

**GOVERNO DO DISTRITO FEDERAL
POLÍCIA MILITAR DO DISTRITO FEDERAL
DEC / INSTITUTO SUPERIOR DE CIÊNCIAS POLICIAIS
DIFORM / PRÓ-REITORIA DE FORMAÇÃO
ACADEMIA DE POLÍCIA MILITAR DE BRASÍLIA**

**PROPOSTA DE ARQUITETURA PARA VIABILIDADE DE IMPLANTAÇÃO DE UM
SISTEMA DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS COM A
UTILIZAÇÃO DE TÉCNICAS DE RECONHECIMENTO ÓTICO DE CARACTERES
PARA O CENTRO DE COMUNICAÇÃO SOCIAL DA PMDF**

PABLO ROGÉRIO GONÇALVES DE OLIVEIRA - ASP OF PM

**BRASÍLIA - DF
2015**



**GOVERNO DO DISTRITO FEDERAL
POLÍCIA MILITAR DO DISTRITO FEDERAL
DEC/INSTITUTO SUPERIOR DE CIÊNCIAS POLICIAIS
DIFORM/PRÓ-REITORIA DE FORMAÇÃO
ACADEMIA DE POLÍCIA MILITAR DE BRASÍLIA**



PABLO ROGÉRIO GONÇALVES DE OLIVEIRA – ASP OF PM

**PROPOSTA DE ARQUITETURA PARA VIABILIDADE DE IMPLANTAÇÃO DE UM
SISTEMA DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS COM A
UTILIZAÇÃO DE TÉCNICAS DE RECONHECIMENTO ÓTICO DE CARACTERES
PARA O CENTRO DE COMUNICAÇÃO SOCIAL DA PMDF**

BRASÍLIA/DF

2015



PABLO ROGÉRIO GONÇALVES DE OLIVEIRA – ASP OF PM

**PROPOSTA DE ARQUITETURA PARA VIABILIDADE DE IMPLANTAÇÃO DE UM
SISTEMA DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS COM A
UTILIZAÇÃO DE TÉCNICAS DE RECONHECIMENTO ÓTICO DE CARACTERES
PARA O CENTRO DE COMUNICAÇÃO SOCIAL DA PMDF**

Trabalho de Conclusão do Curso
apresentado ao Instituto Superior de
Ciências Policiais como requisito parcial
para obtenção do Título de Bacharel em
Ciências Policiais.

Orientador: MSc. Fábio Lúcio Lopes de
Mendonça

BRASÍLIA/DF

2015

PABLO ROGÉRIO GONÇALVES DE OLIVEIRA – ASP OF PM

**PROPOSTA DE ARQUITETURA PARA VIABILIDADE DE IMPLANTAÇÃO DE UM
SISTEMA DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS COM A
UTILIZAÇÃO DE TÉCNICAS DE RECONHECIMENTO ÓTICO DE CARACTERES
PARA O CENTRO DE COMUNICAÇÃO SOCIAL DA PMDF**

Trabalho de Conclusão do Curso apresentado ao Curso de Formação de Oficiais da Polícia Militar do Distrito Federal como requisito parcial para obtenção do Título de Bacharel em Ciências Policiais.

Aprovado em: 08 de Dezembro de 2015

BANCA EXAMINADORA

MSc. Fábio Lúcio Lopes de Mendonça – Orientador
Universidade de Brasília

MSc. Beatriz Campos Santana – Membro
Universidade de Brasília

MSc. Daniel Alves da Silva – Membro
Universidade de Brasília

Dedico esta monografia a minha família, que sempre incentivou e apoiou meus estudos. Dedico ainda a todos aqueles que de alguma forma estiveram e estão próximos a mim e que contribuíram para minha formação.

AGRADECIMENTOS

Agradeço a minha mãe e minhas irmãs que, com muito carinho e apoio, sempre estiveram do meu lado nos momentos difíceis dessa dura caminhada.

Agradeço a todos os instrutores do curso que contribuíram para a minha formação profissional e amadurecimento pessoal.

Ao meu orientador, MSc Fábio Lúcio Lopes de Mendonça, pelas sugestões e incentivo.

A professora de metodologia, MSc Alda Lino, pela paciência, sugestões e incentivos na orientação que tornaram possível a conclusão deste trabalho.

A todos os amigos, companheiros e irmãos que fiz ao longo desses três anos de Curso de Formação de Oficiais. Vocês tornaram essa conquista possível.

E não poderia deixar de agradecer a minha namorada, Ingrid, pelo carinho, ajuda e paciência quando eu estava preocupado com a elaboração do trabalho.

Quanto mais um homem se aproxima de suas metas, tanto mais crescem as dificuldades. (Johann Goethe)

RESUMO

Quanto maior uma instituição maior será a quantidade de documentos gerados em decorrência de suas atividades. Todos os dias as pessoas criam, modificam e visualizam esses documentos. Com a popularização da tecnologia e as diversas opções de ferramentas utilizadas, surge uma grande quantidade de formato de arquivos digitais. O objetivo deste trabalho é propor uma arquitetura de Gerenciamento Eletrônico de Documentos para o Centro de Comunicação Social da PMDF objetivando a melhoria do acesso a informação. A metodologia utilizada primeiramente foi uma pesquisa bibliográfica, com análise do que já foi produzido sobre o tema. A pesquisa *in loco* permitiu observar como é coletada e disseminada a informação no CCS e aspectos que podem ser melhorados a fim de construir um modelo de gerenciamento de informação para o setor e para as organizações como um todo. A utilização de equipamentos de alta tecnologia é um dos requisitos da implantação do GED, sendo fundamental observar a aquisição destes equipamentos como um investimento que auxilia na construção de um sistema de informação atual, moderno que atenda a demanda da sociedade.

Palavras-chave: Informação. Tecnologia. Sociedade. Arquivos digitais.

ABSTRACT

The bigger is an institution, more will be the amount of documents generated as a result of their activities. Every day people create, modify and view these documents. With the popularization of technology and the various tools used options, comes a lot of digital file format. The objective of this work is to propose a Eletronic Document Management architecture for the PMDF Social Communication Center aimed at improving access to information. The methodology used was primarily a bibliographic research, with analysis of what has already been produced on the subject. The research allowed to observe in loco as is collected and disseminated information on CCS and aspects that can be improved in order to build an information management model for the sector and for organizations as a whole. The use of high-tech equipment is one of the EDM deployment requirements making it essential to observe the purchase of this equipment as an investment that helps build a current, modern information system that meets the demands of society.

Key words: Information. Technology. Society. Digital files.

LISTA DE FIGURAS

Figura 1 - Ciclo de Vida dos Documentos	23
Figura 2 - Ambiente Típico GED	26
Figura 3 - Etapas da digitalização	28
Figura 4 - Componentes de um sistema de OCR.....	31
Figura 5 - Interação em Sistema de Banco de Dados	37
Figura 6 - Sistemas Distribuídos	40
Figura 7 - Modelo de Arquitetura Cliente-Servidor	42
Figura 8 - Arquitetura em módulos para um sistema de OCR.....	50
Figura 9 - Subdivisão do módulo principal	51
Figura 10 - Subdivisão do módulo de pré processamento	53
Figura 11 - Detalhamento dos inter-relacionamentos da arquitetura.....	53
Figura 12 - Uso do padrão de projeto Facade.....	55
Figura 13 - Diagrama de classe do protótipo.....	58
Figura 14 - Classe do módulo principal	59
Figura 15 - Classe do módulo extrator	60
Figura 16 - Classe do módulo Classificador	61

LISTA DE TABELAS

Tabela 1 - Efetivo da PMDF	18
Tabela 2 - Descrição dos módulos apresentado na arquitetura	50

LISTA DE ABREVIATURAS E SIGLAS

CCS	Centro de Comunicação Social
CONARQ	Conselho Nacional de Arquivos
DAI	Divisão de Assessoria de Imprensa
GED	Gerenciamento Eletrônico de Documentos
GEPEs	Sistema de Gerenciamento de Pessoal
ICR	<i>Intelligent Character Recognition</i>
IDL	<i>Interface Definition Language</i>
MEC	Ministério da Educação
OCR	<i>Optical Character Recognition</i>
PMDF	Polícia Militar do Distrito Federal
RIM	<i>Records and Information Management</i>
SGDB	Sistema Gerenciados de Banco de Dados

