso, diagramas de actividades, etc. Possui funcionalidades de geração de código por meio de certos tipos de diagramas. Esta ferramenta foi utilizada para o desenvolvimento dos diagramas que fazem parte do sistema.

4.10 Visual Studio Code

O Visual Studio Code é um editor de código-fonte desenvolvido pela Microsoft para Windows, Linux e macOS.

Ele inclui suporte para depuração, controle Git incorporado, realce de sintaxe, complementação inteligente de código, snippets e refatoração de código.

4.11 Android Studio

Android Studio é um ambiente de desenvolvimento integrado (IDE) para desenvolver para a plataforma Android.

4.12 XAMPP

O XAMPP é um pacote com principais servidores independente de plataforma, que consiste principalmente na base de dados MySQL, o servidor Apache e os interpretadores para linguagens de script PHP e Perl. É um método que torna extremamente fácil para os desenvolvedores criar um servidor *web* local para fins de teste durante o desenvolvimento.

4.13 Padrão de desenvolvimento e de arquitectura do software

O paradigma escolhido é o Orientado à Objectos, pela sua difusão e evolução, principalmente por questões voltadas para a segurança, o reaproveitamento do código, a modularização, a estrutura de dados e a manutenção do código.

Por outro lado, trabalhar-se-á com o padrão arquitectural MVC, para se separar a representação da informação da interacção do utilizador com ela. Conforme ilustra a figura 15, pode-se ver a separação das suas três camadas e se ter uma ideia de como este padrão de arquitectura funciona a ajuda na circulação da informação no sistema.

Fonte: google.co.ao

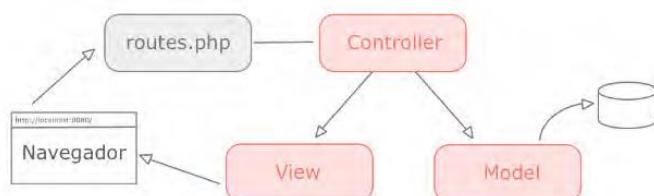


Figura 9: Padrão MVC

5 CONSTRUÇÃO DO ARQUIVO WEB ANGOLANO

A modelação constitui uma fase muito importante no processo de desenvolvimento de qualquer sistema de *software*. A prática da engenharia de *software* consiste em princípios, conceitos, métodos e ferramentas que devem ser considerados ao planear e desenvolver um *software*. Direccionam a acção e estabelecem a infraestrutura a partir da qual o processo de desenvolvimento será conduzido. Este capítulo apresenta as actividades aplicadas ao projecto, de acordo com a metodologia de desenvolvimento escolhida.

O presente capítulo traz uma visão geral sobre o processo de desenvolvimento da solução ao problema proposto para este trabalho. São ilustradas neste capítulo as etapas relevantes sobre a construção do ARQUIVO WEB ANGOLANO.

Conforme indicado no capítulo 3, o processo de desenvolvimento do ARQUIVO WEB ANGOLANO, seguiu uma abordagem ágil, de forma a conseguir-se uma maior adaptação e flexibilidade no processo de desenvolvimento, e também a apresentação de resultados em pouco tempo.

5.1.1 Âmbito do Projecto

O ARQUIVO WEB ANGOLANO é um software voltado ao arquivação e gestão das informações relativas a web angolana.

É de responsabilidade do ARQUIVO WEB ANGOLANO o acompanhamento total das informações da web angolana desde o seu cadastramento primário ate as suas alterações ao longo do tempo, visto que o propósito é melhorar o processo de gestão e acesso a conteúdos da web angolana sem se ter em conta o tempo.

Pretendemos com o passar do tempo disponibilizar os seguintes serviços:

- Pesquisa histórica por termo: permitirá identificar páginas arquivadas ao longo dos anos que contenham determinadas palavras;
- Pesquisa histórica por endereço da web (URL): permitirá identificar várias páginas arquivadas ao longo dos anos referenciadas por um determinado URL;
- Coleções históricas de conteúdos web para fins de investigação: a disponibilização de coleções da web permitirá que estes investigadores possam processar informação localmente nos seus computadores sem terem de realizar recolhas da web;

5.3 Product Backlog

A montagem do *product backlog* foi efectuada com base no levantamento de funcionalidades no *product owner*, onde as prioridades foram definidas com base na importância para a arquivação e gestão da informação da web angolana.

Nota: quanto maior for o número da prioridade, menor prioridade o requisito tem.

Tabela 3: Backlog do projecto

Funcionalidade	Prioridade	Custo-Hora
Modelagen de Dados	1	74
Guardar Arquivo (Crawler)	2	183
Guardar Ficheiro (Crawler)	3	90
Motor de Busca	4	137
Robo de Arquivação	5	147
Gestão de Arquivos	6	92
Web-Service	7	97
Arquivo Mobile	9	199

Fonte: arquivo.pt

Com o *product backlog* elaborado, seguiu-se a definição das metas dos sprints. Esta definição foi realizada por intermédio reuniões de planeamento de sprints, onde o projecto foi apresentado pelo *product owner* aos demais integrantes da equipa *Scrum*, que definiram com base na experiência o custo em horas que cada tarefa levou, levando em conta todos os aspectos técnicos.

Antes de entramos nos sprints é importante afirmar que todo o trabalho sobre as tarefas foi realizado obedecendo ao seguinte ciclo interno:

- Desenvolvimento: analisar, desenhar, implementar, testar e documentar;
- Integração: acoplar as partes desenvolvidas ao produto final;
- Homologação: verificação do que foi feito;
- Ajustes: qualquer mudança nos requisitos ou planos.

Com a concepção e o levantamento dos dados, passamos pela fase da definição das principais funcionalidades que o sistema *Web* proposto vai oferecer e dos actores que vão operar no sistema. Para tal, recorreu-se ao diagrama de casos de uso para uma melhor definição. Com todas as funcionalidades definidas, especificou-se, então, todos os requisitos funcionais e não-funcionais para o sistema.

5.4 Engenharia de requisitos

Implementou-se um protótipo funcional para um Sistema de Coleta e Arquivação da Web Nacional (ARQUIVO WEB ANGOLANO). A análise do ARQUIVO WEB ANGOLANO foi desenvolvida com base aos problemas expostos e com vista na hipótese proposta para solucionar o problema a fim de alcançar os objetivos esperados, recorrendo às técnicas de análise e desenho de sistemas, se utilizou a linguagem de modelação UML para o desenho dos diagramas.

As tarefas e técnicas que levam a um entendimento dos requisitos é chamada engenharia de requisitos e constrói uma ponte para o projecto e para a construção, permitindo que se examine o contexto do *software* a ser desenvolvido, as necessidades específicas que o projecto deve atender, as prioridades que orientam a ordem na qual o trabalho de ser completado e as informações que terão um impacto profundo no produto final resultante.

Fornece o mecanismo apropriado para entender aquilo que o cliente deseja, por se analisar as necessidades e especificar a solução sem ambiguidades. Abrange tarefas distintas, como: concepção, levantamento, elaboração e especificação.

Para o levantamento dos requisitos, conforme abordado no capítulo 2, fez-se uma pesquisa cuidadosa sobre o tema em questão, as sistemas existentes. Além disso, a observação foi determinante para que se conseguisse outros dados, tal como a conversa formal e informal com técnicos da área e com os actores deste sector.

5.5 Requisitos Funcionais

Requisitos funcionais são as declarações dos serviços que o sistema deve fornecer, de como o sistema deve reagir a entradas específicas e de como o sistema deve se comportar em determinadas situações. Em alguns casos, os requisitos funcionais também podem explicitar o que o sistema não deve fazer.

Os requisitos funcionais de utilizadores definem recursos específicos que devem ser fornecidos pelo sistema (SOMMERVILLE, 2008).

Desta forma, foi elaborada uma lista de requisitos funcionais. A tabela abaixo apresenta uma visão total desta lista.

Tabela 4: Requisitos Funcionais

Nº	Descrição	Actor
RF1.	Registrar Utilizador (Administrador)	
RF2.	Actualizar Conta Utilizador (Administrador)	
RF3.	Desactivar Conta de Utilizador (Administrador)	
RF4.	Activar Conta de Utilizador (Administrador)	
RF5.	Listar Utilizadores	
RF6.	Registrar site (Qualquer)	
RF7.	Definir Privilégios Padrão	

RF8.	Atribuir Privilégio	Super Administrador
RF9.	Remover Privilégio	
RF10.	Remover todos os Privilégios	
RF11.	Desablitar Site (Qualquer)	
RF12.	Remover Site (Qualquer)	
RF13.	Registrar Ficheiro (Qualquer)	
RF14.	Desablitar Ficheiro (Qualquer)	
RF15.	Remover Ficheiro (Qualquer)	
RF16.	Listar site (Qualquer)	
RF17.	Listar Ficheiro (Qualquer)	
RF18.	Estatísticas de Acesso (Qualquer)	
RF19.	Estatísticas de Registo (Qualquer)	
RF20.	Estatísticas de Versões (Qualquer)	
RF21.	Registo de Conta (Pedido ao Super Administrador)	Administrador
RF22.	Actualizar Conta (Pedido ao Super Administrador)	
RF23.	Desablitar Conta	
RF24.	Registrar Site (Arquivação de seu Site)	
RF25.	Desablitar Site (seu Site)	
RF26.	Desablitar Versão (seu Site)	
RF27.	Registrar Ficheiro (Arquivação de seu Site)	
RF28.	Desablitar Ficheiro (seu Ficheiro)	
RF29.	Visualizar Sites (seu Site)	
RF30.	Pesquisar por Informação	
RF31.	Estatísticas de Acesso (seu Site)	
RF32.	Estatística de Versões (seu Site)	
RF33.	Pesquisar Por Informações	Utilizador (normal)
RF34.	Resultado de Pesquisa	
RF35.	Visualizar Resultado	
RF36.	Estatística de Versões (Registado)	
RF37.	Estatística de Acesso (Registado)	

Fonte: autor

5.6 Requisitos Não Funcionais

Os requisitos não funcionais são aqueles que não dizem respeito diretamente às funcionalidades fornecidas pelo sistema. Podem estar relacionados a propriedades de sistemas emergentes, como confiabilidade, tempo de resposta, espaço em disco, desempenho e outros atributos de qualidade do produto (PAULA FILHO, 2000). Às vezes podem dizer respeito ao sistema como um todo.

Sommerville (2008) classifica os requisitos não funcionais em:

- Requisitos de produto que especificam o comportamento do produto. Podem restringir, por exemplo, a liberdade dos projetistas a utilizar uma determinada linguagem.
- Requisitos organizacionais que são procedentes de políticas e procedimentos adotados nas organizações do cliente e do desenvolvedor. Especifica que o sistema deve ser de acordo com um processo-padrão da empresa.
- Requisitos externos que abrange tópicos advindos de fatores externos ao sistema. Dentre eles destacam-se os requisitos de interoperabilidade, os requisitos éticos e os requisitos legais que devem ser observados a fim de garantir que o sistema opera de acordo com a lei.

Foi elaborada uma lista de requisitos funcionais. A tabela abaixo apresenta uma visão total desta lista:

Tabela 5: Requisitos não funcionais

Requisito do Produto	
RNF01	Sistema deve ser estruturado em três camadas: Servidor Web, Servidor Aplicacional e Servidor Base de dados.
RNF02	Os clientes devem ser um navegador web ou Dispositivo Movel
Requisitos Organizações	
RNF03	Super Administrador, Administrador e Utilizador (Registado) devem autenticar-se com as suas credências (username e palavra-passe) antes de usar serviços oferecidos pelo sistema.
RNF04	O sistema operacional para os servidores tem de ser de base Unix (MacOS, Linux) ou Windows.
RNF05	O sistema deve ser construído usando as linguagens script PHP, JavaScript e Python.
Requisitos externos	
RNF06	O Sistema deve ser imune aos ataques de SQL Injection, PHP Injection, DoS, Session Hijacking, XSS, Broken Authentication and Session Management, Insecure Direct Object References, Cross-Site Request Forgery (CSRF)
RNF07	O sistema deve implementar o processo de assinatura e certificado digital para garantir a integridade dos utilizadores
RNF08	As senhas de utilizadores devem ser armazenadas de forma criptografada
RNF09	O sistema deve registar em uma base de dados as operações críticas realizadas no sistema para permitir auditoria
RNF10	O sistema deve registar em um log as operações realizadas no sistema para permitir auditoria
RNF11	Impedir acesso ao sistema a Utilizadores desactivados

Fonte: autor

5.7 Casos de Uso

Os casos de uso descrevem o comportamento do sistema sob várias condições à medida que o sistema responde a uma solicitação de um dos seus utilizadores. Basicamente, indicam como um utilizador final (que pode desempenhar um ou muitos de vários papéis possíveis) interage com o sistema sob um conjunto de circunstâncias específicas.

Apesar de cada instância de um caso de uso ser independente, a descrição de um caso de uso pode ser factorizada em outros, mais simples. Um caso de uso pode simplesmente incorporar o comportamento de outros casos de uso como fragmentos do seu próprio comportamento. A isto chamamos de relação de inclusão.

Um caso de uso pode também ser definido como extensão incremental de um caso de uso base. A isto chamamos de relação de extensão. Um caso de uso pode também ser especializado em um ou mais casos filhos. Esta relação é chamada de generalização.

Os objectivos da utilização dos casos de uso são:

- Definir os requisitos funcionais e operacionais do sistema, definindo os cenários de utilização do sistema.
Fornecer uma clara e não ambígua descrição de como o utilizador final e o sistema interagem um com o outro.
- Fornecer uma base para testes de validação.

Para escrever um caso de uso, primeiro têm que ser definidos o conjunto de actores envolvidos, as diferentes pessoas ou dispositivos que usam o sistema que representam os papéis que desempenham enquanto o sistema opera.

As seguintes questões ajudam muito na definição dos casos de uso:

- Que precondições devem existir antes de uma história começar?
- Que tarefas ou funções principais são realizadas pelo actor?
- Quais são as variações possíveis na interação do actor?
- Que informações de sistema o actor adquire, produz ou modifica?

A figura abaixo fornece uma ilustração de quais são os caso de uso desenvolvidos no ARQUIVO WEB ANGOLANO de forma individual, bem como a interacção entre os mesmos.