SUMÁRIO

1 INTRODUÇÃO	14
2 REFERENCIAL TEÓRICO	17
2.1 Breve Histórico da Polícia Militar do Distrito Federal	17
2.1.1 O Centro de Comunicação Social da PMDF.....	19
2.2 Informações Arquivísticas e a Gestão de Arquivos.....	20
2.2.1 Ciclo de Vida dos Documentos	22
2.3 Gestão Eletrônica de Documentos	24
2.3.1 História e Definições	24
2.3.2 Tecnologias envolvidas	27
2.3.3 Mecanismos de Digitalização	27
2.3.4 Sistema de Reconhecimento de Caracteres.	30
2.3.5 Vantagens	33
2.3.6 Desvantagens	34
2.3.7 Sistema Informatizado de Gestão de Arquivos de Documentos	34
2.4 Banco de Dados	35
2.4.1 Sistema de Gerenciamento de Banco de Dados - SGBD	37
2.5 Sistema Distribuídos	39
2.6 Metodologia.....	43
2.6.2 Instrumentos de Pesquisa	45
2.7 Análise dos Dados e Proposta de GED	47
3 PROPOSTA DE ARQUITETURA	49
3.1 Representação da Arquitetura	49
3.1.1 Módulo de Interface Gráfica.....	51
3.1.2 Módulo principal.....	51
3.1.3 Módulo de pré-processamento	52
3.1.4 Inter-Relacionamento entre os módulos	53
3.1.5 Padrões utilizados.....	54
3.1.5.1 Padrões <i>Facade</i>	54
3.1.5.2 Padrões <i>Singleton</i>	55
3.2 Detalhamento do objeto	55
4 DESENVOLVIMENTO DO PROTÓTIPO	57

4.1.1 Pontos Iniciais.....	58
4.1.2 Classe GUI.....	59
4.1.3 Classe OcrManager.....	59
4.1.4 Classe Extrator.....	60
4.1.5 Classe Classificador.....	60
5 CONCLUSÃO.....	62
5.1.1 Trabalhos futuros.....	64
REFERÊNCIAS.....	65

4.1.1 Pontos Iniciais.....	58
4.1.2 Classe GUI.....	59
4.1.3 Classe OcrManager	59
4.1.4 Classe Extrator	60
4.1.5 Classe Classificador	60
5 CONCLUSÃO	62
5.1.1 Trabalhos futuros	64
REFERÊNCIAS.....	65

1 INTRODUÇÃO

A pesquisa possui como tema de estudo a Gestão Eletrônica de Documento (GED) no Centro de Comunicação Social (CCS) da Polícia Militar do Distrito Federal (PMDF), tendo em vista ser dever do Poder Público a gestão documental e a proteção especial a documentos produzidos pelos órgãos públicos, de acordo com o Art. 1º da Lei n.º 8.159 de 08 de janeiro de 1991.

O trabalho está amparado no Núcleo de Ensino e Pesquisa em Segurança Pública, Violência e Conflitualidades (NEPES) na área de concentração da Atividade Policial Reflexiva. A pesquisa concentra seus esforços no estudo das atividades que subsidiam a prática policial militar, por entender que as atividades burocráticas são importantes quando percebidas como parte de um complexo sistema que acaba por resultar na sensação de segurança.

O que impulsionou a realização deste trabalho foi entender a importância da aplicação do gerenciamento eletrônico de documentos dentro do Centro de Comunicação Social da Polícia Militar do Distrito Federal, apresentando conceitos, definições e ferramentas necessárias à tomada de decisão para implantação e manutenção de um sistema de GED, voltado para os policiais militares e a população em geral que buscam informações sobre ocorrências ou outros informativos a respeito da instituição.

Os documentos constituem o grande volume de conhecimento e de valor histórico de uma empresa, organização ou instituição e essa quantidade de conhecimento produzida pode ser o diferencial para o sucesso dos seus negócios. Devemos levar em consideração que os documentos produzidos no desenrolar da atividade das organizações servem, além de registro histórico, para mostrar suas decisões, projetos e procedimentos. Além disso, possuem caráter informativo e probatório para prestação de contas de suas atividades ao cidadão.

Como todo órgão público, a PMDF também produz muitos documentos importantes para o planejamento e execução dos seus serviços e o armazenamento em papel demanda um grande espaço físico e maior dificuldade na localização de uma informação existente no documento. Em contrapartida, o armazenamento digital oferece maior confiabilidade no manuseio do documento, menor espaço físico para armazenamento e acesso rápido às informações contidas no documento.

No entanto, não é suficiente apenas armazenar os documentos em formato digital, é preciso que os documentos digitais sejam geridos de forma eficiente, armazenados adequadamente e que seja possível localizar dados de forma rápida e eficiente.

Para isso a tecnologia da informação nos oferece um conjunto de tecnologias que permite uma instituição gerenciar seus documentos no formato digital, é o Gerenciamento Eletrônico de Documentos - GED.

Gerenciar eletronicamente documentos é uma solução que traz resultados bastantes significativos para as instituições. Para Baldam, Valle e Cavalcanti (2004), as principais razões para implementar um ambiente GED são a redução do tempo de processamento, redução do manuseio do papel, alta velocidade e precisão na localização de documentos e melhor atendimento ao cliente.

Ao observar os fatos supracitados, faz-se necessário perguntar: "a implantação da tecnologia GED no Centro de Comunicação Social da PMDF pode otimizar a busca, o armazenamento e recuperação das notícias?"

Nesse sentido, há indícios de que a implantação de um sistema GED pode vir a contribuir significativamente para a forma como são geridas as publicações de ocorrências e os informativos feitos pelo Centro de Comunicação Social da PMDF.

Dessa forma, o objetivo geral deste trabalho é propor uma arquitetura para viabilidade de implantação de um sistema de gerenciamento eletrônico de documentos com a utilização de técnicas de OCR para o Centro de Comunicação Social da PMDF, de maneira a possibilitar a indexação do conteúdo das notícias e o armazenamento em bases de dados distribuído, podendo assim estar disponível para consulta para público interno da PMDF e a população em geral.

Para a efetividade deste trabalho, a pesquisa gerou os seguintes objetivos específicos: apresentar a Polícia Militar do Distrito Federal com a configuração atual, e o histórico de sua evolução, desde sua primeira instalação na cidade do Rio de Janeiro; apresentar o Centro de Comunicação Social da PMDF, com suas atribuições e principais atividades; conceituar arquivo, documentos e gestão de documentos; definir o conceito de Gerenciamento Eletrônico de Documentos, passando por suas tecnologias, as principais ferramentas, vantagens e desvantagens; definir o que é um banco de dados e um Sistema de Gerenciamento de Banco de Dados e conceituar sistemas distribuídos.

A presente pesquisa foi estruturada em cinco capítulos. No primeiro capítulo, foi apresentado o tema da pesquisa, a área de concentração e a linha de pesquisa de acordo com o NEPES, em seguida a justificativa da pesquisa, apontando a importância da mesma e o que motivou o pesquisador a trabalhar o tema em questão, logo após a contextualização do tema e por último o objetivo geral e os objetivos específicos da pesquisa.

No segundo capítulo, será apresentado o referencial teórico contemplando as informações disponíveis na literatura referente ao tema, consistindo na descrição da instituição, na identificação dos principais assuntos relacionados à gestão eletrônica de documentos, para em seguida abordar conceitos, mecanismos, técnicas, vantagens e desvantagens do GED, bem como as principais técnicas de reconhecimento de caracteres e os conceitos de bancos de dados e sistemas distribuídos. O segundo capítulo mostra ainda a metodologia usada para se chegar ao entendimento de como é feito o processo de gestão de documentos no CCS além de expor uma proposta de gerenciamento eletrônico para os documentos de divulgação de notícias.

No terceiro capítulo, será apresentado a proposta de arquitetura lógica com descrição dos principais módulos propostos para a implementação de um sistema GED com técnicas de OCR, com os diagramas de casos de uso, bem como a implementação e alguns resultados encontrados.

No quarto capítulo, será apresentado um protótipo da arquitetura tratada no terceiro capítulo.

O quinto capítulo trará o resultado de todas as questões levantadas, com uma breve conclusão do trabalho com os objetivos atingidos através da arquitetura proposta.

2 REFERENCIAL TEÓRICO

2.1 Breve Histórico da Polícia Militar do Distrito Federal

A Polícia Militar do Distrito Federal tem suas origens no início do século XIX com a vinda da corte portuguesa para o Brasil. Com a invasão de Portugal pelas tropas de Napoleão Bonaparte, o príncipe regente, Dom João VI, decide se mudar para o Brasil, juntamente com a sua corte. Ao chegar ao Rio de Janeiro, Dom João VI promove uma reorganização do Estado e cria diversas instituições que ajudaram na estruturação. Dentre essas instituições, foi criada a Divisão Militar da Guarda Real de Polícia. Criada nos moldes das forças armadas, com base na hierarquia e disciplina, com postos e graduações semelhantes aos utilizados pelo Exército Brasileiro. A Guarda Real substituiu os "quadrilheiros", que eram os responsáveis pela segurança pública do Estado (BRETAS, 1998).

Enquanto Brasília estava sendo criada e a polícia do Distrito Federal ainda não havia sido transferida para a nova capital, o serviço de policiamento urbano e rural foi exercido por várias polícias com diversas denominações, dentre elas a Guarda de Polícia, Guarda Rural e o Serviço de Polícia Metropolitana (SILVA, 2009).

Ainda de acordo com Silva (2009), no final de 1961, foi criada a Polícia Metropolitana do Distrito Federal (utilizando a sigla PMDF pela primeira), tendo como função o policiamento ostensivo, o policiamento judiciário com serviços de natureza política e social, fiscalização de trânsito e outros.

Com a transferência dos policiais militares do Rio De Janeiro, a lei n.º 4.483, de 16 de novembro de 1964, que reorganiza o Departamento Federal de Segurança Pública, acrescentou a Polícia Militar do Distrito Federal à sua estrutura e a composição da nova polícia foi formada basicamente pelos militares que vieram do Rio de Janeiro e optaram por continuar no Distrito Federal, oficiais da reserva remunerada do Exército Brasileiro e mais alguns remanejados de outras instituições que faziam parte da segurança pública.

Segundo o Decreto Nº 31.793, de 11 de junho de 2010, que dispõe sobre a organização básica, a Polícia Militar do Distrito Federal conta com uma estrutura de 23 (vinte e três) batalhões operacionais, 19 unidades especializadas, além das unidades médico-hospitalares e de ensino.

Todo o aparato policial é algo que deve ser considerado pela Segurança Pública do Distrito Federal, tendo em vista o crescimento populacional, vale salientar que a Polícia Militar do Distrito Federal se encontra com uma proporcionalidade razoável em relação às demais Unidades Federativas, cerca de um policial militar para cada 135 habitantes.

Segundo o Sistema de Gerenciamento de Pessoal - GEPES - a PMDF conta hoje com um efetivo de aproximadamente 14.523 homens, responsáveis pela segurança pública de todo o Distrito Federal, distribuídos conforme tabela abaixo.

Tabela 1 - Efetivo da PMDF

Lotação	Qtd	Lotação	Qtd	Lotação	Qtd	Lotação	Qtd	Lotação	Qtd
1° BPEsc	427	3° BPTran	349	CCS	46	DCC	142	DSAP	40
1° BPM	458	4° BPM	328	CI	147	DEA	4	EM	62
1° BPTM	222	5° BPM	292	CLDF	9	DEC	46	GCG	56
1° BPTran	310	6° BPM	378	CM-GDF	195	DEEC	21	GOV AMAPÁ	1
10° BPM	421	7° BPM	215	CMan	162	DEOF	42	GSII/PR	80
11° BPM	519	8° BPM	459	CME	38	DGP	61	MJ	3
12° BPM	158	9° BPM	448	CMed	200	DiCC	10	MP	2
13° BPM	328	ABIN/PR	25	CMT	89	DiForm	10	MPDFT	19
14° BPM	440	Adm Regional	8	CMV	44	DIPC	67	MPF	17
16° BPM	202	APMB	168	CPCDH	85	DiPro	37	OUTR ORGÃO	18
17° BPM	334	Aud. Militar	4	CPRL	52	DiTel	68	Ouvidoria	54
19° BPM	227	Auditoria	30	CPRM	30	DIV DESTIN	88	RPMON	325
2° BPM	475	BAvOp	79	CPRO	37	DLF	30	SG	69
20° BPM	287	BOPE	141	CPRS	29	DOP	47	SSP-DF	177
21° BPM	253	BPCães	116	CPSO	30	DPAD	21	STJ	2
24° BPM	189	BPCHOQUE	350	CTEsp	60	DPGC	59	STM	2
25° BPM	229	BPMA	490	DAE	12	DPM	90	TSE	1
26° BPM	369	C ODONTO	89	DALF	65	DPMT	79	TST	2
27° BPM	277	CAEAp	158	DAM	18	DPPHC	76	VGDF	38
28° BPM	259	CASO	86	DAO	15	DPPP	66	VPR	1
3° BPM	457	CCF	29	DAP	30	DRS	14	Total	14523

FONTE: SISTEMA DE GERENCIAMENTO DE PESSOAL - 09/11/2015

O art. 144, § 5º, da C.F, disciplina que, "Às policias militares cabem a polícia ostensiva e a preservação da ordem pública; aos corpos de bombeiros militares, além das atribuições definidas em lei, incumbe a execução de atividades de defesa civil". De acordo com o texto constitucional, fica demonstrado que a polícia militar exerce a função de polícia administrativa, sendo responsável pelo policiamento preventivo e pela manutenção da ordem pública.

De acordo com Soibelman (1994, p. 278), Polícia Ostensiva, "é a que age de uma forma visível pelo público. Opõe-se a polícia secreta [...]. é a que obtém resultados preventivos pela simples ação da presença". Sobre Polícia Preventiva, Soibelman (1994, p. 278) esclarece que são: "Medidas adotadas pela administração pública para prevenir comprometimento da segurança, higiene, moralidade ou economia pública".

2.1.1 O Centro de Comunicação Social da PMDF

O Centro de Comunicação Social é um órgão de apoio ao Comando Geral. O Decreto Lei n.º 31.793, de 11 de junho de 2010, define órgão de apoio como órgãos que prestam assistência direta e imediata ao Comandante Geral e ao Subcomandante-Geral.

O Centro de Comunicação Social da PMDF está estruturado sob as seguintes divisões:

- Divisão Administrativa;
- Divisão de Planejamento;
- Divisão de Marketing e Criação;
- Divisão de Assessoria de Imprensa;
- Divisão de Relações Públicas.

O acesso à informação é fundamental para a construção de uma sociedade democrática e cidadã. A Lei n.º 12.527, de 18 de novembro de 2011, prevê em seu art. 6º, que cabe aos órgãos e entidades do poder público, assegurar gestão transparente da informação, propiciando amplo acesso a ela e sua divulgação.

A comunicação social engloba várias áreas: relações públicas, jornalismo, publicidade e propaganda, audiovisuais e multimídia, entre outras. O profissional da comunicação social poderá trabalhar em agências de publicidade, na promoção de eventos, em assessoria de empresa, em editoras, institutos de pesquisa, gráficas, departamentos de comunicação e marketing de empresas, etc.

O Art. n.º 107 do Decreto Lei supracitado, dispõe que compete ao Centro de Comunicação Social da PMDF: planejar, desenvolver, coordenar e executar as atividades de comunicação social da PMDF; propor a edição de publicações

destinadas ao público interno e externo; promover a imagem institucional e mercadológica da Polícia Militar do Distrito Federal nos ambientes sociais e de segurança pública, dentre outros.

Além da comunicação interna, aquela que vai transmitir o pensamento e ações da empresa para o público interno, o Centro de Comunicação Social também é responsável pela comunicação externa da PMDF. De acordo com Bahia (1995) a comunicação externa é a comunicação feita pela empresa para um público fora dos limites internos. É o uso que a empresa faz dos meios de comunicação para o fim de tornar conhecida suas mensagens.

Dessa forma, é possível observar a importância do Centro de Comunicação Social como ferramenta para melhorar a imagem da PMDF junto à população do Distrito Federal.

2.2 Informações Arquivísticas e a Gestão de Arquivos

Analisando o aspecto semântico da palavra, Fernández e Arroyo (1982) definem o sentido de “documento” como uma instrução que é dada a alguém sobre qualquer matéria. Os mesmos autores definem documento também como um testemunho escrito de épocas passadas, que servem para reconstruir sua história.

Dessa forma, os documentos podem ser considerados como a memória de uma instituição. Neles estarão contidos todos os tipos de informações, desde registros históricos até escritos que contenham valores probatórios.

O Dicionário Houaiss da Língua Portuguesa (2001) segue a mesma linha de Fernández e Arroyo e define documento como sendo:

[...] qualquer escrito usado para esclarecer determinada coisa; qualquer objeto de valor documental (fotografia, peças, papéis, filmes, etc.) que elucide, instrua prove ou comprove cientificamente algum fato, acontecimento, dito, etc.

Já o arquivo pode ser considerado como um conjunto de documentos. E tem como finalidade a disponibilização, organização e preservação, independentemente da forma e da data que foi criado. A lei n.º 8.159, de 08 de janeiro de 1991, define arquivo como sendo:

Art. 11 Os conjuntos de documentos produzidos e recebidos por órgãos públicos, instituições de caráter público e entidades privadas, em decorrência do exercício de atividades específicas, bem como por pessoa física, qualquer que seja o suporte da informação ou a natureza dos documentos.

A principal função do arquivo é proteger e disponibilizar as informações contidas no seu interior. O CONARQ, em seu Dicionário Brasileiro de Terminologia Arquivística (2005, p. 1), define arquivo como "um conjunto de documentos produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades, independente da natureza dos suportes".

Dessa forma, é possível perceber a importância de preservação de um documento para uma instituição. Luciana Duranti (2002) cita que o documento é gerado no curso de uma atividade prática e serve como fonte de prova de ação que o gerou, sendo que o valor desta fonte depende da fidedignidade e autenticidade do documento.

Entretanto, as instituições não utilizam apenas os documentos em papéis para exercerem as suas atividades, elas também fazem uso do computador para criação de vários documentos, são os documentos digitais. O documento digital é a informação registrada, codificada em códigos binários e acessível por meio de sistema computacional (CONARQ, 2011).