Fonte: autor

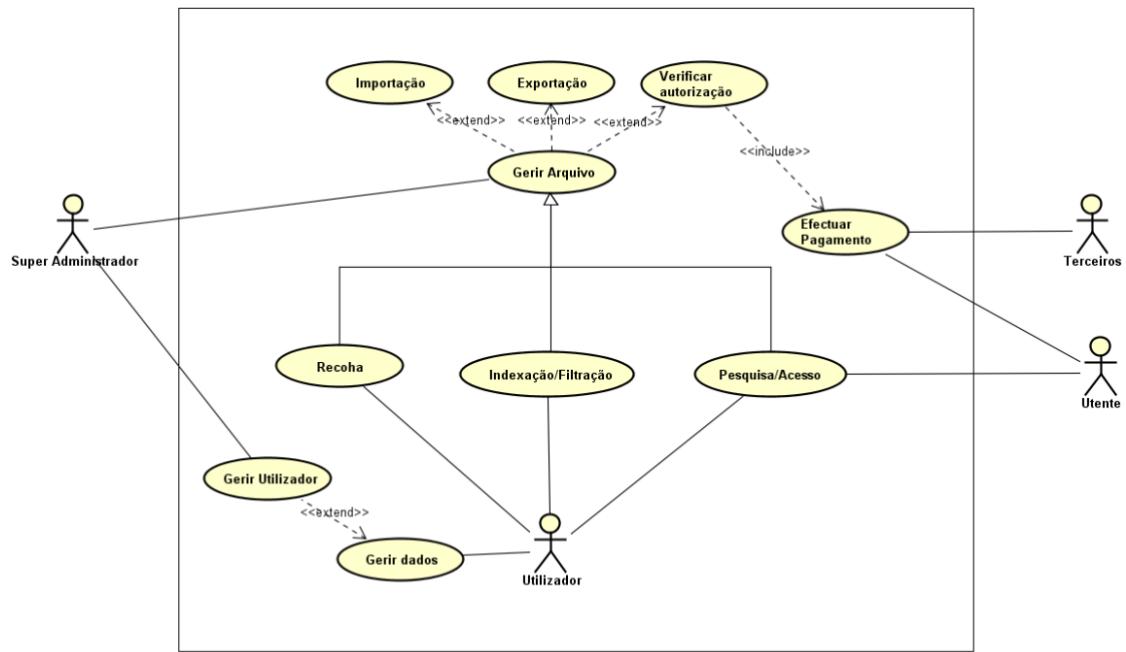
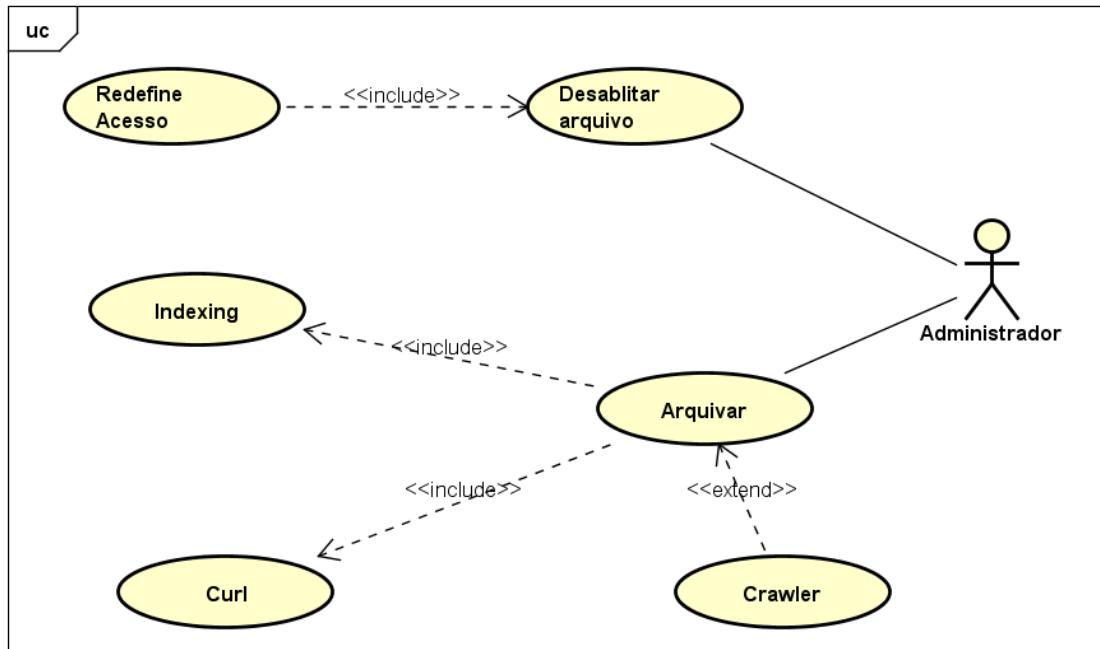


Figura 10: Diagrama de Casos de Uso do Sistema

A figura a seguir nos mostrará todo processo de gestão de arquivos segundo o nosso modelo de negocio, para facilitar a comunicação entre as partes envolvidas na manutenção e controle da aplicação.

Fonte: autor

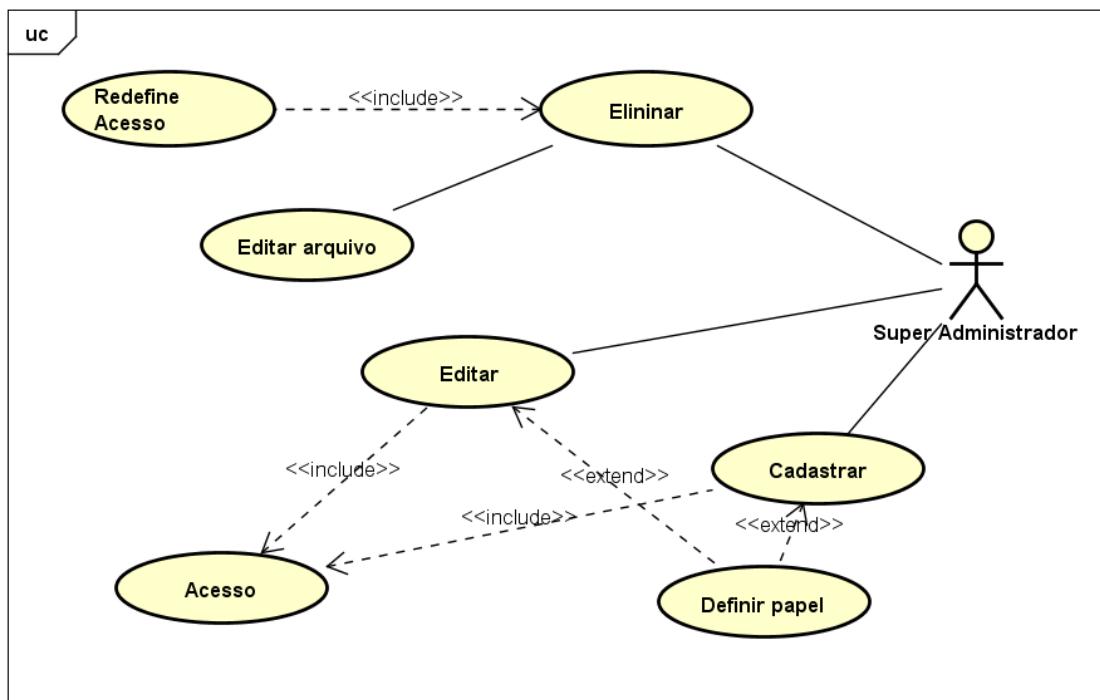


powered by Astah

Figura 11: Caso de uso gestão de arquivo

Tal como outros sistemas normais o arquivo web angolano também tem utilizadores e estes têm o seu nível de gestão desde o irão aceder ate mesmo o que irão modificar:

Fonte: autor



powered by Astah

Figura 12: Caso de uso gestão de arquivo

5.8.1 Especificação de Casos de Uso

Na busca por um melhor entendimento do diagrama de casos de uso é fundamental a existência de uma documentação acoplada.

Esta documentação descreve a interacção entre os actores e os casos de usos por meio de uma linguagem simples. Foi adoptado um formato de documentação que julgamos ser de fácil entendimento para leitores.

Nas páginas vamos mostrar as especificações de alguns casos de uso do sistema:

Tabela 6: Descrição dos casos de uso

Objectivo	Gestão de diferentes utilizadores desde papel ate acesso
Requisitos	Nome de utilizadore, senha, contacto e email
Actores	Super utilizador
Prioridades	Nenhum
Frequencia de uso	Sempre
Condição de entrada	Ser registado pelo super utilizador
Fluxo principal	nenhum
Nome do caso de uso	Gerir utilizadores

Fonte: autor

Tabela 7: Descrição do caso de uso pesquisar

Objectivo	Procurar por ficheiros /sites arquivados
Requisitos	Entrar no sistema
Actores	Super utilizadores, utilizadores e agentes
Prioridades	Nenhuma
Frequencia de uso	Nenhuma
Condição de entrada	Nenhuma
Fluxo principal	Pesquisar
Nome do caso de uso	Pesquisar

Fonte: autor

Tabela 8: Descrição do caso de uso crawler

Objectivo	Indexing um ficheiro /site
Requisitos	Nenhuma
Actores	Super administrador, administrador
Prioridades	Ficheiro/site existir
Frequencia de uso	Constante
Condição de entrada	Estar logado
Fluxo principal	Arquivaç Ao de ficheiro/site
Nome do caso de uso	Crawler

Fonte: autor

Tabela 9: Descrição do caso de uso cadastrar

Objectivo	Cadastra um utilizador
Requisitos	Nenhuma
Actores	Super administrador, administrador
Prioridades	Ficheiro/site existir
Frequencia de uso	Constante
Condição de entrada	Estar logado
Fluxo principal	Arquivaç Ao de ficheiro/site
Nome do caso de uso	Cadastrar

Fonte: autor

Tabela 10: Descrição do caso de uso Arquivar

Objectivo	Arquivar um ficheiro /site
Requisitos	Nenhuma
Actores	Super administrador, administrador
Prioridades	Ficheiro/site existir
Frequencia de uso	Constante

Condição de entrada	Estar logado
Fluxo principal	Arquivação de ficheiro/site
Nome do caso de uso	Arquivar

Fonte: autor

Actores em caso de utilização, refletem as funções dos utilizadores que interagem com o sistema, com base o nível de acesso dos recursos que cada um possui.

Tabela 11: Actores

NOME DO ACTOR	DESCRIPÇÃO DO ACTOR
Utilizador (normal)	Pessoa registada ou não no Sistema, capaz de visualizar arquivos e fazer pesquisa no arquivo.
Administrador	Pessoa registada no Sistema, arquivar sistemas, ver estatísticas
Super Administrador	Pessoa registada no sistema, responsável pela Gestão e Manutenção do Sistema.

Fonte: autor

5.9 DIAGRAMA DE SEQUÊNCIA

Os diagramas de interacção ajudam a documentar e a entender os aspectos dinâmicos do sistema de software. Mais especificamente, eles descrevem a sequência de mensagens enviadas e recebidas pelos objectos que participam em um caso de uso (Eduardo Bezerra, 2010).

Entre os diagramas de interacção destacam-se o de sequência e o de colaboração. Especificando o diagrama de sequência, ele é utilizado para modelar a lógica de um cenário de caso de uso.

5.9.1 Mensagens

A notação para uma mensagem em um diagrama de sequência é uma flecha (geralmente desenhada na horizontal) ligando uma linha de vida a outra. O objeto do qual parte a seta é aquele que está enviando a mensagem (objeto remetente). O objeto para o qual a seta aponta é aquele que está recebendo a mensagem (objeto receptor). O formato da ponta da seta indica o tipo de mensagem sendo enviada (síncrona ou assíncrona). O rótulo da mensagem é posicionado acima dessa seta.

5.9.2 Conceitos

Actores: São entidades externas que interagem com o sistema e que solicitam serviços. Normalmente, o ator primário é o responsável por enviar a mensagem inicial que inicia a interação entre os objetos.

Objetos: Representam as instâncias das classes representadas no processo.

Gate: Indica um ponto em que a mensagem pode ser transmitida para dentro ou para fora.

Fragmento: Fragmentos de interacção como: Alt (Alternativa), Opt (Opcional), Break (Parar), Loop (Repetição) e outras.

Linha de vida: As linhas de vida compõem a dimensão vertical. Uma linha de vida é composta de duas partes, a cabeça e a cauda. A cabeça é representada por um retângulo com dois compartimentos, no compartimento superior a identificação do objeto é exibida e no compartimento inferior (cuja utilização é opcional), aparecem valores para os atributos definidos na classe do objeto. A cauda corresponde a uma linha vertical tracejada.