Conseqüentemente, instituições públicas e privadas têm produzido e acumulado documentos eletrônicos de caráter arquivístico em grande volume e velocidade. Esse é um fenômeno que se acentuou a partir da década de 80, com a disseminação das tecnologias da informação, e isso representa um grande desafio porque, se não houver um tratamento adequado para esses documentos, haverá riscos de perda de provas ou perda histórica.

O art. 3º da lei 8.159/91 define gestão de documentos como:

[...] o conjunto de procedimentos e operações técnicas que engloba a produção, a tramitação, o uso, a avaliação e o arquivamento em fase corrente e intermediária, visando sua eliminação ou recolhimento para a guarda permanente.

Rondinelli (2005) ressalta que o conceito de gestão de documentos restaura e dinamiza a concepção dos arquivos como instrumentos facilitadores da administração.

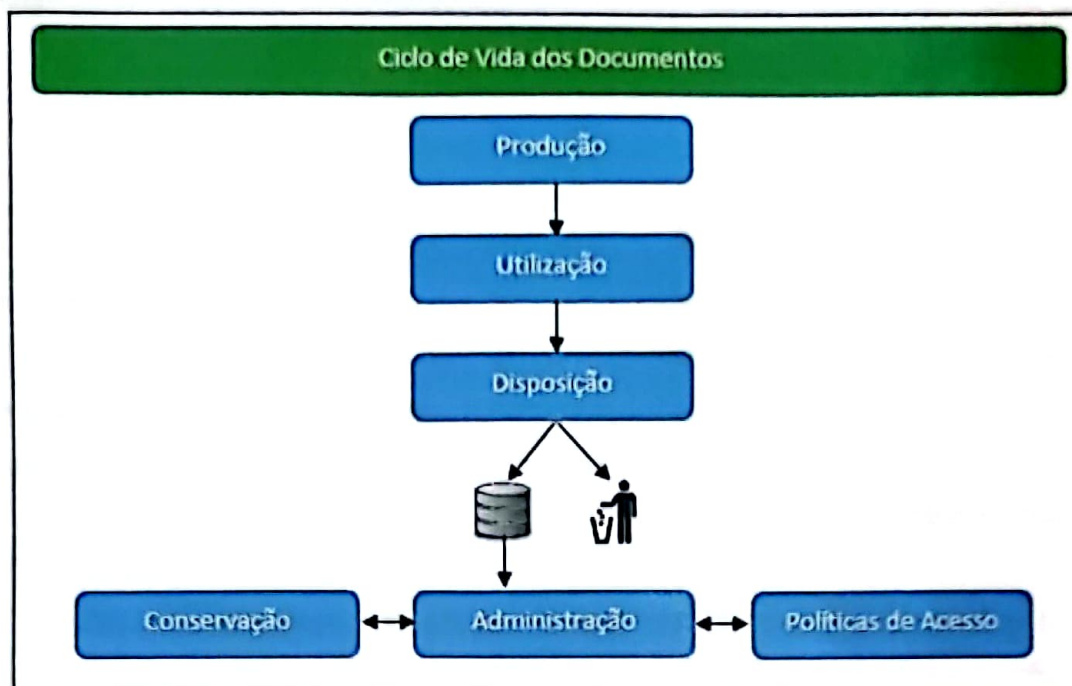
Assim, é preciso que as instituições façam a gestão dos seus documentos durante todo o ciclo de vida do documento. A gestão de documentos eficiente deve ser o objetivo de qualquer instituição.

2.2.1 Ciclo de Vida dos Documentos

O ciclo de vida de um documento pode ser definido, segundo Rhoads (1989, p. 3), em quatro fases, assim resumidas:

- a) **Elaboração/Criação** – é fase que contempla o desenho e produção de documentos, estipulação de diretrizes e fomento de sistemas de gestão e aplicação de tecnologia moderna aos processos;
- b) **Utilização e manutenção do documento** – é a fase que vai do nascimento à vida produtiva e compreende a utilização de documentos necessários para as atividades de uma instituição. Engloba a recuperação de documentos, a gestão de arquivos, a análise de documentos vitais, etc.
- c) **Disponibilização do documento** – é a fase de destinação. Engloba a identificação e a descrição do documento. É ainda nessa fase que o documento vai receber a decisão sobre a sua conservação.
- d) **Gestão de documentos de caráter permanente** – é fase que engloba os mecanismos de armazenamento, os processos de conservação e preservação dos arquivos. É também a fase que de planejamento para as políticas de acesso às informações dos arquivos.

Figura 1 - Ciclo de Vida dos Documentos



Fonte: Elaborado pelo autor

Os documentos podem ser armazenados por três meios: papel, microfimes e mídia eletrônica. Cada um com suas vantagens e desvantagens (FANTINI, 2001).

O papel é considerado uma das mídias mais usadas devido ao seu baixo custo e a facilidade de acesso. Além disso, apresenta uma facilidade de leitura e possui uma facilidade de alteração. Todavia, o papel possui algumas desvantagens, como o excesso de volume, a facilidade de extravio, deterioração e a necessidade de uma grande quantidade pessoas para o seu gerenciamento.

A utilização de microfimes possui a vantagem de utilizar um espaço de armazenamento menor do que o original e possui um prazo de validade maior. Outra vantagem sobre o armazenamento em papel é a garantia de autenticidade, uma vez que não é possível fazer alterações no documento gerado com o microfilme. Possui como desvantagem a necessidade de equipamentos especiais para o seu manuseio e leitura. Podem ainda ocorrer problemas de incompatibilidade, causados pela geração da tecnologia dos equipamentos e, conseqüentemente, aumentando o custo de utilizar essa tecnologia.

As mídias eletrônicas têm um custo elevado e por isso deve ser feita uma análise para saber o que vai ou não ser convertido para o formato digital. Em alguns casos é aconselhável manter uma estrutura de armazenamento mista.

2.3 Gestão Eletrônica de Documentos

2.3.1 História e Definições

Após a Segunda Guerra Mundial, o uso do computador não ficou restrito apenas aos militares, essa tecnologia foi incorporada às instituições públicas e privadas. Até a década de 70 o uso do computador ficou restrito às grandes empresas. Apenas com o surgimento do computador pessoal, na década de 80, o computador começou a fazer parte da vida das pessoas.

Com a difusão dos documentos digitais, surge a necessidade de estudo sobre a gestão de documentos eletrônicos.

O GED foi introduzido no Brasil pelo CENADEM (Centro Nacional da Gestão da Informação), entidade extinta em 2010. Segundo Rondinelli (2005, p. 116), "o GED surgiu no Brasil por parte do Poder Executivo Federal, de programas voltados para a disponibilização de serviços e de informações ao cidadão via internet". Era o final da década de 80 e começo da década de 90 e ainda não possuía as características e tecnologias que fazem parte dos sistemas de gestão eletrônica de documentos nos dias atuais, mas foram os primeiros passos para a popularização do GED no Brasil.

De acordo com Avedon (1999), GED é a configuração de equipamento, *software* e de recursos de telecomunicações baseada em computador e automatizada, que armazena e gerencia imagens de documentos, e seus índices codificados. Estas informações podem ser lidas por máquinas e processadas por computador para recuperação sob solicitação.

Sprague Jr. (1995), define GED analisando cada um dos seus termos de forma isolada. Gerenciamento pode ser entendido como a criação, armazenamento, organização, transmissão, recuperação e manipulação de documentos para cumprir um propósito organizacional. O segundo termo – Eletrônico – é o uso de modernas tecnologias de informação. E o último termo – Documentos – pode ser entendido como um conjunto de informações relativas a um tópico, estruturado para a compreensão humana, representada por uma variedade de símbolos, armazenado e manuseado de forma unitária.

As ferramentas GED servem principalmente para gerenciar documentos desde a sua criação e todos os estágios do seu ciclo de vida. Para Baldam, Valle e

Cavalcanti (2004), documento é qualquer papel que contenha informação relevante, ou qualquer arquivo digital que possa ser impresso de forma legível e, conseqüentemente, transformado em documento em papel. Dessa forma, o GED atuará desde a sua criação, passando pela emissão de cópias, distribuição, revisão e até a fase em que ele vai ser descartado ou armazenado.

Quanto maior uma instituição, maior será a quantidade de documentos gerados em decorrência de suas atividades. Todos os dias as pessoas criam, modificam e visualizam esses documentos. Com a popularização da tecnologia e as diversas opções de ferramentas utilizadas, surge uma grande quantidade de formatos de arquivos digitais. Segundo Baldan, Valle e Cavalcanti (2004, p. 101) os formatos mais comum utilizados para grande produção de documentos e/ou objetos em GED são:

- Texto: .doc, .rtf, .txt;
- Planilha: .xls;
- CAD: .dwg, .dgn, .vsd;
- Planejamento: .mmp.

Levando em consideração o exposto acima, GED é o conjunto de todas as tecnologias voltadas para gerenciar informações de forma eletrônica. Koch (1998) cita que essas informações podem ser:

Voz – informações geradas de forma verbal;

Texto – informações mais formais, desde cartas a contratos, planilhas, etc;

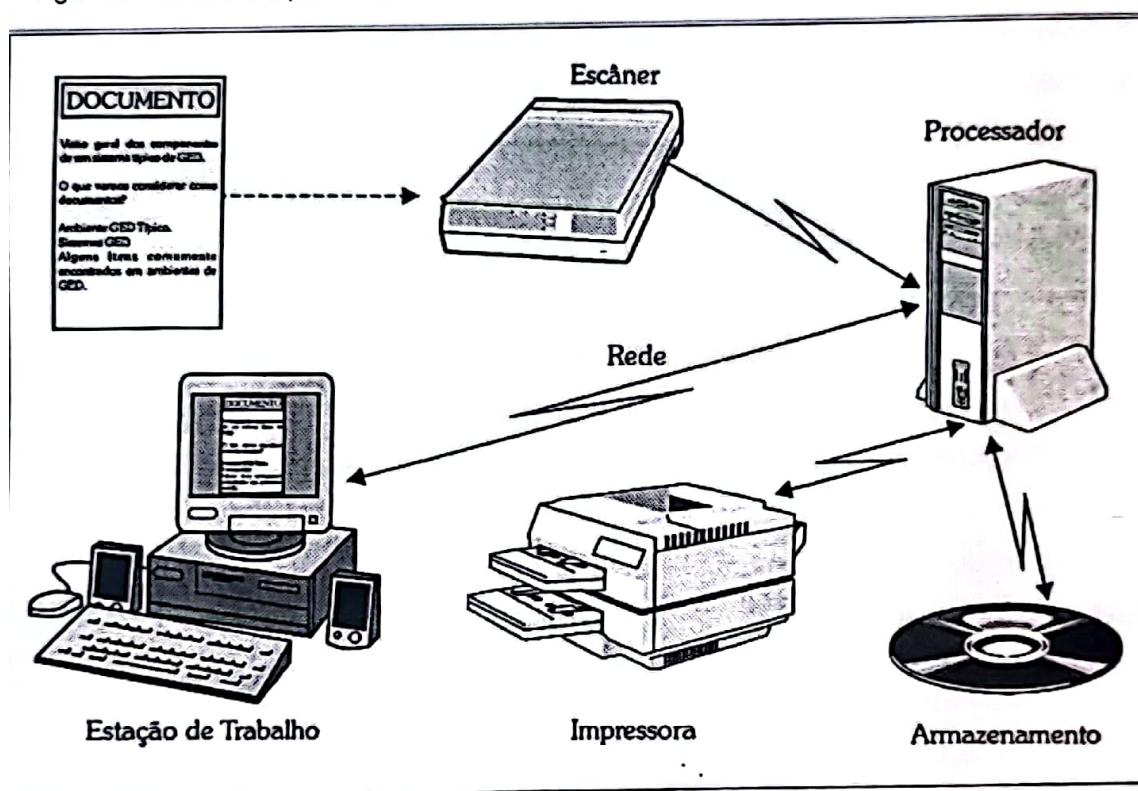
Imagem – informações que não podem ser representadas nas formas anteriores, como mapas, vídeos, fotografias, assinaturas, etc.

Essas informações podem estar armazenadas em mídias analógicas ou mídia digital. Caso as informações estejam em mídias analógicas, as mesmas serão convertidas para o formato digital. O GED é ferramenta que possibilita a conversão dos documentos em seu formato original para o documento eletrônico e permite o seu gerenciamento através de um software que irá admitir a indexação, armazenamento, pesquisa e visualização.

Embora um ambiente GED possa ter configurações totalmente variadas, dependendo da aplicação desejada, Baldan, Valle e Cavalcante (2004, p. 39) mostram o que pode ser considerado um ambiente GED típico, conforme apresentado na figura 2, e a descrição dos seus componentes:

- **documento:** pode estar em papel ou nativamente digital;
- **escâner:** equipamento utilizado para digitalizar o documento, ou seja, uma imagem do documento físico a ser armazenado eletronicamente;
- **processador:** normalmente os sistemas GED são instalados em um ou mais computadores, normalmente servidores, em rede para facilitar a distribuição de informação;
- **rede:** meio de comunicação entre os diversos componentes do sistema;
- **armazenamento:** pode ser o próprio servidor de imagens ou outro ambiente computacional;
- **impressora:** sempre usada quando da necessidade de obter uma cópia física do documento;
- **estação de trabalho:** computador para acesso ao servidor que pode permitir

Figura 2 - Ambiente Típico GED



Fonte: Baldam, Valle e Cavalcanti (2004, p. 40)

Para a implementação do GED em uma instituição, poderão ser necessário vários *softwares* para ter uma solução completa. O conjunto de programas que administram um ambiente GED pode ser considerado um sistema. Um Software de GED dificilmente possui todas estas funções, mesmo porque em separado eles dão melhor resultado e permitem que você integre as melhores solução para o seu problema (BALDAM, VALLE e CAVALCANTI, 2004).

2.3.2 Tecnologias envolvidas

Várias são as tecnologias envolvidas no processo de conversão para o formato digital, dentre elas podemos citar as técnicas *document imaging* (DI), *forms processing* (processamento de formulários), *Records and Information Management* – RIM, gerenciamento corporativo de relatórios, automação de fluxo de trabalho e técnicas de digitalização (BALDAM, VALLE e CAVALCANTI, 2004).

Documento Imagin é o gerenciamento da imagem dos documentos. É utilizado para agilizar a consulta processamento e distribuição dos documentos.

Os *forms processing* ou processamento de formulários é utilizado principalmente por bancos para agilizar o processo dos formulários de abertura de contas e concessão de créditos. Permite reconhecer as informações contidas em um formulário e relacioná-las a um banco de dados, automatizando o processo de digitação.

O RIM – *Records and Information Management* – é o gerenciamento de arquivos. É a tecnologia responsável pelo gerenciamento do ciclo de vida dos documentos, independente da mídia em que ele se encontra.

O gerenciamento corporativo de relatórios possibilita que os relatórios sejam gerados e gerenciados no formato digital. Permite ainda fazer anotações sobre o relatório sem afetar o documento original.

O *workflow* ou automação de fluxo de trabalho garante a integração dos departamentos envolvidos em processos com vários documentos, permitindo a execução das tarefas pelas pessoas corretas no tempo previamente definido.

2.3.3 Mecanismos de Digitalização

O objetivo principal do GED é a redução da quantidade de papéis produzidos nas empresas ou organizações, utilizando técnicas de digitalização para transformar para o formato eletrônico.

Segundo o CONARQ (2011), digitalização é um processo de conversão de um documento para o formato digital, por meio de dispositivo apropriado. Geralmente o dispositivo utilizado é o escâner. No final do processo, teremos um documento

eletrônico contendo uma imagem do documento em papel. A figura abaixo mostra as principais etapas de um processo de digitalização.

Figura 3 - Etapas da digitalização



Fonte: Site Delta – Gestão de Documentos

As principais vantagens da digitalização são a redução da deterioração com o tempo e natureza, redução do espaço físico para armazenamento e dificuldades contra fraude.

Outra técnica de conversão de documentos para o formato digital é a microfilmagem. A microfilmagem é o serviço de armazenamento e preservação de informações, através da captação das imagens dos documentos por processo fotográfico. A microfilmagem é mais indicada para garantir a preservação com qualidade arquivística. (CABRAL, 2002).

As vantagens da microfilmagem são os custos previsíveis com baixo custo de produção e de manutenção, o baixo custo do equipamento necessário à leitura e a sua durabilidade.

Para o GED, não basta que a imagem seja digitalizada e armazenada no computador, é preciso que o arquivo seja colocado em banco de dados e que o usuário possa localizar facilmente informações contidas no documento.

Uma das principais soluções de GED, nesse sentido, é o *Document Imaging*. De acordo com Baldam, Valle e Cavalcanti (2004, p. 47), os principais objetivos de um sistema de DI (*Document Imaging*) são:

- Capturar documentos em formato eletrônico;
- Armazená-lo em ambiente seguro;
- Recuperar esses documentos quando necessários;
- Permitir manipular esses documentos de acordo com os processos de negócios do setor/empresa em questão.

Os sistemas de DI podem variar conforme as necessidades de cada organização, mas é possível citar algumas características comuns para este tipo de

aplicação. Baldam, Valle e Cavalcanti (2004) citam que as mais relevantes são a interface de pesquisa, visualização e indexação, banco de dados para armazenar os índices, visualizador de imagens, recurso de customização, recurso de programação e segurança.