Fonte: autor

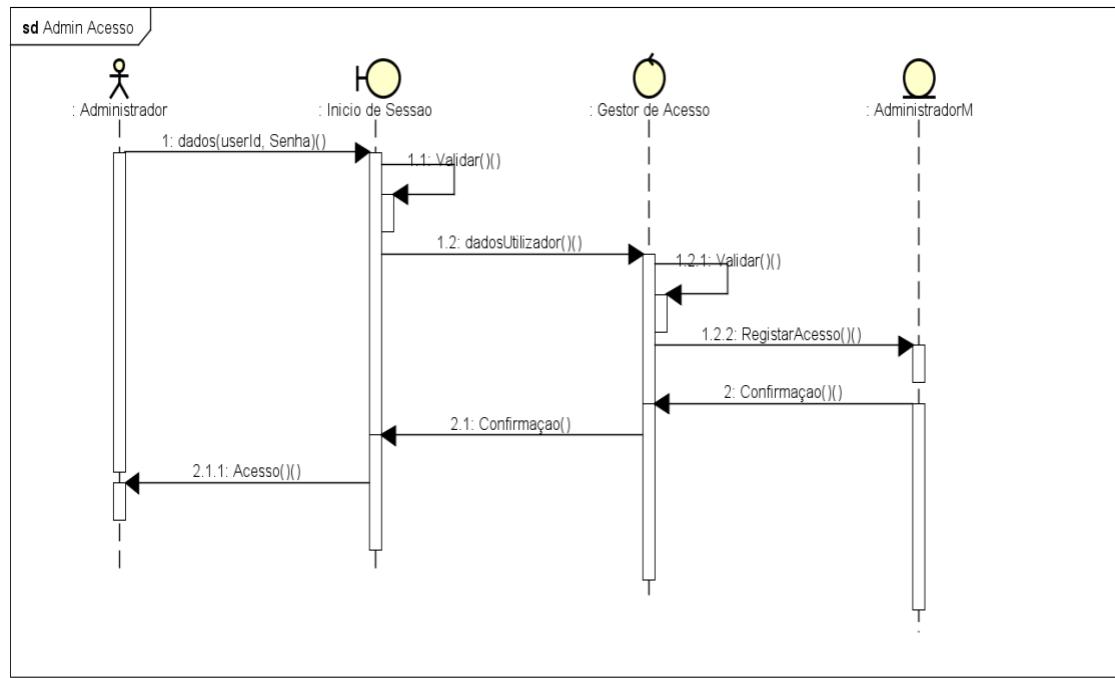


Figura 13: Diagrama de sequência do caso de uso autenticação do utilizador

powered by Astah

Fonte: autor

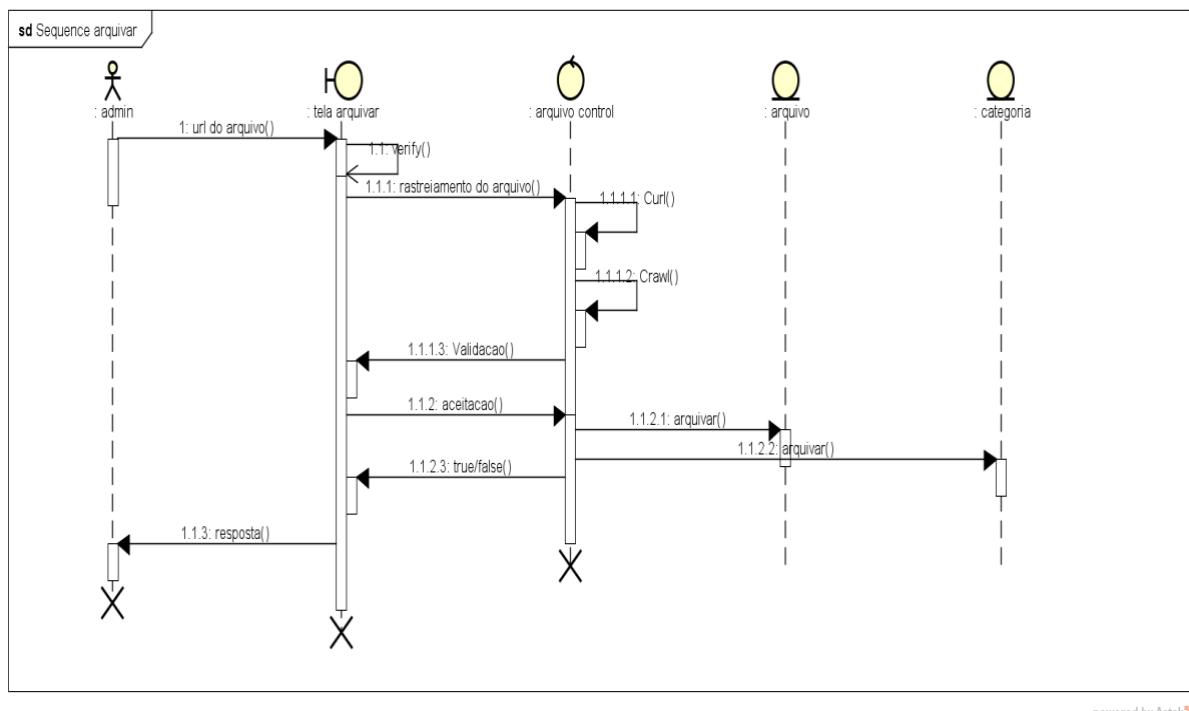


Figura 14: Diagrama de sequência do caso de uso arquivar

powered by Astah

Fonte: autor

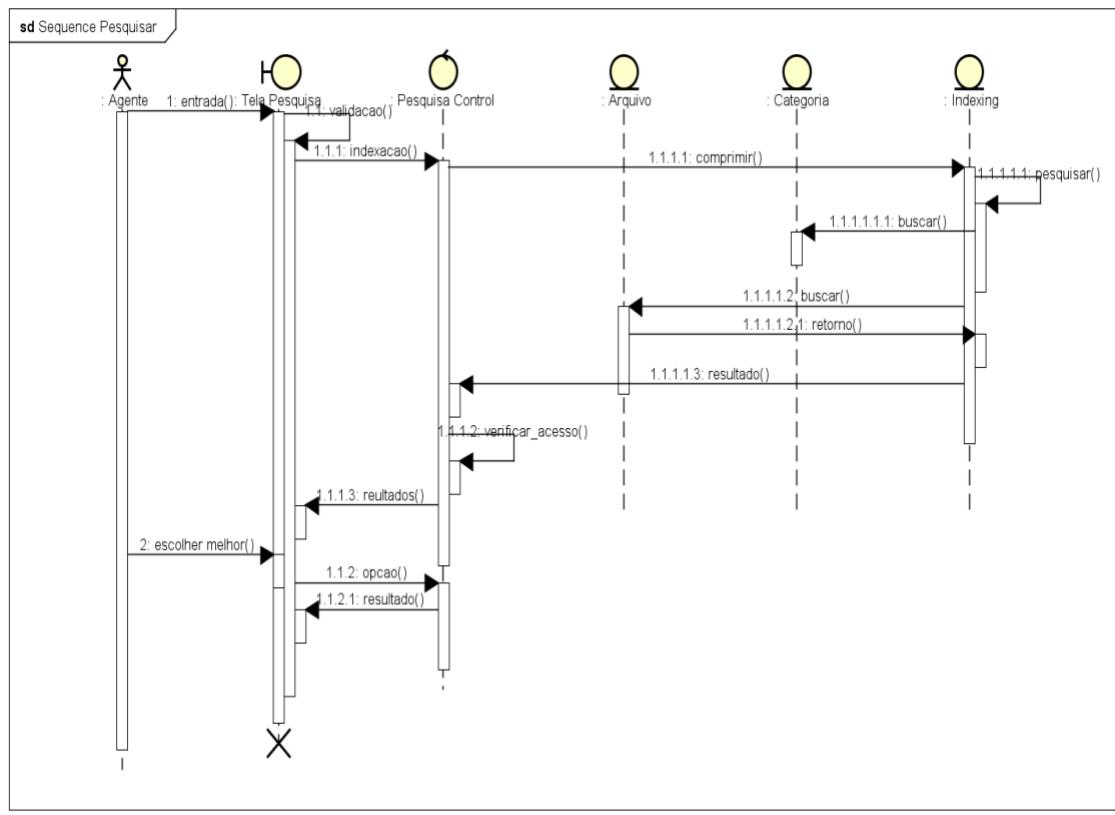


Figura 15: Diagrama de sequência do caso de uso pesquisar

powered by Astah

5.10 DIAGRAMA DE ESTADOS

Em um diagrama de estado, um objeto possui um comportamento e um estado. O estado de um objeto depende da atividade na qual ele está processando. Um diagrama de estado mostra os possíveis estados de um objeto e as transações responsáveis pelas suas mudanças de estado.

Fonte: autor

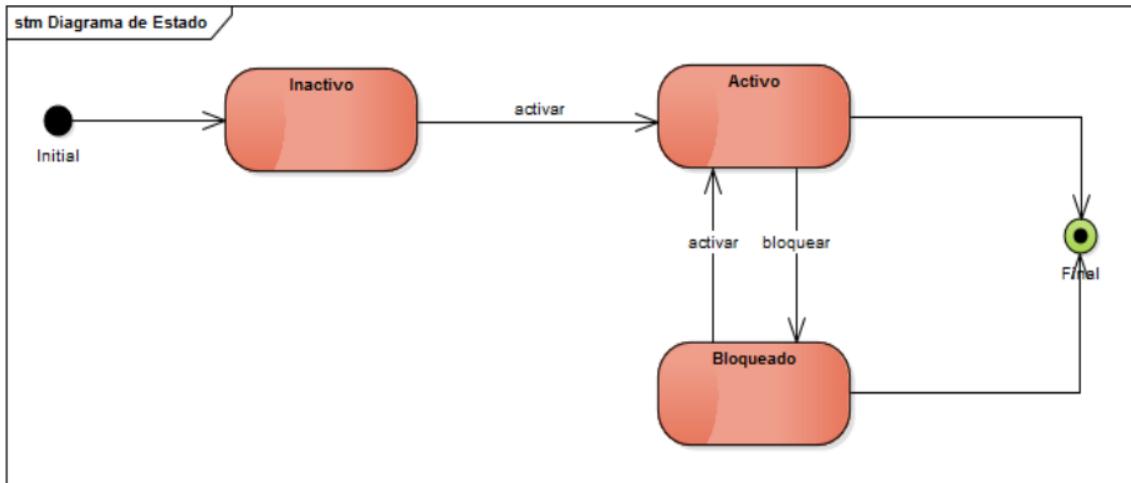


Figura 16: Diagrama de estado – autenticação

5.11 MODELO DE CLASSES DE DESENHO

No ARQUIVO WEB ANGOLANO o diagrama de classes foi utilizado para ter uma visão estática dos objectos que interagem com o sistema em um todo, fazendo menção das relações entre as classes, e os seus respectivos atributos, pondo sempre em evidência as obrigatoriedades e multiplicidades dos mesmos.

Um diagrama de classe serve para modelar os conceitos dentro do universo de discurso do problema que a aplicação se propõe resolver, bem como os conceitos internos inventados como parte da implementação da aplicação.

Um diagrama de classe é considerado como sendo estático porque não descreve o comportamento temporal (dinâmico) do sistema, que é descrito em outros tipos de diagramas. Os principais constituintes de um diagrama de classes são as classes e as suas relações: associações, generalizações e dependências.

Uma classe representa um conceito discreto dentro da aplicação a ser modelada. Tanto pode ser algo físico, um negócio (um arquivo), algo lógico (como o horário de realização de uma arquivação), ou algo comportamental (uma actividade). A classe serve como descritor para um conjunto de objectos com estrutura, comportamento e relações similares.

Um objecto é uma entidade com identidade, estado e comportamento. Os objectos são os pedaços constituintes de um sistema em execução.

As classes são os conceitos individuais a partir dos quais entende-se e descreve-se os inúmeros objectos individuais. Podemos também dizer que uma classe define um conjunto de objectos que possuem estado e exibem comportamento.

Um estado é descrito através dos atributos e associações.

Os atributos representam dados puros sem identidade, como números inteiros ou cadeia de caracteres.

As associações são usadas para relacionar objectos com identidade.

Fonte: autor

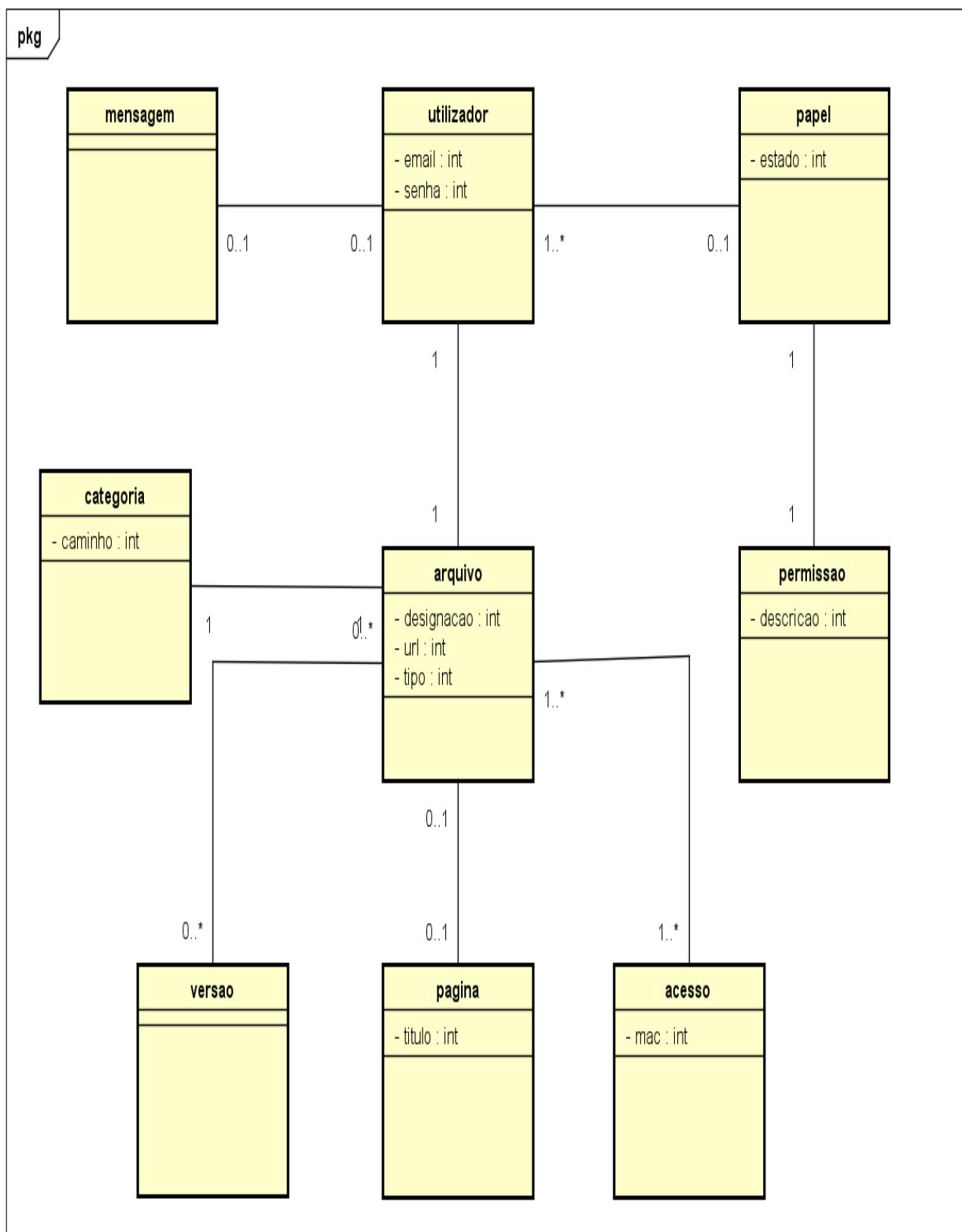


Figura 17: Diagrama de classe

5.12 MODELO DE CLASSES PERSISTENTE

As classes que precisam armazenar seu estado em uma mídia permanente são denominadas 'persistentes'. A necessidade de armazenar seu estado pode ser para registrar permanentemente as informações de classe, para fins de backup em casos de falhas do sistema ou para intercâmbio de informações.

O modelo de classes persistentes é resultante da construção do diagrama de classes anterior, onde são mostradas as classes de persistência que foram utilizadas na construção do ARQUIVO WEB ANGOLANO, conforme ilustra a figura a seguir:

Fonte: autor

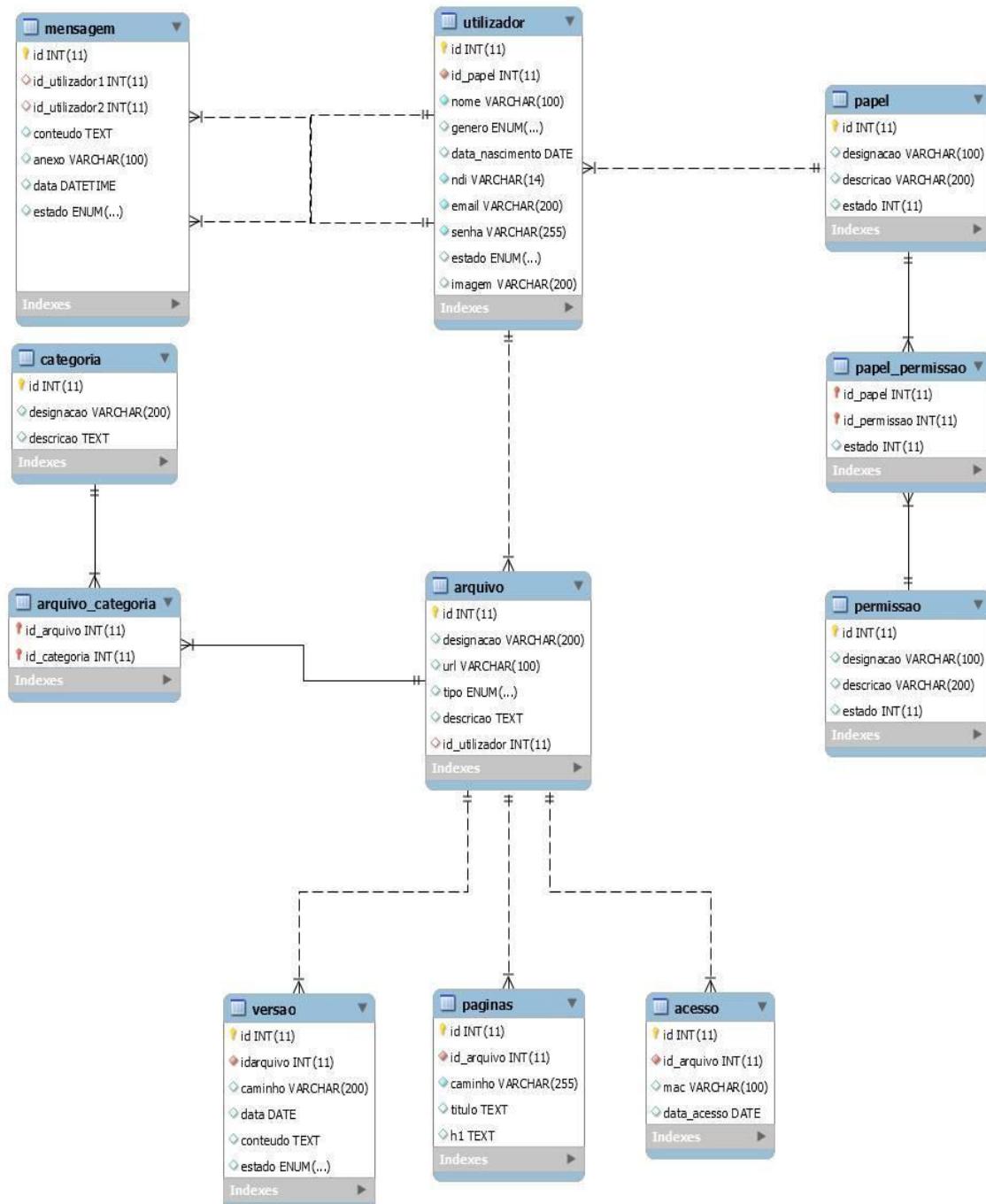
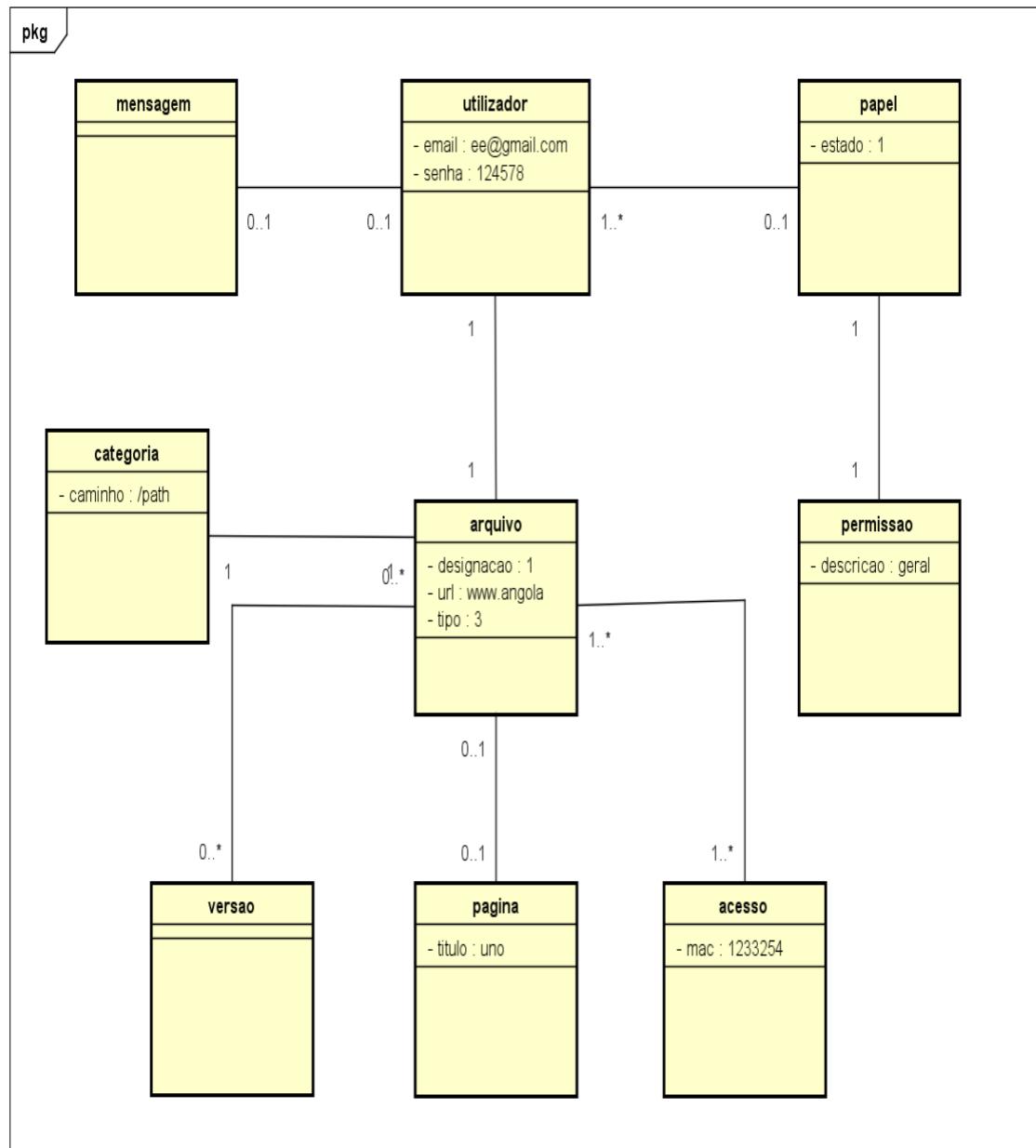


Figura 18: Modelo de classe persistente

5.13 DIAGRAMA DE OBJECTOS

Utilizamos o diagrama de objectos como uma concretização do diagrama de classes ilustrado acima, tendo em conta um extracto das classes principais do ARQUIVO WEB ANGOLANO, omitindo os nomes dos objectos e alguns conceitos dos mesmos simplificando assim a notação.

Fonte: autor



powered by Astah

Figura 19: Diagrama de objeto

5.14 Arquitectura

O Arquivo Web Angolano apresenta uma arquitectura cliente-servidor, como ilustra a Figura.. O servidor está instalado no arquivo, o cliente é o computador do utilizador.

Fonte: autor

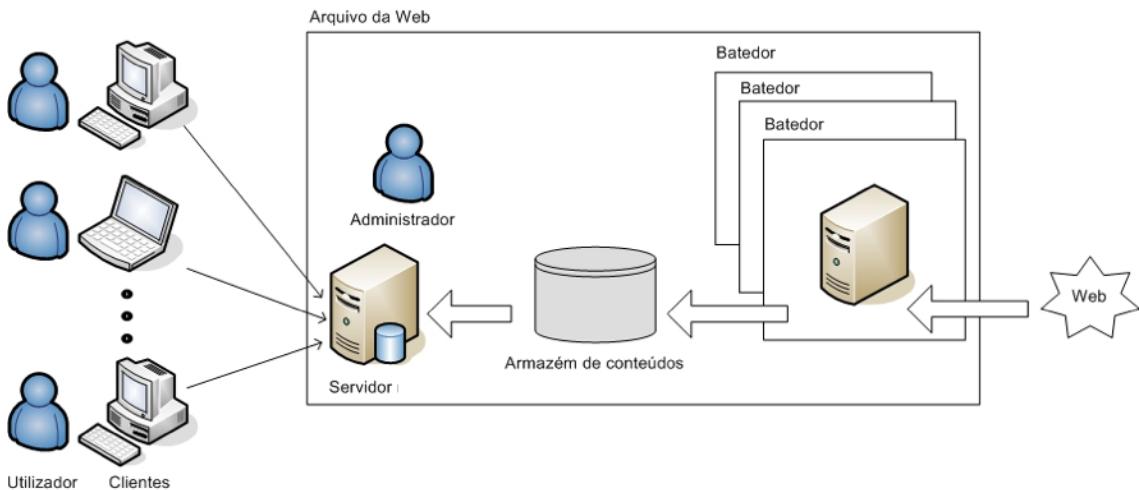


Figura 20: Arquitectura

O arquivo um batedor que é responsável por recolher os conteúdos da Web. Os conteúdos recolhidos são armazenados no armazém de conteúdos sob a forma de ficheiros. São este ficheiros que são visualizados pelos computadores dos utilizadores.

As tecnologias usadas para a concretização do AWA são exclusivamente de código-aberto porque dão maior garantia de preservação do sistema e consequentemente dos dados arquivados a longo prazo, em comparação com soluções de código-fechado que são normalmente dependentes das empresas proprietárias. O facto de o código ser aberto permite também que as ferramentas sejam alteradas para responder aos requisitos específicos do AWP.

5.15 PROTÓTIPOS DE BAIXA FIDELIDADE

Com o intuito de se definir e experimentar o projecto de forma rápida e económica, foi realizada a prototipagem de baixa fidelidade, concebida na fase inicial do projecto, com o objectivo de definir o projecto e de forma superficial definir a interacção com o utilizador.

5.15.1 Página Inicial

Esboço da página inicial do sistema, ou seja, logo que aceder a aplicação, essa tela representa a sua primeira interacção com o sistema, onde o utilizador pode efectuar a pesquisa de um arquivo dentro sistema.

Fonte: autor

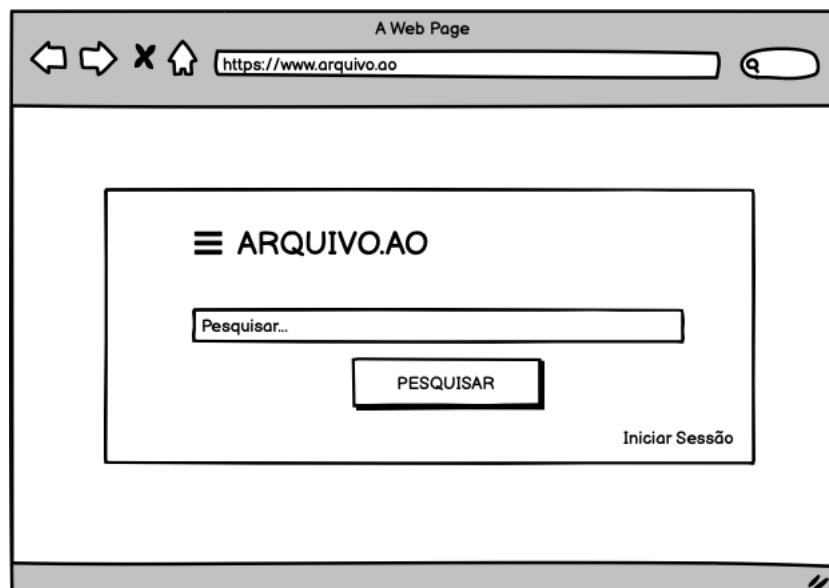


Figura 21: Protótipo da página inicial

5.15.2 Página Resultado da pesquisa

Esboço da página onde o utilizador do sistema, poderá ver os resultados da pesquisa de um site ou uma palavra.

Fonte: autor

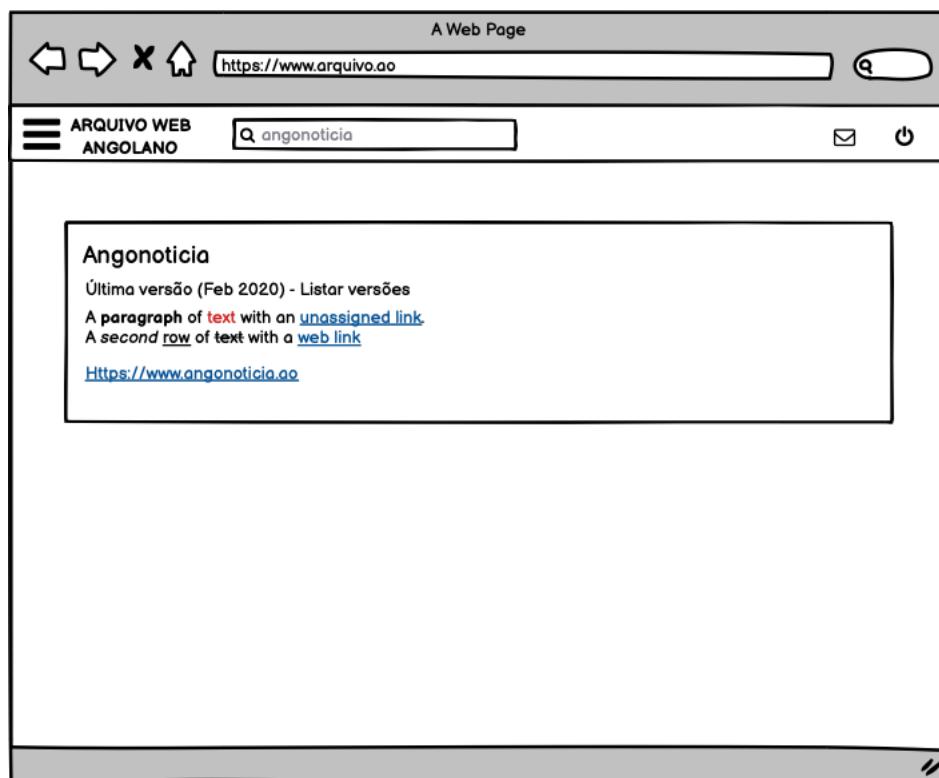


Figura 22: Protótipo da página resultado da pesquisa

5.15.3 Pagina de Resultado da pesquisa final

Esboço da página onde o utilizador do sistema, poderá ver os resultados da final da pesquisa de um site e as suas versões de arquivo disponíveis.

Fonte: autor

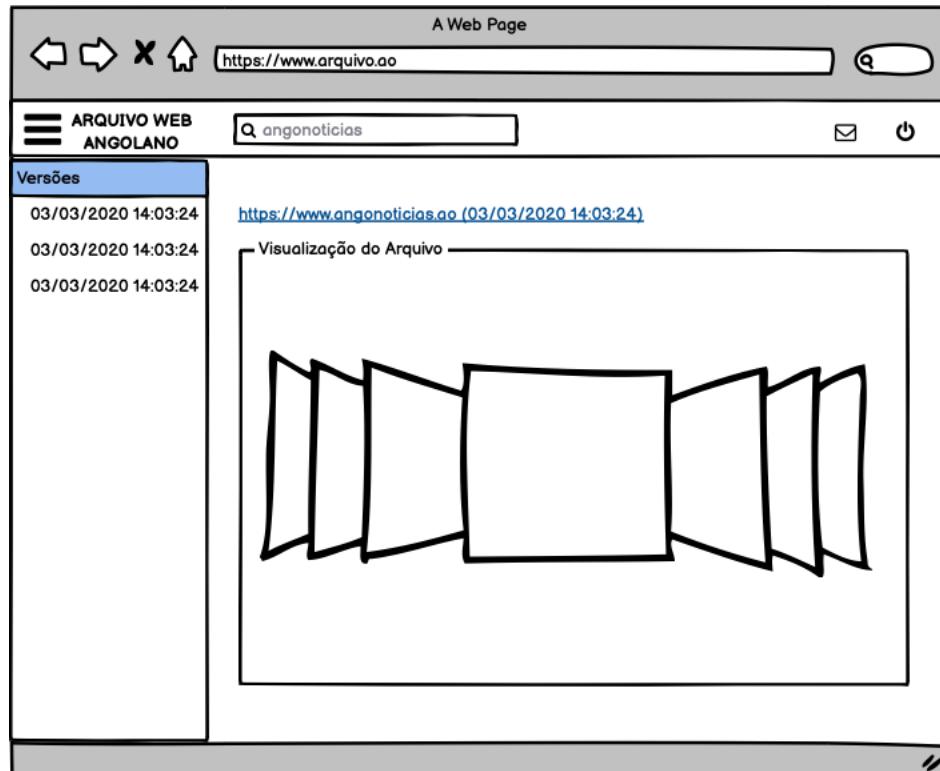


Figura 23: Protótipo do resultado de uma pesquisa

5.15.4 Página de Autenticação - Login

Esboço da página onde o utilizador com perfil de administrador do sistema, poderá fazer a autenticação para entrar no backoffice do arquivo.

Fonte: autor

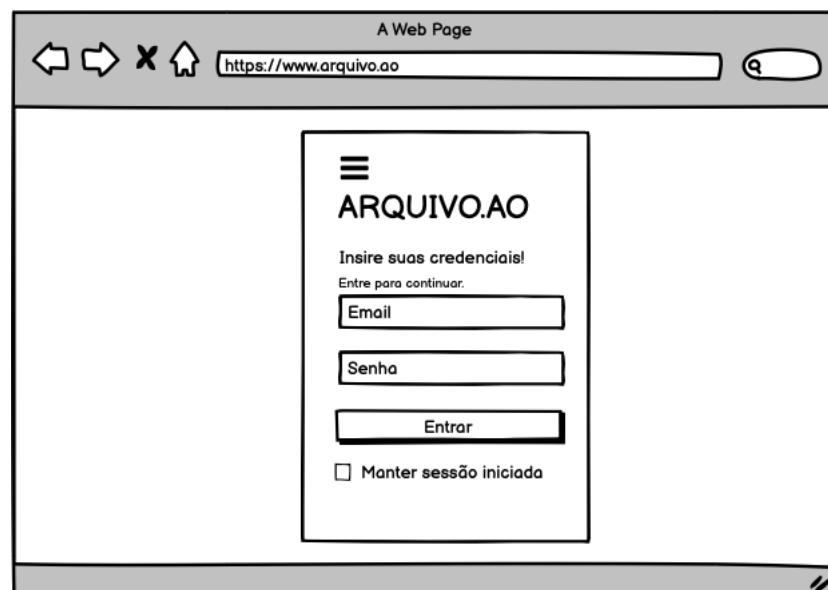


Figura 24: Protótipo da tela de login

5.15.5 Tela de Inical da Administração do Sistemas

Esboço da página inicial do administrador do sistema, onde poderá encontrar os menus de: arquivo, mensagem, configuração, documentos e um dashboard com o resumo das principais atividades do sistema.

Fonte: autor



Figura 25: Protótipo da tela inicial do BackOffice

5.15.6 Tela da página de Arquivamento

Esboço da página de arquivamento, ou seja, na área administrativa do sistema o administrador poderá escolher a opção arquivo para arquivar um site ou ficheiro.