- **Interface de pesquisa** – deve possuir meios de facilitar para o usuário a pesquisa aos documentos. Sendo assim, a interface de pesquisa deve conter campos de auxílio à pesquisa por meio de índices, áreas para apresentar os resultados das pesquisas e deve conter comandos que permitam indexação dos documentos, descrevendo e identificando de acordo com o seu assunto.
- **Banco de dados** – é o lugar onde vai ser armazenado o índice de pesquisa. Um índice é o resultado do processo de indexação, que proporciona um resumo do documento que foi digitalizado. Um sistema GED necessita de um bom sistema de índice, para criar um modo organizado dos arquivos eletrônicos, possibilitando a pesquisa futura de maneira fácil e simples.
- **Visualizador de imagens** - muito importante para uma aplicação de DI. Os visualizadores mais comuns são aqueles que permitem abrir formatos de imagem digitalizada e possuem funções como: zoom, rotação, marca-texto, deslocamento de imagem, etc. Podem ser utilizados visualizadores externos, no case de arquivos CAD ou podem ser utilizados visualizadores do próprio fabricante, como no caso de arquivos PDF.
- **Recursos de Customização** – É importante que a ferramenta escolhida tenha uma interface para customização, permitindo que sejam feitas pequenas modificações e se adequando as reais necessidades do cliente.
- **Recursos de Programação** – Como é difícil um desenvolvedor prever toda a gama de necessidades do cliente, é interessante que a ferramenta seja programável.
- **Segurança** – Como os sistemas de DI são muito utilizados para arquivar grande variedade de documentos, é essencial que o sistema

possua regras de acesso. O administrador do sistema deve determinar quais diretórios e documentos poderão ser acessados pelos usuários.

Para capturar os documentos não físicos, podem ser utilizados diversos dispositivos, como os escâneres, câmera fotográficas e fax.

Para Baldam, Valle e Cavalcanti (2004), os dispositivos mais comuns que permitem transformar a imagem impressa no papel em formato digital são os escâneres.

Vários são os modelos de escâneres disponíveis no mercado. Escâneres para mercado doméstico, escâneres para cheques, documentação técnica, radiografias e escâneres 3D.

Quando um documento é digitalizado, ele se torna apenas uma cópia do documento em papel. Para que seja possível aferir a autoria, origem e autenticidade de um documento digital é necessário a assinatura digital. "Em situações onde não há confiança plena entre emissor e receptor, algo mais do que a autenticação é necessária. A solução mais atraente para este problema é a assinatura digital (STALLINGS, 2010). Essa tecnologia permite a segurança para o envio, recebimento e manuseio do documento eletrônico. Segundo o mesmo autor, a assinatura digital deve ter as seguintes propriedades: deve verificar o autor, data e hora da assinatura; deve autenticar os conteúdos no momento da assinatura e deve ser verificável por terceiros, para resolver disputas.

2.3.4 Sistema de Reconhecimento de Caracteres.

Os documentos provenientes da digitalização são normalmente uma imagem do documento original em papel. Eles não podem ser processados por sistema comum de processamento de dados. Para que esse documento digitalizado possa ser processado é preciso que ele passe por um processo de reconhecimento de caracteres. Dessa forma, o reconhecimento de padrões permite que sejam identificados em uma imagem digitalizada, caracteres, números, código de barra, etc. Depois de identificados esses caracteres, os dados podem ser processados de forma convencional por um computador, podendo ser editado por algum programa de texto. (BALDAM, VALLE e CAVALCANTI, 2004).

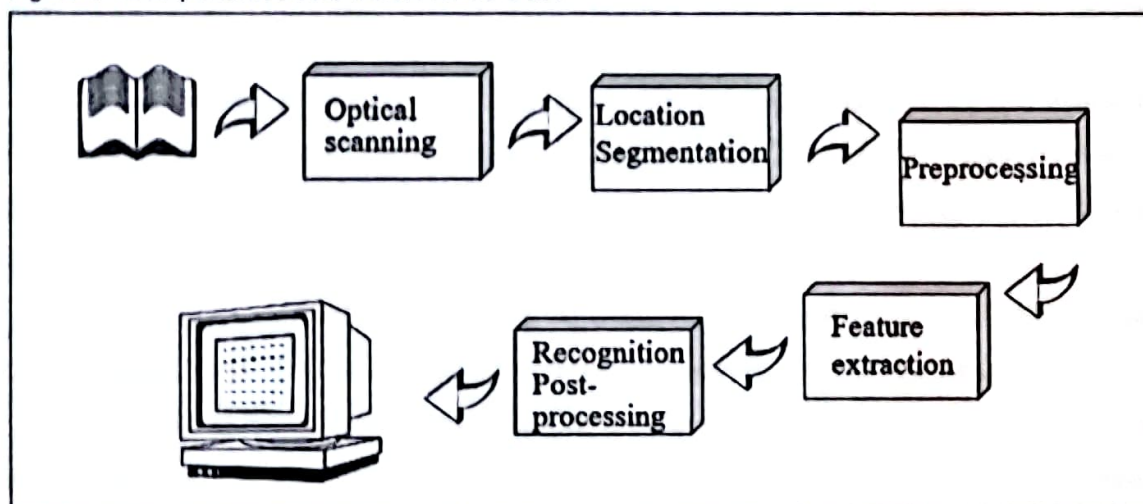
O objeto do reconhecimento de padrões é detectar, utilizando dispositivos e algoritmos (códigos de programação), elementos que possam ser comparados e feitas relações entre eles.

Os sistemas mais utilizados para reconhecimento de caracteres são: OCR, *Optical Character Recognition* (Reconhecimento Óptico de Caracteres) e o ICR, *Intelligent Character Recognition* (Reconhecimento Inteligente de Caracteres).

O OCR é uma tecnologia que permite fazer o reconhecimento de caracteres a partir de um arquivo de imagem. Segundo Baldam, Valle e Cavalcanti (2004, p. 117) o OCR permite reconhecer caracteres impressos ou datilografados em dados processáveis pelo computador.

Segundo Eikvil (1993, p. 11), "um sistema de OCR é formado pelos seguintes passos: Escaneamento Óptico; Localização e Segmentação; Pré-processamento; Extração de Características; Classificação e Pós-processamento". O primeiro passo do processo é digitalizar o documento analógico. Depois o sistema localiza a parte que contém o texto e faz extração dos caracteres usando a segmentação. Os caracteres extraídos vão ser processados para eliminar o ruído. O caractere é identificado ao comparar suas características com símbolos já cadastrados previamente no sistema. Por último a informação obtida é usada para reconstruir as palavras e números do texto original. O processo pode ser visualizado na figura 4.

Figura 4 - Componentes de um sistema de OCR



Fonte: Eikvil (1993, p. 11)

O ICR, reconhecimento inteligente de caracteres, é um avanço da tecnologia de OCR, permitindo que fontes diferentes, estilos diferentes de escritas a mão sejam

aprendidas pelo computador durante o processo de processamento, para melhorar a precisão do reconhecimento.

Segundo J. Mantas (1986, p. 425), um sistema de reconhecimento de escrita pode ser enquadrado em uma das seguintes categorias:

1. **Reconhecimento de caracteres com fontes fixas:** é o reconhecimento de caracteres escritos com fontes de tipos gráficos específicos como Pica, Courier e etc.
2. **Reconhecimento *on-line*:** é o reconhecimento de caracteres escritos à mão, onde é levado em conta não só o traço do caractere, mas também o tempo e a pressão que o autor impõe no processo de escrita do mesmo.
3. **Reconhecimento de caracteres manuscritos:** é o reconhecimento de caracteres à mão, porém com letra de forma e não conectados.
4. **Reconhecimento de caracteres cursivos:** é o reconhecimento de caracteres manuscritos sem restrição, isto é, são cursivos e podem estar conectados.

Para realizar o reconhecimento de caracteres de um documento, são utilizadas implementações de diversos algoritmos, que são usualmente baseadas nas técnicas citada em seguida (BALDAM, VALLE E CAVALCANTI 2004):

- *Pattern Matching* (Comparação de Padrões);
- *Feature Extraction* (Extração de Característica);
- *Neural Network* (Redes Neurais).

O *Pattern Matching* (Comparação de Padrões) compara o caractere digitalizado com uma biblioteca de padrões disponíveis. Dessa forma, quanto maior a biblioteca, maior será a probabilidade de acertar o caractere que se deseja reconhecer.

O *Feature Extraction* (Extração de Característica) funciona com comparação de caracteres mas utiliza algumas características percebidas nos traços.

A *Neural Network* (Redes Neurais) monta uma estrutura que o *software* vai armazenar as palavras que não são reconhecidas para serem analisadas separadamente e depois elas são incorporadas ao *software*. Em outras palavras a utilização da técnica de Redes Neurais faz com o que o *software* "aprenda" com o passar do tempo.

Existem alguns métodos auxiliares de reconhecimento. Baldam, Valle e Cavalcanti (2004, p.119-120), citam 3 (três) métodos: "análise de contexto, análise estatística e o apoio de bases de dados".

O método análise de contexto faz uma avaliação utilizando não só a palavra que o caractere se encontra, mas também a sentença formulada. Em documentos que envolvem número ou sequências numéricas, na maioria das vezes, não é utilizado esse método.

O método análise estatístico leva em consideração a probabilidade de um caractere ser utilizado após o uso de outro caractere.

O apoio de bases de dados ocorre em forma de consulta a uma base de palavras existentes em um banco de dados, que possibilitam indicar se a palavra encontrada é coerente ou não, indicando uma palavra mais próxima caso não seja encontrada.