Fonte: autor

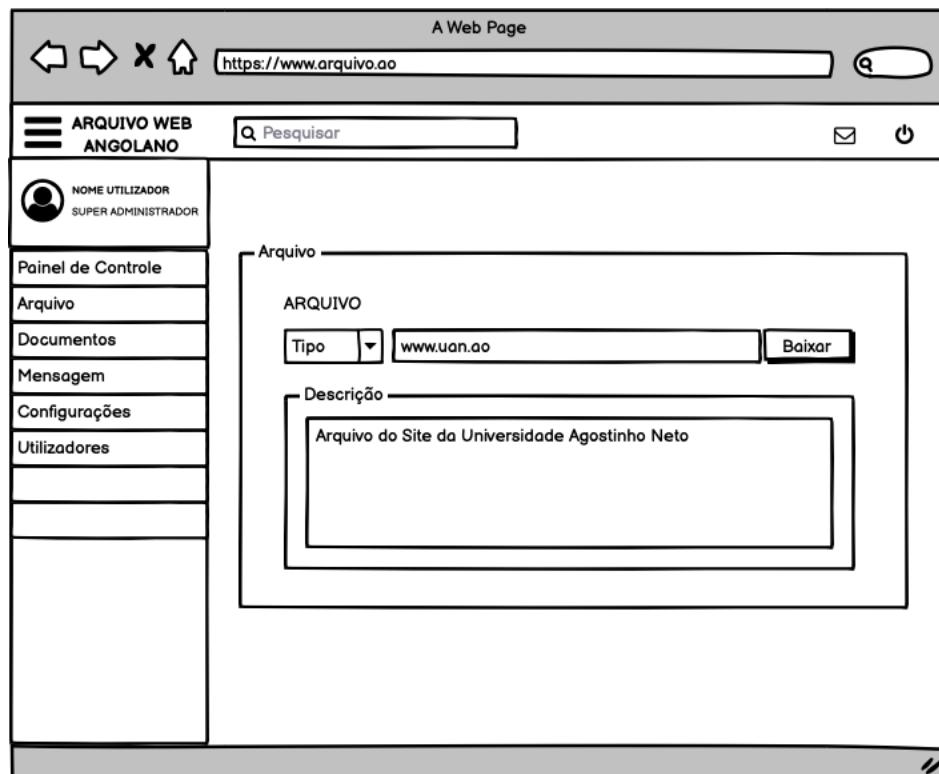


Figura 26: Tela da página de arquivamento

5.15.7 Erro no domínio

Esboço da tela de erro ao fazer um arquivo: nota-se que se o domínio for diferente de um domínio angolano não se pode fazer o arquivo.

Fonte: autor

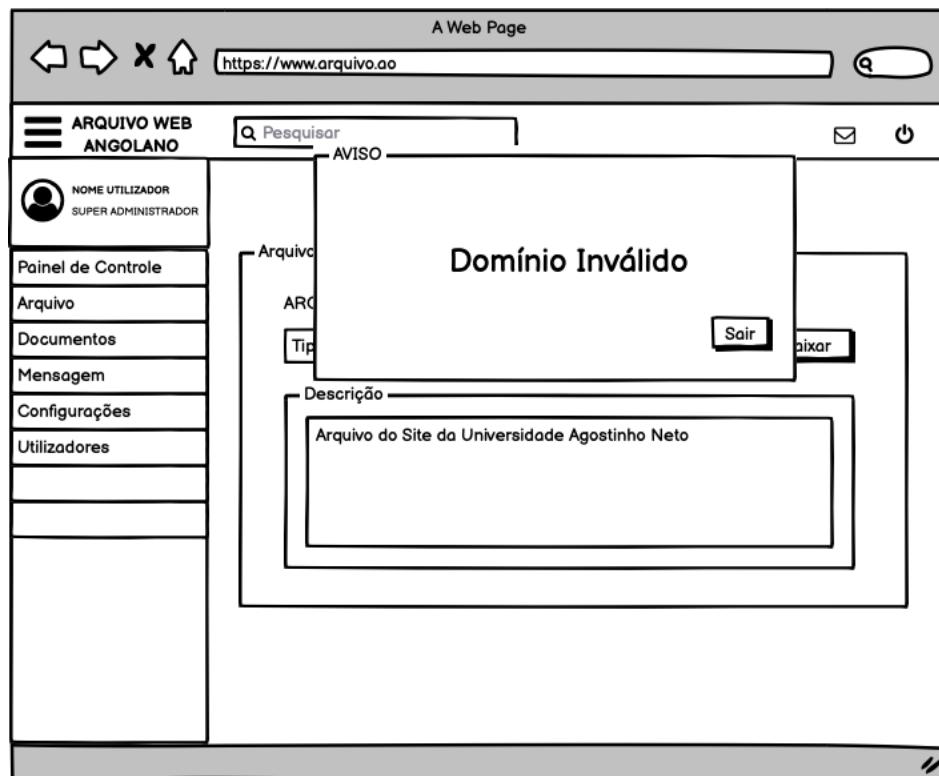


Figura 27: Tela de erro ao arquivar

5.15.8 Página de Estatística

Esboço da tela da página de estatísticas onde se tem a visão geral de dados estatísticos do sistema e ainda poderá se exportar as estatísticas por meio de um relatório.

Fonte: autor

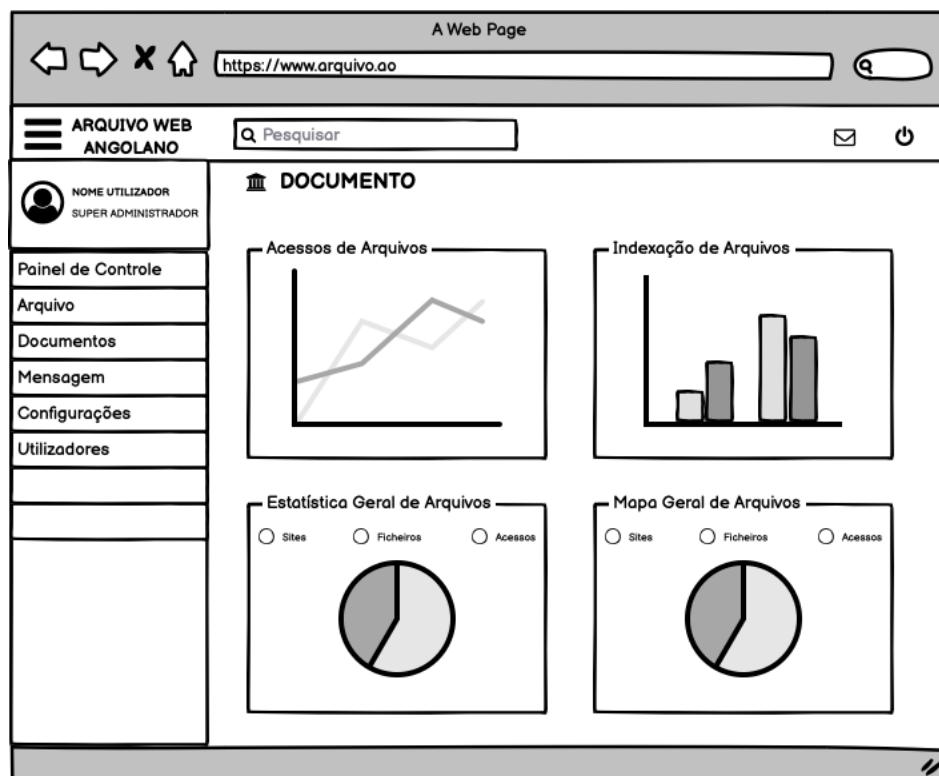


Figura 28: Tela da página de estatística

5.15.9 Gerar Relatório

Esboço da tela da página de estatísticas onde se tem a visão geral de dados estatísticos do sistema e ainda poderá se exportar as estatísticas por meio de um relatório.

Fonte: autor

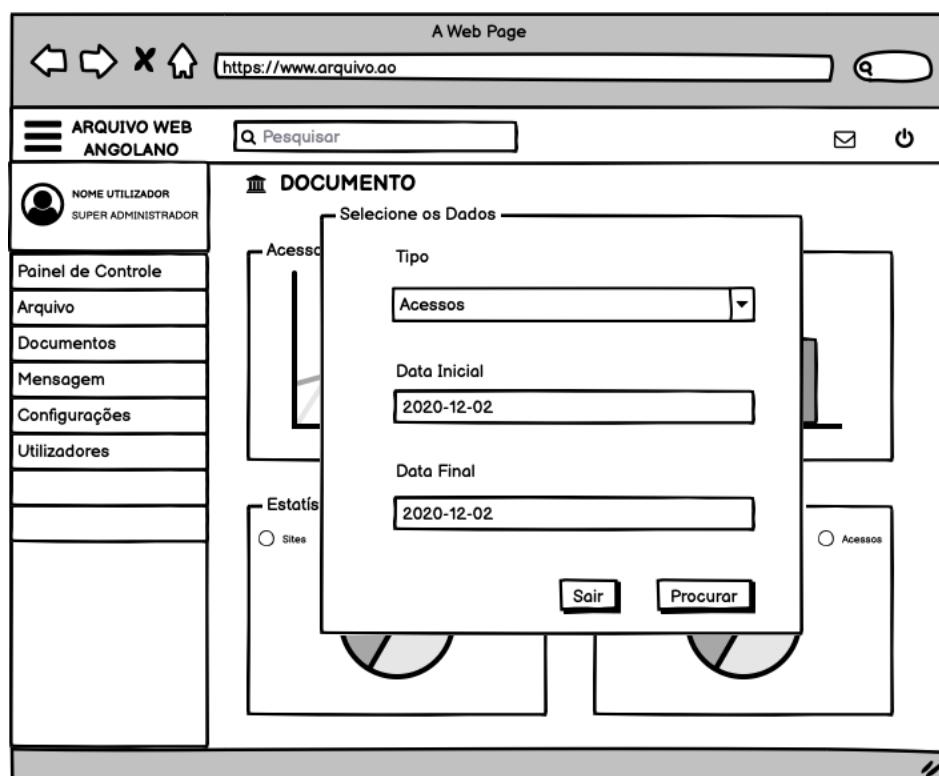


Figura 29: Tela da opção de exportar relatório

Fonte: autor

The screenshot shows a web browser window titled 'A Web Page' with the URL <https://www.arquivo.ao>. The page is titled 'ARQUIVO WEB ANGOLANO' and features a title 'RELATÓRIO DE 2020 À 2020'. Below the title is a table with the following data:

Name (job title)	Age	Nickname	Employee
Giacomo Guilizzoni Founder & CEO	40	Peldi	<input checked="" type="radio"/>
Marco Botton Tuttofare	38		<input checked="" type="checkbox"/>
Mariah MacLachlan Better Half	41	Patata	<input type="checkbox"/>
Valerie Liberty Head Chef	:)	Val	<input checked="" type="checkbox"/>

Figura 30: Tela com relatório

5.15.10 Página de Mensagem

Esboço da tela da página de mensagens, onde os administradores poderão ter uma interação colaborativa.

Fonte: autor

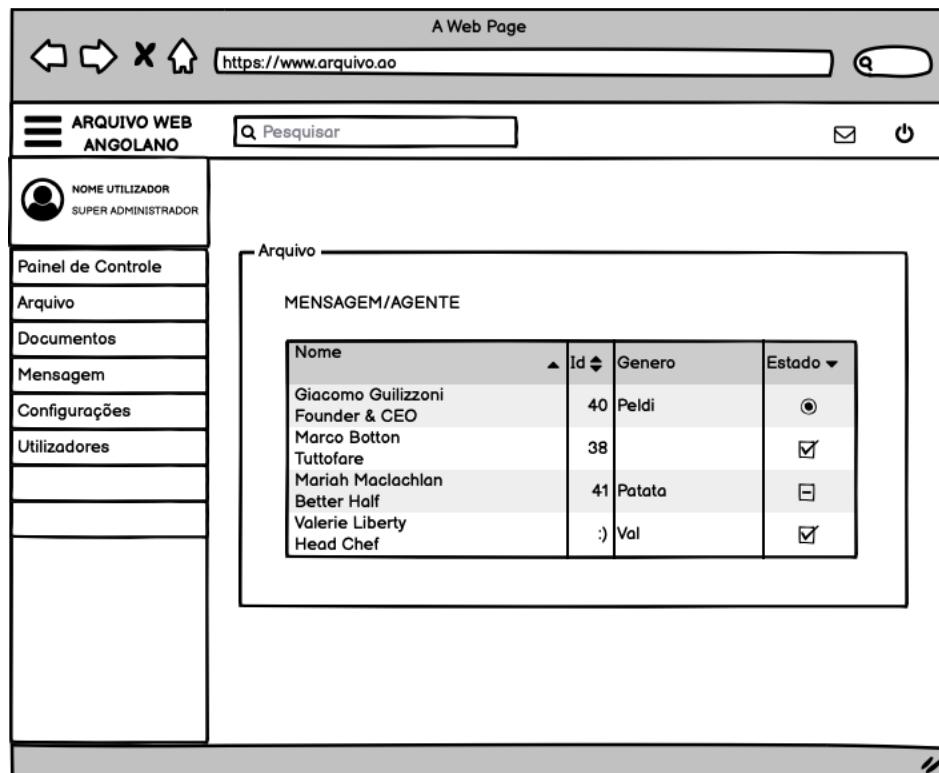


Figura 31: Tela da página de mensagem

5.15.11 Página dos Utilizadores

Esboço da tela da página de utilizadores do sistemas, onde os administradores poderão ver quais os utilizadores cadastrados no sistema e quais as suas permissões de acesso.

Fonte: autor

A Web Page

https://www.arquivo.ao

ARQUIVO WEB
ANGOLANO

NOME UTILIZADOR
SUPER ADMINISTRADOR

Pesquisar

Arquivo

UTILIZADOR/NOVO

Nome

Data Nascimento

Não Identificação

Ex: 00111112LA0111

Tipo

Gênero

Imagen

Carregar Imagem

Email

Email

Figura 32: Tela de cadastro de usuário

Fonte: autor

The screenshot shows a web browser window titled 'A Web Page' with the URL <https://www.arquivo.ao>. The page is titled 'ARQUIVO WEB ANGOLANO'. On the left, there is a vertical sidebar with a user icon and the text 'NOME UTILIZADOR SUPER ADMINISTRADOR'. Below this are menu items: 'Painel de Controle', 'Arquivo', 'Documentos', 'Mensagem', 'Configurações', and 'Utilizadores'. The main content area is titled 'Arquivo' and contains a section titled 'MENSAGEM/AGENTE'. It features a table with the following data:

Nome	Id	Genero	Estado
Giacomo Guilizzoni Founder & CEO	40	Peldi	<input type="radio"/>
Marco Botton Tuttofare	38		<input checked="" type="checkbox"/>
Mariah MacLachlan Better Half	41	Patata	<input type="checkbox"/>
Valerie Liberty Head Chef	:)	Val	<input checked="" type="checkbox"/>

Figura 33: Tela com a listagem dos utilizadores do sistema

6 RESULTADOS E DISCUSSÃO

No presente capítulo apresentam-se os principais resultados da implementação da proposta em ordem de precedência. São espelhados neste ponto do capítulo detalhes relativos à aplicação desenvolvida como solução do problema proposto.

6.8 Acesso a Aplicação

O controlo de acesso faz parte dos elementos que formam as políticas de segurança e procedimentos utilizados para evitar o acesso indevido a pessoas não autorizadas.

Mas para o nosso caso existem três grupos de utilizadores:

- Super Administrador;
- Administrador;
- Utilizador (pessoa comum que visita o nosso Sistema)

Logo que é acedida a aplicação é redirecionado a uma tela de pesquisa, conforme apresentado a seguir.

Fonte: autor

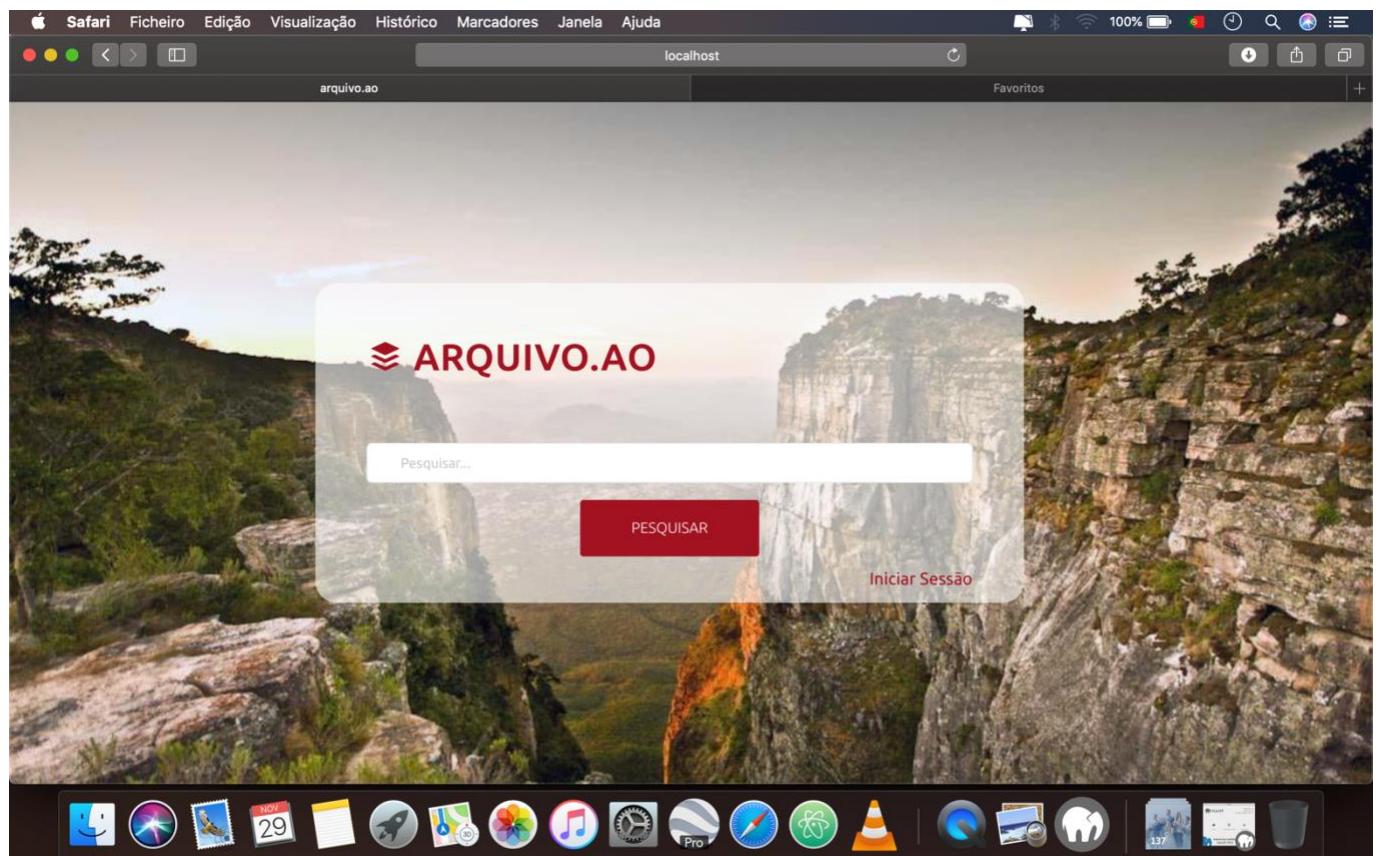


Figura 34: Tela no motor de busca do Arquivo Web Angolano

Após uma pesquisa feita o utilizador (agente) será redirecionado a uma tela de possíveis resultados da pesquisa (varios itens), esperando assim a escolha do utilizador.

Fonte: autor

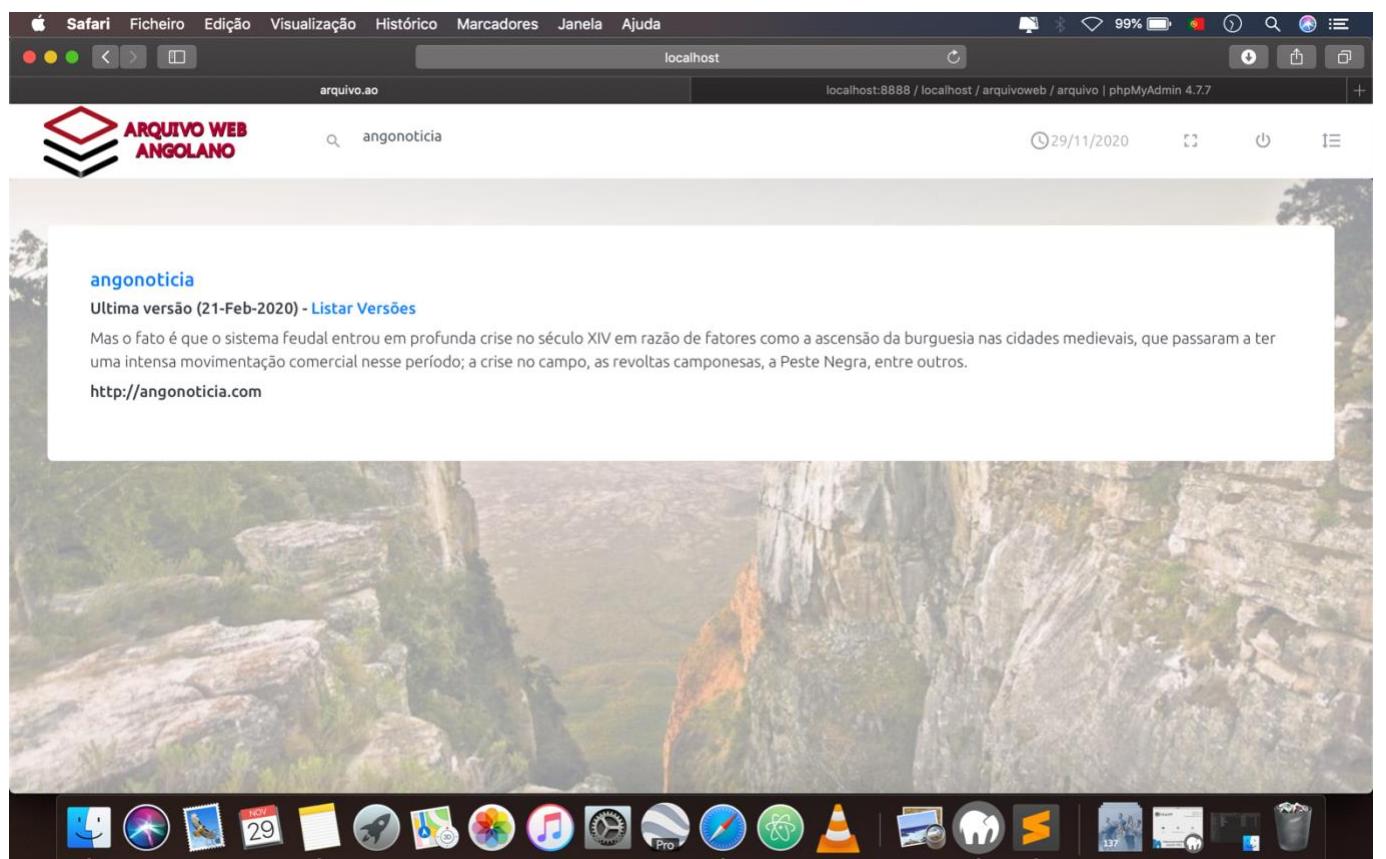


Figura 35: Resultado de uma pesquisa no AWA

Fonte: autor

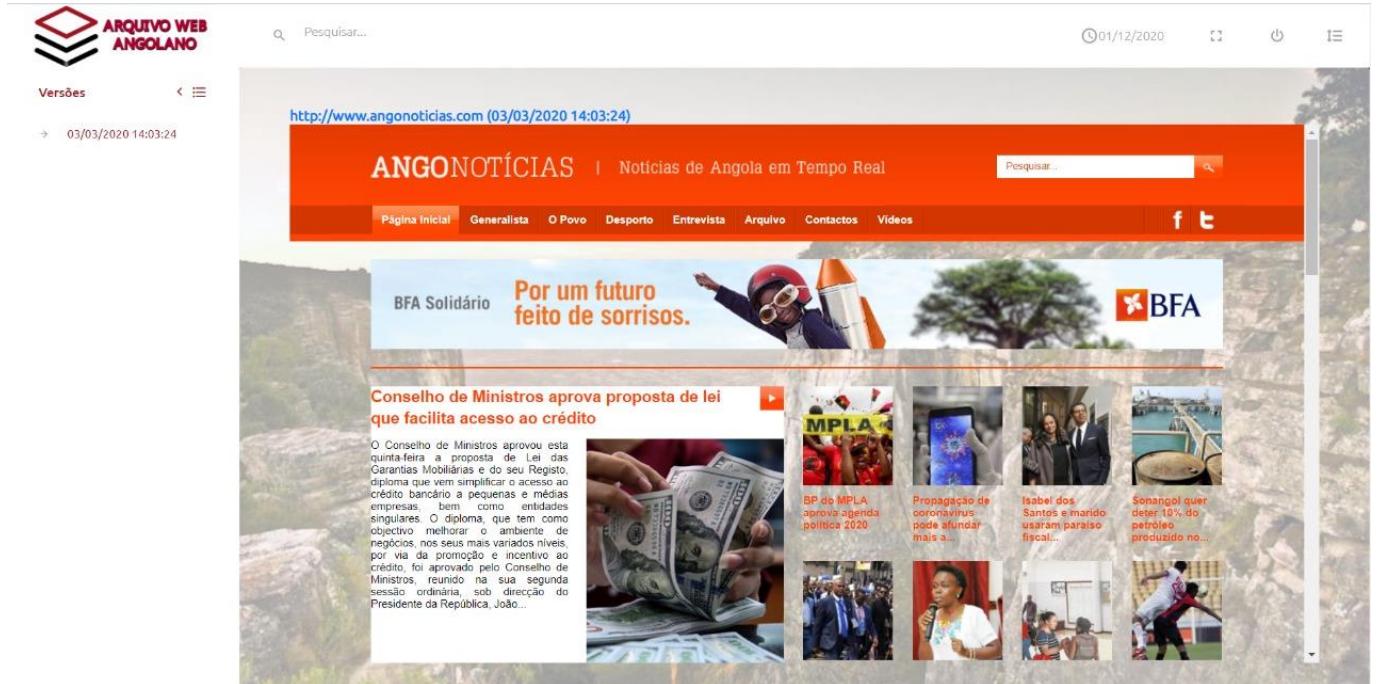


Figura 36: Visualização de uma pesquisa no AWA

6.2 Administrador do Sistema

O utilizador com mais privilégios no sistema, acede muito mais recursos que os outros, pelo papel que possui, lhe permite fazer a configuração do sistema e gerir os utilizadores, atribuindo-lhes papéis que os possibilitará aceder aos seus recursos e para que o sistema tenha o mínimo de condições para arrancar em pleno.