Dessa forma, observa-se que o banco de dados pode ter uma grande importância num sistema de reconhecimento de caracteres.

2.3.5 Vantagens

O GED é indicado para instituições que produzem um grande fluxo de informação constante e grande procura pelos usuários. Um sistema GED possibilita melhorias e ampliação do alcance da informação, como também do controle, acesso e recuperação da informação.

Segundo o CONARQ (2011), a aplicação do GED traz alguns benefícios para as empresas, tais como:

- A recuperação rápida de documentos, caso aconteça algum desastre;
- A redução de custos com papéis e impressão, porque um dos objetivos do GED é que todo o ciclo do documento seja feito no formato eletrônico;
- Gerenciamento automatizado de processos, minimizando o gasto de recursos humanos;
- Melhor atendimento ao cliente, pois proporciona respostas rápidas;
- Possibilidade de integração com outros sistemas;

- Redução de fraudes com documentos e outros.

Todas essas vantagens irão depender do sistema escolhido pela instituição, pois cada sistema tem suas peculiaridades e, por isso, deve ser feito para atender as especificações no contexto que a empresa está inserida.

2.3.6 Desvantagens

Há algumas desvantagens que devem ser levados em consideração ao se pensar em aplicar as tecnologias GED. A implantação de um sistema exige muitos recursos tanto financeiros como humanos, devido à necessidade de profissionais qualificados. Nesse sentido, Baldam, Valle e Cavalcanti (2004) afirmam que podem ocorrer problemas na implantação do GED, quando não são consultados profissionais especializados.

Outro problema que pode ocorrer é necessidade de duplicidade de documentação gerada pelos casos de os documentos eletrônicos não serem aceitos como autênticos, como é o caso de documentos com valores probatórios ou jurídicos.

A mudança constante de tecnologia e o surgimento de novos formatos de arquivos digitais também podem ser considerados uma desvantagem para o sistema GED. Nesse sentido, Koch (2005) afirma que o mais complicado na aplicação GED é lidar com obsolescência tecnológica, o que é inevitável, visto que os meios físicos para guardar a informações duram muito tempo, mas os *softwares* de acesso duram poucos anos e as mudanças dos meios físicos são inevitáveis.

2.3.7 Sistema Informatizado de Gestão de Arquivos de Documentos

Segundo CONARQ (2011), o Sistema Informatizado de Gestão de Arquivos – SIGAD é um conjunto de procedimentos e operações técnicas que visam o controle do ciclo de vida dos documentos, desde a produção até a destinação final, seguindo os princípios da gestão arquivística de documentos e apoiado em um sistema informatizado.

O que diferencia um SIGAD de um GED é o trato dado ao documento arquivístico, garantindo confiabilidade, autenticidade e o acesso, ao longo do tempo.

O SIGAD é aplicável em sistemas que utilizam documentos digitais e não digitais.

Um SIGAD inclui operações como: captura de documentos, aplicação do plano de classificação, controle de versões, controle sobre os prazos de guarda e destinação, armazenamento seguro e procedimentos que garantam o acesso e a preservação a médio e longo prazo de documentos arquivísticos digitais e não digitais confiáveis e autênticos.

2.4 Banco de Dados

Dado pode ser definido como uma representação simbólica, quantificada ou quantificável. Assim um texto é um dado, pois as nossas letras latinas formam um sistema numérico discreto e, portanto, quantificado (SETZER e SILVA, 2005).

Um banco de dados pode ser definido como um conjunto de dados relacionados armazenados juntos, sem redundâncias nocivas ou desnecessárias, que modela um sistema do mundo real e pode servir para diferentes aplicações (CANTU, 2005).

Um sistema de banco de dados é um sistema de armazenamento de dados baseado em computador; isto é, um sistema cujo objetivo global é registrar e manter informação (DATE, 1983).

Um sistema de banco de dados é projetado para armazenar grandes volumes de informações. Além disso, o sistema de banco de dados precisa proporcionar segurança ao armazenamento de informações diante de falhas do sistema e acesso não autorizado.

Para Silberschatz, Korth e Sudarshan (2004, p. 1), "o principal objetivo de um sistema de banco de dados é proporcionar aos usuários uma visão abstrata dos dados, isto é, o sistema esconde determinado detalhes de como o dado é mantido e armazenado".

Os principais componentes de um sistema de banco de dados são: o *hardware*, o *software* e os usuários. Os dados devem estar integrados, ou seja, deve permitir que vários usuários possam acessar os mesmos dados.

O *hardware* consiste no volume de memórias – discos, tambores, etc. – nos quais reside o banco de dados.

O *software* é o que vai permitir ao usuário fazer solicitações ao banco de dados.

Para Date (1983, p. 29), são considerados três grandes classes de usuários: “o programador de aplicação, o administrador de banco de dados e o usuário final”.

O programador de aplicação é o responsável por escrever os códigos utilizando as diversas linguagem de programação disponível.

O administrador de banco de dados é o responsável pelo controle centralizado do banco de dados, e também por criar o banco de dados original e por fornecer autorização de acesso ao sistema.

O usuário final é o aquele que tem acesso ao banco de dados a partir de um terminal, fazendo consultas e interagindo com o banco de dados.

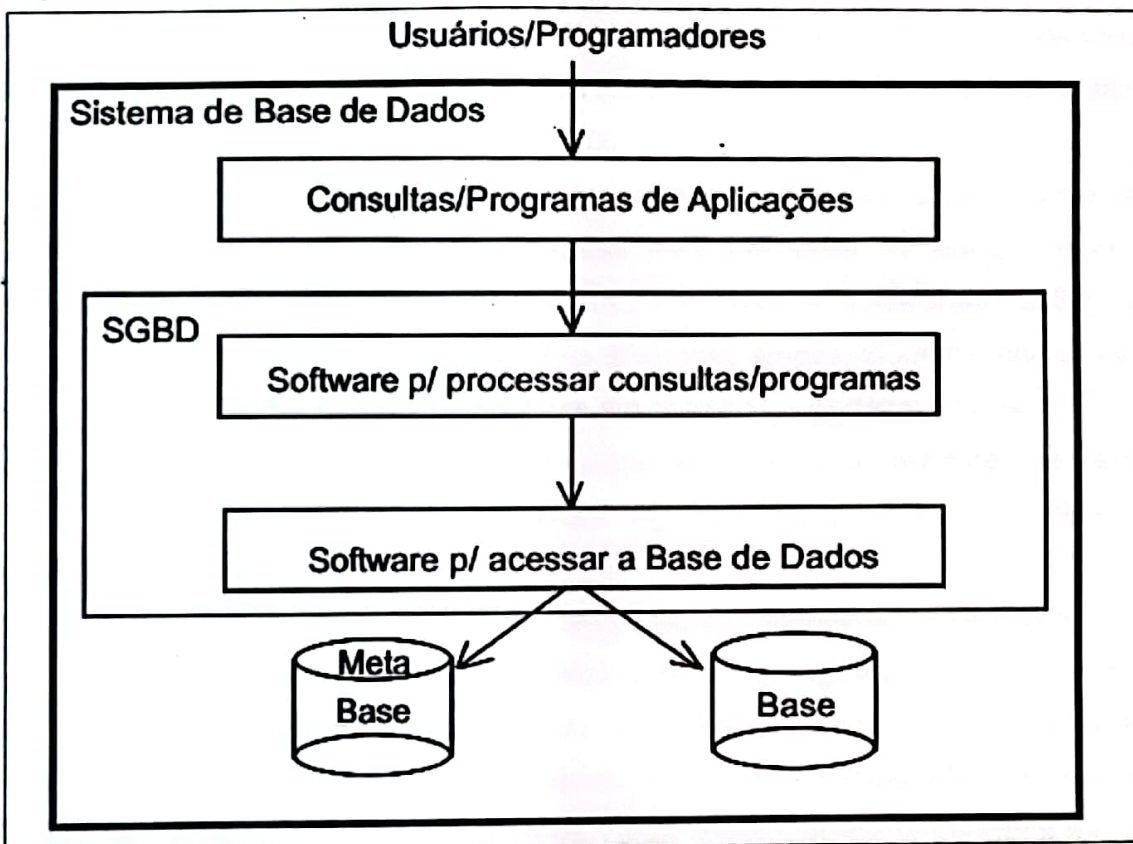
Segundo Navathe (2005), as principais características de um banco de dados são:

- **Controle de redundância.** A redundância é o armazenamento de uma informação em mais de um local no banco de dados, provocando inconsistência.
- **Compartilhamento e dados.** O sistema de banco de dados deve possuir um software que garanta a leitura e escrita de dados sem falha, mesmo em caso de concorrência de acesso.
- **Controle de acesso.** O SGBD deve possuir recursos de controle segurança e controle de autorização. O controle de autorização vai garantir quais dados o usuário pode ou não acessar. O controle de segurança são medidas que visam resguardar o banco de possíveis perdas ou destruição de dados em caso de falhas do software ou equipamento.
- **Interface.** Um banco de dados deve possuir uma interface que possibilite o acesso via menus de acesso, para facilitar o acesso.
- **Esquematização.** Um bando de dados deve fornecer mecanismos que possibilitem a compreensão dos relacionamentos existentes entre suas tabelas.