Fonte: autor

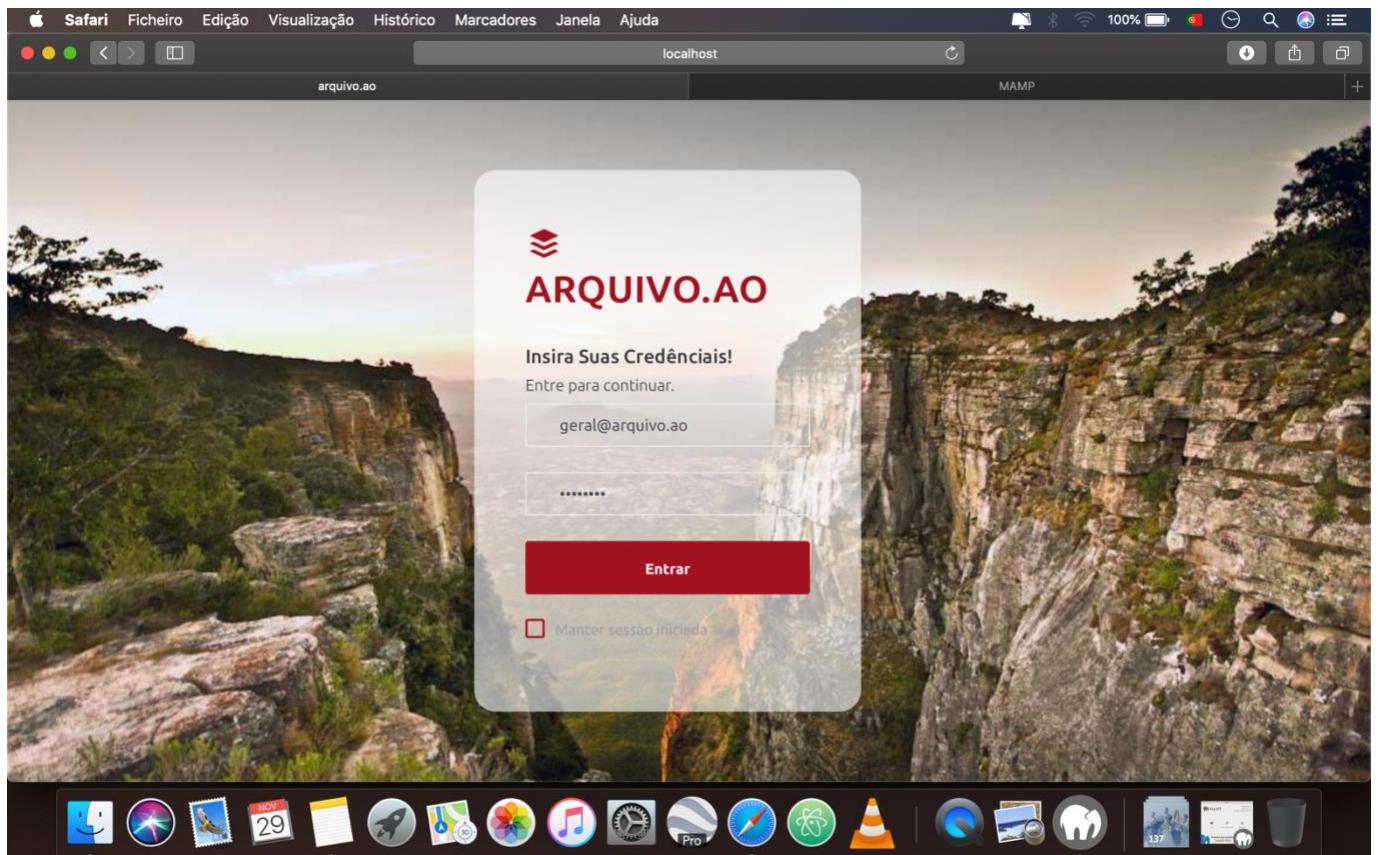


Figura 37: Tela de login para o backoffice do AWA

Fonte: autor

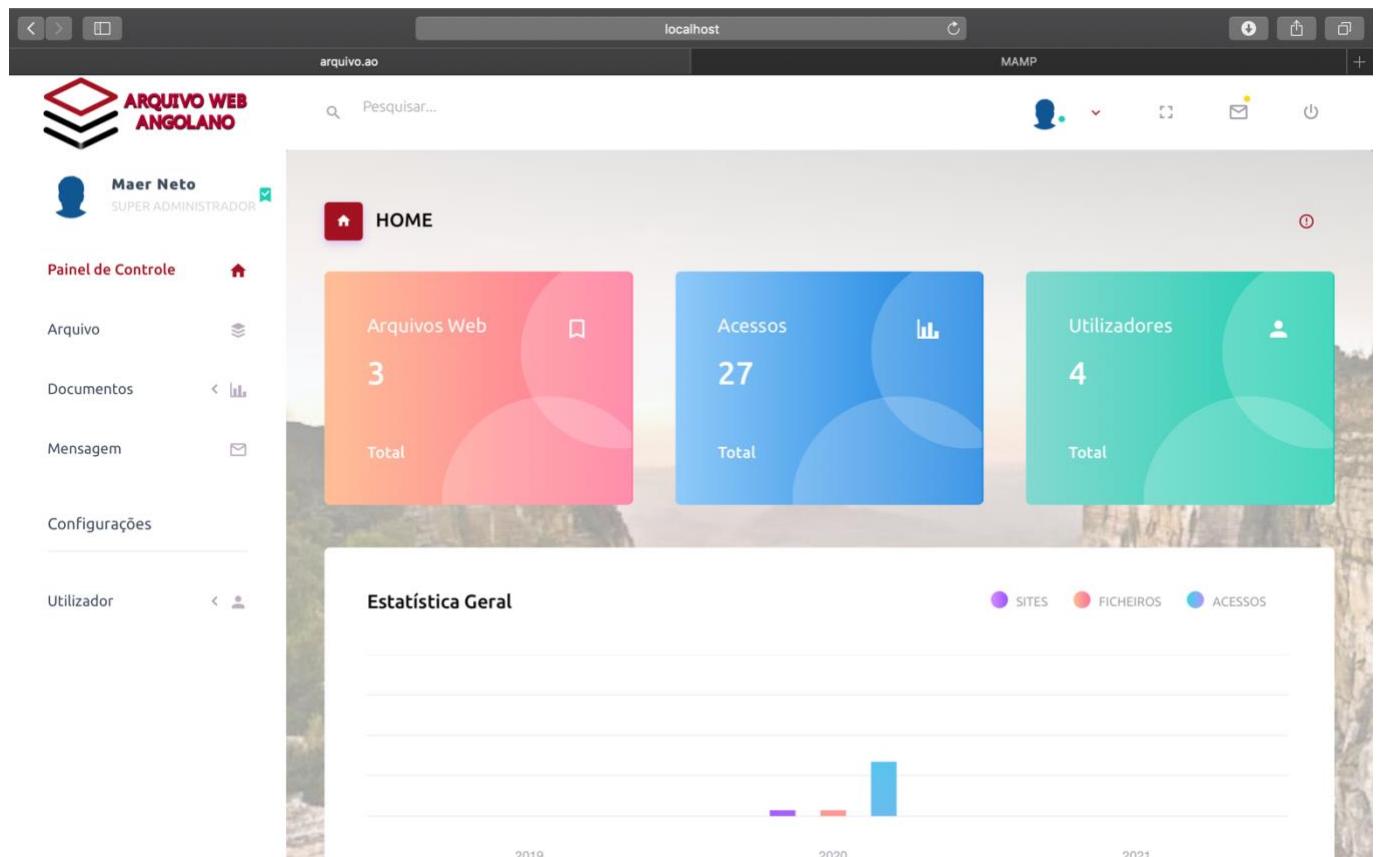


Figura 38: Painel de dashboard do AWA

Fonte: autor

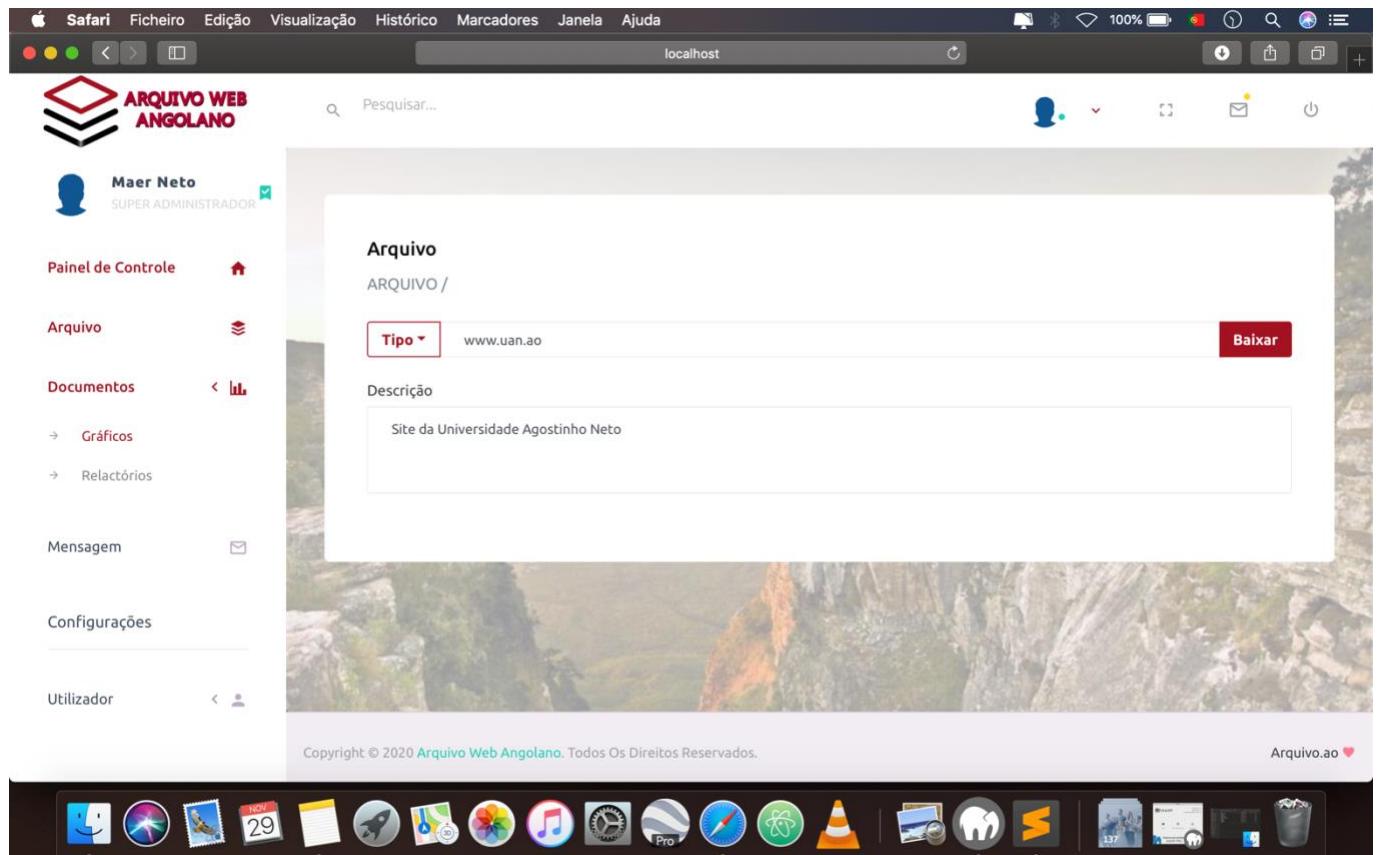


Figura 39: Tela de arquivamento do AWA

Fonte: autor

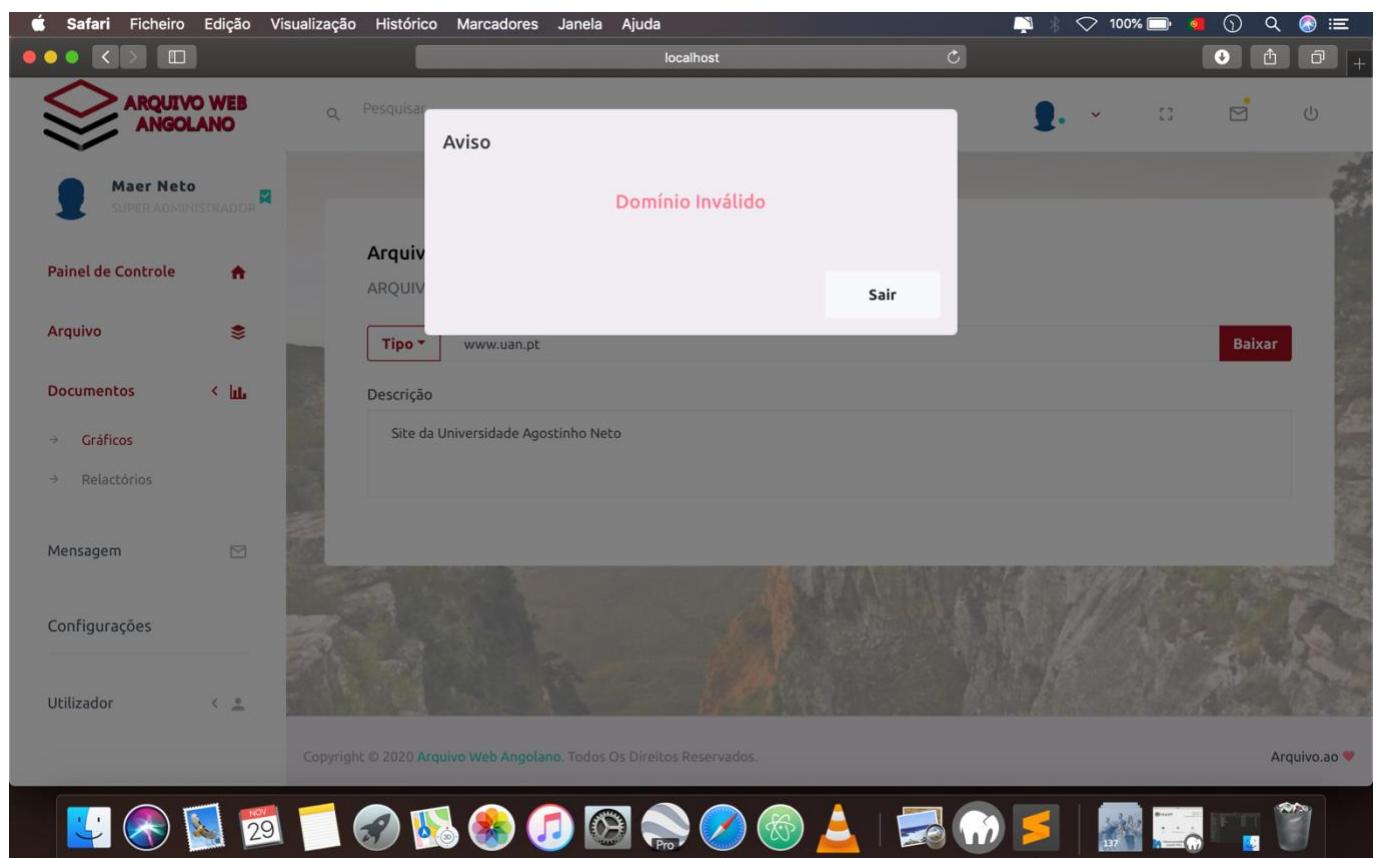


Figura 40: Erro de domínio

Fonte: autor

The screenshot shows a web browser window for 'arquivo.ao' on 'localhost'. The title bar indicates the URL is 'localhost:8888/arquivo_ao/documento/relatorio'. The left sidebar has a dark theme with white text and icons. It includes links for 'Painel de Controle', 'Arquivo', 'Documentos', 'Mensagem', 'Configurações', and 'Utilizador'. The 'Utilizador' link is currently selected, showing a blue icon. The main content area is titled 'Arquivo' and 'UTILIZADOR / NOVA'. It contains several input fields: 'Nome' (Name), 'Data Nascimento' (Birth Date), 'Nº de Identificação' (Identification Number) with a placeholder 'Ex: 11111111MO035', 'Genero' (Gender) with a dropdown menu showing 'Masculino', 'Imagen' (Image) with a placeholder 'Carregar Imagem' and a red button 'Carregar Imagem' to its right, and 'Email' (Email). The overall interface is clean and modern.

Figura 41: Registo dos utilizadores no AWA

Fonte: autor

The screenshot shows a web browser window with the URL `localhost:8888/arquivo_ao/documento/relatorio`. The page title is "Arquivo". The main content area displays a table titled "Arquivo" under the heading "MENSAGEM / AGENTE". The table has columns: #, Agente, Identificação, Gênero, Estado, and Opcões. There are three rows of data:

#	Agente	Identificação	Gênero	Estado	Opcões
1	Hilquias Chitazo	11111111111111	MASCULINO	ATIVADO	✉
2	estima sauma	1234567890LA12	MASCULINO	ATIVADO	✉
3	Maer Neto	1898717817LA91	MASCULINO	ATIVADO	✉

The left sidebar includes links for "Painel de Controle", "Arquivo", "Documentos", "Mensagem", "Configurações", and "Utilizador". The top navigation bar shows the logo "ARQUIVO WEB ANGOLANO", a search bar, and user profile icons.

Figura 42: Listagem dos utilizadores no AWA

Fonte: autor

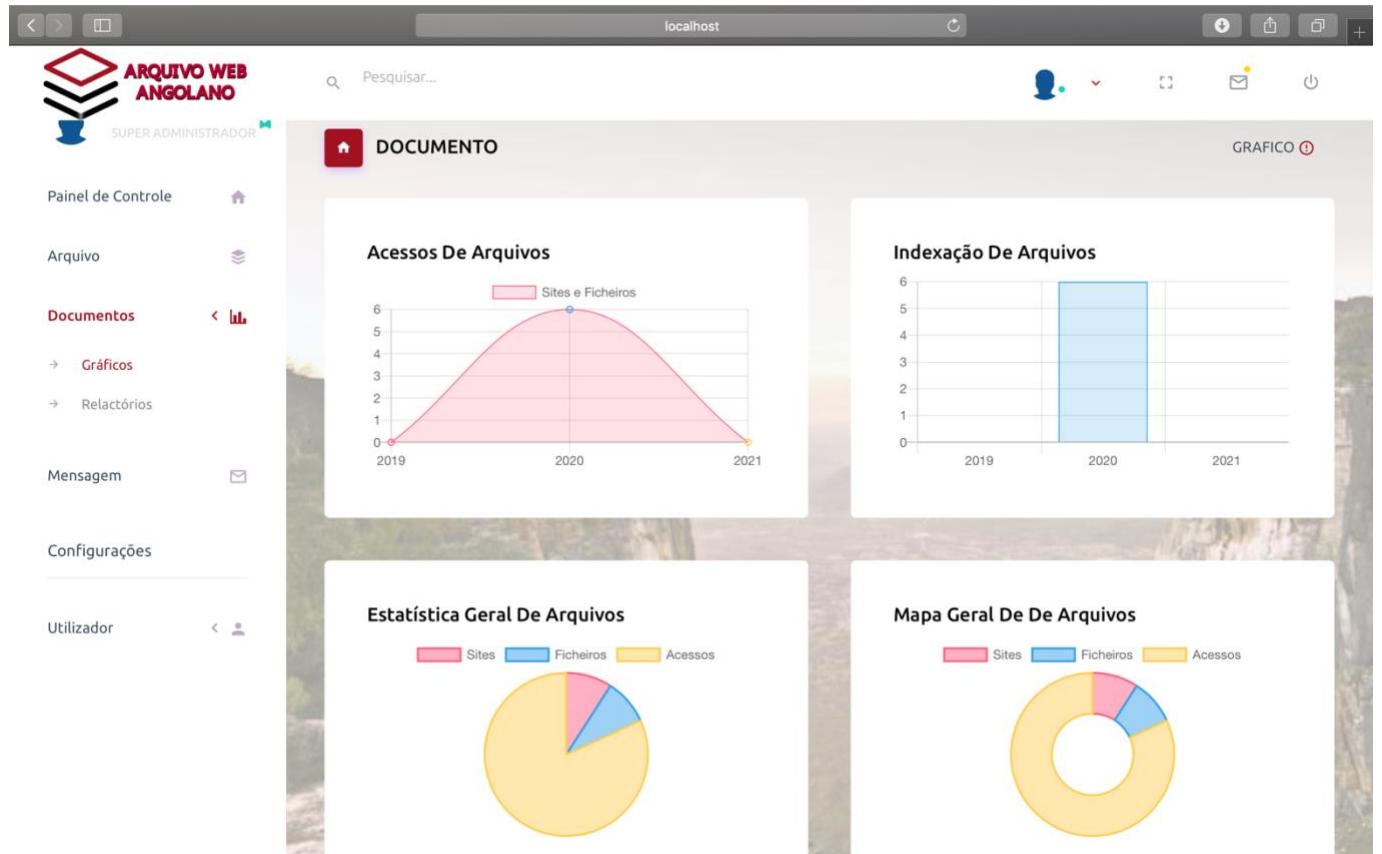


Figura 43: Tela de estatísticas do AWA

Fonte: autor

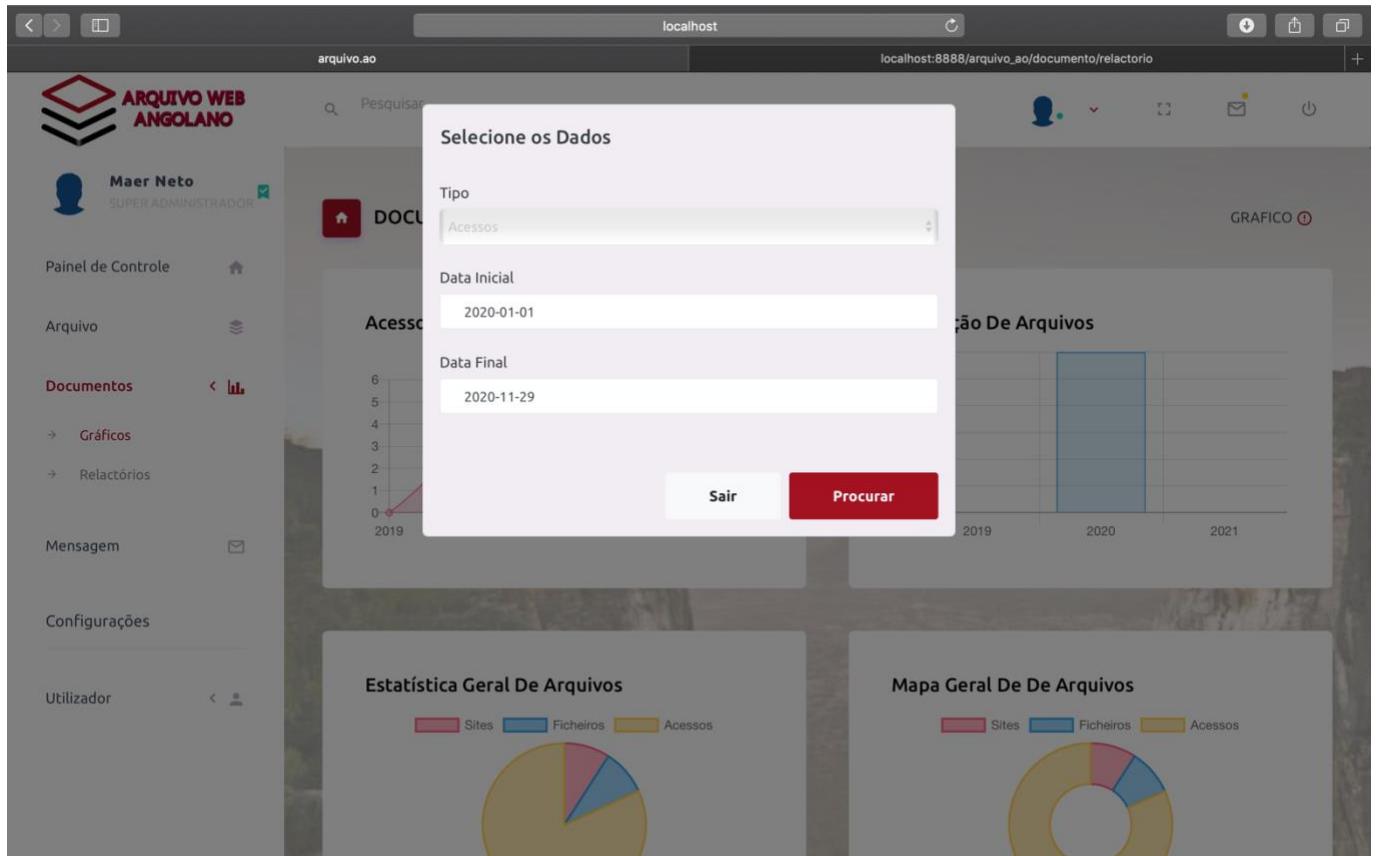


Figura 44: Imprimir relatório no AWA

Fonte: autor

#	Designacao	URL	Tipo	Host	Data
1	ANGONOTICIA	http://angonoticia.com	SITE	0A-00-27-00-00-17	01/03/2020 12:03:36
2	ANGONOTICIA	http://angonoticia.com	SITE	0A-00-27-00-00-17	01/03/2020 12:03:41
3	PLATINALINE	http://platinoline.com	SITE	0A-00-27-00-00-17	29/02/2020 17:02:27
4	PLATINALINE	http://platinoline.com	SITE	0A-00-27-00-00-17	29/02/2020 17:02:10
5	PLATINALINE	http://platinoline.com	SITE	0A-00-27-00-00-17	29/02/2020 23:02:40
6	PLATINALINE	http://platinoline.com	SITE	0A-00-27-00-00-17	01/03/2020 12:03:05
7	PLATINALINE	http://platinoline.com	SITE	0A-00-27-00-00-17	01/03/2020 12:03:19
8	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 12:03:24
9	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:51
10	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:19
11	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:55
12	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:35
13	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:31
14	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:32
15	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:34
16	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:02
17	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:59
18	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:03
19	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:05
20	FICHEIRO		FICHEIRO	0A-00-27-00-00-17	01/03/2020 13:03:09

Figura 45: Relatório de uma pesquisa no AWA

O ArquivoAO é a representação móvel do ARQUIVO WEB ANGOLANO:

Fonte: autor



Figura 46: Tela inicial do AWA Mobile

Fonte: autor



Figura 47: Resultado de uma pesquisa no AWA Mobile

6.9 TESTES NO ARQUIVO WEB ANGOLANO

Com foco em objectivos particulares distintos, no ARQUIVO WEB ANGOLANO foram realizados variados tipos de testes que tiveram como alvo as funcionalidades a ser realizadas pela aplicação e/ou características não funcionais.

Os testes foram feitos pela equipa desenvolvedora do projecto.

Objectivo do Teste:	Garantir que os métodos e processos de acesso a base de dados funcionam apropriadamente e sem corrupção dos dados.
Técnica:	Invocar cada método e processo de acesso a base de dados, alimentando cada um com dados ou requisições de dados válidos e inválidos. Inspeccionar a base de dados para garantir que os dados foram populados como pretendido, e que todos os eventos da base de dados ocorreram apropriadamente, ou revisar os dados retornados para garantir que os dados correctos foram recuperados pelas razões correctas.
Critério de Finalização:	Todos os métodos e processos de acesso a base de dados funcionam como projectados e sem nenhuma corrupção de dados.
Considerações Especiais:	Existiu a necessidade durante o teste de um ambiente de desenvolvimento e/ou drivers de SGBD para inserir ou modificar os dados directamente na base de dados Processos foram invocados manualmente Bases de dados pequenas ou minimizadas (número de registros limitados) foram utilizadas para aumentar a visibilidade de eventos não-aceitáveis.
Resultados:	APROVADO.

Teste de Integridade de Dados e Base de Dados

Objectivo do Teste:	Segurança do Nível de Aplicação: verificar que um actor pode aceder apenas aquelas funcionalidades e/ou dados para os quais o seu tipo de utilizador tem permissão. Segurança do Nível de Sistema: verificar que apenas aqueles actores com acesso ao sistema e aplicações têm permissão de acedelos.
Técnica:	Segurança do Nível de Aplicação: identificar e listar cada tipo de utilizador e as funcionalidades e/ou dados para os quais cada tipo tem permissão. Criar testes para cada tipo de utilizador e verificar cada permissão criando transacções específicas para cada tipo de utilizador. Modificar o tipo de utilizador e refazer os testes para os mesmos utilizadores. Em cada caso, verificar que funcionalidades e/ou dados adicionais estão correctamente disponíveis ou não.
Critério de Finalização:	Para cada tipo de actor conhecido as funcionalidades e/ou dados apropriados estão disponíveis, e todas as transacções funcionam como esperado e rodam nos Testes de Funcionalidades anteriores.
Considerações Especiais:	Nenhum
Resultados:	APROVADO.

Teste de Segurança e Controle de Acesso.

6.10 SOLUÇÕES DE SEGURANÇA

Um sistema sem conceitos de segurança aplicada é como se fosse uma casa sem porta onde todo mundo entra. Por isso para evitar esse tipo de vandalismo devemos aplicar medida para acautelar esses tipos de situações, garantindo assim a confidencialidade, integridade, autenticidade e a disponibilidade dos recursos da nossa aplicação, foram aplicados vários conceitos sobre segurança computacional para suprir essa necessidade.

Os principais mecanismos de segurança implementados tiveram como objectivo garantir a confidencialidade, a integridade e a autenticidade dos conteúdos armazenados, garantir a autenticidade dos utilizadores que participam no arquivamento e, por último, a privacidade na comunicação entre o cliente e o servidor.

6.11 CONTROLO DE ACESSO

Para garantir o pilar confidencialidade e integridade dos dados “segurança computacional”, utilizamos o modelo de proteção baseada em papel, de modos que os privilégios sejam controlados, permitindo assim alteração a qualquer momento com apenas um clique.

6.12 NEGAÇÃO DE SERVIÇO (DoS)

A disponibilidade dos recursos à tempo e hora é um dos principais factores no mundo das tecnologias, para garantir esse tipo de problema, temos uma configuração apropriada e personalizada para os servidores que velam sobre números de requisições, equipamentos autorizados, firewall habilitado, memória de consumo, tempo de execução, filtração de pacotes.

6.13 CSRF

A CSRF (do inglês *Cross-Site Request Forgery*) é um tipo de ataque que, ao executar um script malicioso no browser da vítima, acede a outro website, sem que esta se aperceba. O atacante consegue, assim, roubar a sessão da vítima. Para prevenção desse tipo de ataque utilizados conceitos de tokens temporários para cada requisição.

6.14 CRIPTOGRAFIA

A encriptação dos dados sensíveis é um factor primordial para garantir a confiança dos utilizadores de modo que somente os proprietários têm acesso aos seus recursos no sistema. Para cumprir com estes requisitos, os dados na base de dados foram aplicados técnicas de criptografia (simétricas e outras assimétricas) de modo que mesmo tempo acesso directo a base de dados não há percepção sobre os mesmos dados.

6.15 SQL INJECTION

Os ataques à base de dados hoje em dia são frequentes, logo a necessidade de protecção a estes dados estão cada vez mais rigorosos. Uma base de dados pode relevar dados sensíveis ou até mesmo permitir a alteração dos resultados. Para prevenir esta situação implementamos técnicas de protecção desse ataque malicioso.

6.16 PHP INJECTION E XSS

Devemos ter muito cuidado quando se trata de entrada de dados a partir da aplicação (principalmente via http ou https). Essas entradas podem conter códigos maliciosos que podem deixar a nossa aplicação instável ou mesmo inoperável. Para prevenir ataques do gênero, fizemos validação em todos os campos que permitem a entrada de dados de modo a certificar se tudo está sobre as nossas regras, impedindo assim qualquer entrada maliciosa.

7 CONCLUSÃO

Existem cada vez mais publicações disponíveis exclusivamente na Web. No entanto, a informação publicada neste meio de comunicação é muito mais volátil do que a publicada em documentos impressos. Os arquivos da Web são sistemas informáticos criados para realizar o arquivo automático da informação proveniente na Web e partilham o mesmo objetivo que os arquivos ou bibliotecas nacionais: a preservação de conhecimento para as gerações futuras. Existem várias iniciativas de arquivo da Web em todo o mundo que colaboram para enfrentar os vários e complexos desafios de preservar a informação publicada na Web.

O desenvolvimento do sistema de Arquivamento da *Web* Angolana para o apoio aos processos na área digital representa, consideravelmente, um contributo muito grande para preservação da história da web angolana;

Com os resultados obtidos, é possível diminuir as dificuldades que os investigadores científicos enfrentam na divulgação e na procura de conteúdo da sua produção;

Também é possível possibilitar aos utilizadores individuais um maior acesso às informações sobre a produção nacional, aumentando as suas opções em várias áreas;

7.1 Recomendações e Trabalho Futuro

O ARQUIVO WEB ANGOLANO foi concebido, desenhado, implementado, entretanto, há questões sempre por se melhorar, pelo facto de que actualmente ele consegue atender e suprir a maior parte das necessidades pontuais e dos problemas encontrados na área de arquivamento da Web, certamente que ainda faltam outras necessidades que de algum modo não puderam ser contemplada durante essa abordagem, pois optamos por priorizar os módulos tendo em vista a necessidades imediatas da web angolana, o que vai incidir com que, para essas eventuais necessidades não contempladas possam ser supridas através da implementação de novos módulos, funcionalidades ou ainda a melhoria de algumas funcionalidades já existentes.

Para uma maior robustez deste projecto, será bom estudar e aplicar as seguintes recomendações:

- Desenvolver a funcionalidade de arquivo de documentos digitais;
- Desenvolver a funcionalidade de arquivo de imagens digitais;
- Desenvolver um modelo matemático mais eficiente e aplicar no crawler, para possibilitar aos clientes possam fazer um benchmarketing;
- Aprimorar a funcionalidade de pesquisa por palavra ou conjunto de palavras;
- Aprimorar a aplicação *Mobile*, com as mesmas funcionalidades.
- Infraestrutura para processamento paralelo dos dados arquivados

Quando um motor de busca procura um termo dentro de uma única recolha da Web são encontrados milhões de textos mas os utilizadores normalmente analisam no máximo algumas dezenas de resultados. Os motores de busca sobre a Web usam algoritmos de ordenação dos resultados das pesquisas para mostrarem primeiro os resultados mais relevantes. No entanto, o estudo de algoritmos de ordenação para pesquisas de âmbito histórico sobre várias recolhas realizadas ao longo do tempo é um tópico de investigação pouco aprofundado até à data, mas fundamental para a criação de mecanismos de pesquisa eficientes sobre arquivos da Web. O estudo de algoritmos de ordenação adequados à Web angolana está em curso.

Para o sucesso do projecto do AWA nas suas múltiplas vertentes são necessários conhecimentos profundos sobre diferentes áreas de conhecimento. Por isso, é necessário cativar a participação de entidades externas à comunidade do arquivo da Web, principalmente a nível nacional. Pretende-se que o AWA, além de ser um fornecedor de recursos para investigação, possa contribuir para fomentar a colaboração dentro da comunidade científica nacional e servir como prova de conceito para resultados de investigações.

7.2 Impacto Social do Arquivo Web Angolano

O arquivamento da web tem como objetivo adquirir, preservar e fornecer acesso a informações históricas publicadas online. Em abril de 2013, havia pelo menos sessenta e quatro iniciativas de arquivamento da web em todo o mundo. Em conjunto, essas coleções arquivadas de documentos da web formam uma imagem abrangente de nossa história cultural, comercial, científica e social. O arquivamento da web também tem um impacto sociológico importante porque os cidadãos comuns estão publicando informações pessoais online sem preocupações de preservação. No futuro, os arquivos da web provavelmente serão a única fonte de memórias pessoais para muitas pessoas. Fornecemos alguns exemplos de ferramentas que facilitam a pesquisa histórica em arquivos da web, destacando seu potencial para as Humanidades.

7.3 Principais Dificuldades

Grande parte do esforço atual no desenvolvimento de arquivos da web se concentra na aquisição, armazenamento, gerenciamento e preservação de dados. No entanto, este é apenas o começo. Os dados devem ser acessíveis ao público para vê-los e explorá-los.

Nossa perspectiva é que os arquivos da web não atendem às necessidades de informação dos usuários. Este é o resultado de múltiplas causas:

- Primeiro, não há um entendimento claro de quem são os usuários e do que eles realmente precisam.
- Em segundo lugar, os arquivos da web são baseados nos mecanismos de pesquisa da web. No entanto, ambos não são projetados para lidar com a dimensão temporal criada pelos sucessivos instantâneos da web como um cidadão de primeira classe.
- Terceiro, não existem ferramentas para agilizar a coleta e a exploração das informações dos conteúdos arquivados.
- Quarto, não há avaliação da tecnologia de IR de arquivo da Web atual. Cada uma dessas causas representa um desafio de pesquisa que deve ser abordado.

ANEXOS

Anexo A - Cronograma de actividades

Quando se faz o planeamento, é uma boa prática definir as fases e/ou actividades que vão constituir o processo de desenvolvimento, o tempo e a duração de cada uma. As tabelas 1 e 2, mostram as actividades que foram desenvolvidas ao longo do processo.

Tabela 12: Cronograma do projecto

Fase	Actividade(s)	Tempo abrangido
Levantamento dos requisitos	Realização de entrevistas	01/07/2019 à 03/08/2019
	Definição dos actores do sistema e suas funções	15/08/2019 à 01/09/2019
	Definição dos casos de uso	15/08/2019 à 01/09/2019
Desenho e Implementação	Concepção da base de dados	11/09/2019 à 24/03/2020
	Desenvolvimento do protótipo da aplicação	12/09/2019 à 30/10/2019
	Desenvolvimento	25/01/2020 à 20/09/2020
Testes	Teste caixa-preta	07/05/2020 à 27/05/2020

Fonte: Autor

REFERÊNCIAS

- Universidade Agostinho Neto (2019). REGULAMENTO DE TRABALHOS DE FIM DO CURSO.
- Sommerville, I. (2011). *Engenharia de Software* (Vol. 9º). Pearson Education, Inc.
- Calado, M. P. (2015). Normas para Formatação e Apresentação de Monografias de Ciências da Computação. Luanda: Departamento de Ensino e Investigação de Ciências da Computação – Faculdade de Ciências – Universidade Agostinho Neto.
- Miguel Costa, Daniel Gomes, Mário J. Silva, The Evolution of Web Archiving, International Journal of Digital Library Systems, Abril de 2016
- Daniel Gomes, André Nogueira, João Miranda, Miguel Costa, Introducing the Portuguese web archive initiative, 8th International Web Archiving Workshop, Aarhus, Dinamarca, Setembro de 2008