- **Controle de integridade.** O SGBD deve impedir que as aplicações pelas interfaces possam comprometer a integridade dos dados.
- **Backups.** O SGBD deve possuir mecanismos que possibilitem a recuperação de dados perdidos por falhas no equipamento ou software.

A base de dados e o software de gerenciamento da base de dados compõem o chamado sistema de base de dados. A figura 5 apresenta um esquema genérico de um sistema de banco de dados em sua interação com seus usuários (DATE, 1983).

Figura 5 - Interação em sistema de banco de dados



Fonte: Date, 1983.

2.4.1 Sistema de Gerenciamento de Banco de Dados - SGBD

Como o banco de dados trabalha com grandes volumes de dados, é preciso de um sistema que gerencie essas informações. Nesse sentido o gerenciamento de informações implica a definição das estruturas de armazenamento das informações e

a definição dos mecanismos para a manipulação dessas informações (SILBERSCHATZ, KORTH E SUDASRSHAN, 2004).

Dessa forma, para o gerenciamento de todas essas informações, é necessário o SGDB – Sistema Gerenciador de Banco de Dados. O banco de dados é criado e mantido pelo SGDB. Para Elmasri e Navathe (2002, p. 4), SGDB “é um sistema de *software* de propósito geral que facilita os processos de definição, construção, manipulação e compartilhamento de dados entre vários usuários e aplicações”.

O SGBD passou a utilizar diferentes formas de representação, ou modelos de dados, para descrever a estrutura das informações contidas em seus bancos de dados. Atualmente os seguintes modelos de dados são normalmente utilizados pelos SGBD's: modelo hierárquico, modelo em redes, modelo relacional e modelo orientado a objeto (TAKAI, ITALIANO E FERREIRA, 2005).

Sob a estrutura do banco de dados está o modelo de dados, que é um conjunto de ferramentas conceituais usadas para descrever os dados, fazer o relacionamento entre eles, etc. Para Silberschatz, Korth e Sudarshan (2004), os modelos de dados são classificados em três diferentes grupos: modelos lógicos com base em objetos, modelos lógicos com base em registros e modelos físicos.

Os modelos lógicos com base em objeto são usados na descrição de dados no nível lógico e de visões. Sua principal característica é permitir recursos de estruturação mais flexíveis.

Os modelos lógicos com base em registros, ao contrário do modelo com base em objetos, são usados tanto para especificar a estrutura lógica do banco de dados como para implementar uma descrição de alto nível. Os modelos com base em registro são assim chamados porque o banco de dados é estruturado por meio de registros de formato fixo para todos os tipos. Cada registro define um número fixo de campos ou atributos, e cada campo possui normalmente um tamanho fixo. (SILBERSCHATZ, KORTH E SUDASRSHAN, 2004).

Os modelos físicos de dados são usados para descrever os dados no nível mais baixo. São os modelos menos utilizados.

2.5 Sistema Distribuídos

Os primeiros sistemas de computadores eram centralizados. Um computador era responsável pelo gerenciamento e distribuição dos recursos de uma rede.

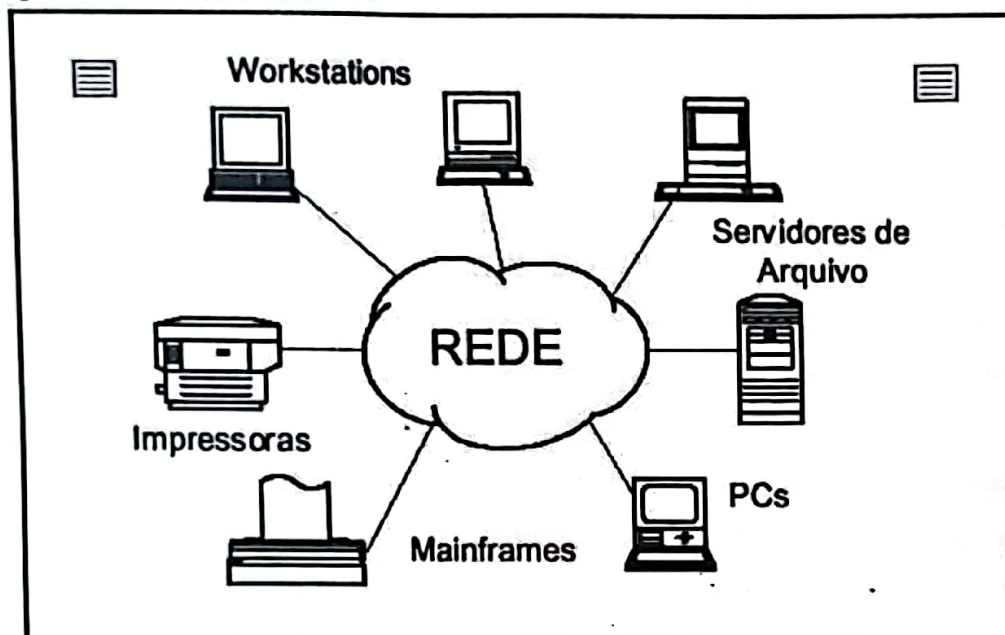
A evolução das redes de computadores e a popularização das máquinas possibilitaram a redução dos custos de comunicação entre computadores, gerando um aumento de usuários necessitando de serviços. Isso fez que com que crescesse a necessidade de interligar os computadores para compartilhar os recursos entre esses usuários. Assim começaram a surgir os sistemas distribuídos.

Atualmente, os sistemas distribuídos têm sido muito utilizados nos sistemas de computação em geral. Isso ocorre devido à grande difusão do uso de redes de computadores.

Um sistema distribuído é aquele no qual os componentes que estão interligados em rede, se comunicam e coordenam suas ações apenas passando mensagens entre si (COULOURIS, DOLLIMORE E KINDBERG, 2007).

O *software* de um sistema distribuído permite que os computadores coordenem suas atividades e compartilhem recursos do sistema como o *hardware*, *software* e os dados. Vale ressaltar a importância que o sistema distribuído seja projetado de maneira que possa ser percebido como um sistema único integrado. Nesse sentido, Tanenbaum e Van Steen (2008) consideram que sistema distribuído é o resultado da integração de sistemas computacionais autônomos combinados de forma a atingir um objetivo em comum, ou seja, sistemas que compartilham recursos físicos, recursos de software, com o objetivo de realizar alguma tarefa. A figura abaixo ilustra um sistema distribuído.

Figura 6 - Sistemas Distribuídos



Fonte: Tanenbaum e van Steen (2008)

Para suportar vários computadores e redes diferentes e, simultaneamente, oferecer uma visão de sistema única, os sistemas distribuídos costumam ser organizados por meio de uma camada de *software* composta por usuários e aplicações, no nível mais alto, e uma camada de *software* no nível mais baixo, formado pelos sistemas operacionais e facilidades de comunicação. (TANENBAUM; VAN STEEN, 2008). Por isso, o sistema distribuído, às vezes, é denominado de *middleware*.

Dessa forma, um sistema distribuído deverá conter componentes que são distribuídos em vários computadores. Para Emmerich (2000), cada computador que hospeda esses componentes é chamado de *host*.

Quando esses computadores são ligados por meio de uma rede num sistema distribuído, podem ocorrer algumas consequências (COULOURIS, DOLLIMORE E KINDBERG, 2007):

- **Concorrência:** a execução de programas de forma concorrente num sistema distribuído é a regra. Portanto, sempre será avaliada a capacidade do sistema em lidar recursos compartilhados
- **Inexistência de um relógio global:** a cooperação de recursos num sistema distribuído é feito por trocas de mensagens e como é difícil

estabelecer uma hora precisa para todos os computadores, pode gerar consequências para o sistema.

- Falhas independentes: o sistema deve garantir que quando um componente de um sistema falhe independentemente, os outros componentes continuem funcionando.

Para a construção de um sistema distribuído, Tanenbaum e van Steen (2008) apontam quatro metas importantes que devem ser cumpridas. O sistema distribuído deve oferecer acesso fácil aos seus recursos; deve ser transparente; deve possuir escalabilidade.

- O compartilhamento de recursos é a principal motivação para a construção de um sistema distribuído. Os recursos podem ser gerenciados por servidores e acessados por clientes. Os recursos podem ser discos, impressoras, arquivos, banco de dados e outros. Há várias razões para querer compartilhar recursos, uma delas é a economia.
- A transparência significa que o sistema deve ocultar razoavelmente o fato de que os recursos são distribuídos por uma rede. Ele deve se apresentar para o usuário como um sistema único.
- A abertura do sistema é um sistema que oferece serviços de acordo com regras padronizadas que descrevem sintaxe e semântica desses serviços. É mostrar, por meio de uma linguagem natural, o que realmente fazer os serviços que estão sendo oferecidos pela linguagem da interface, IDL (*interface definition language*).
- A escalabilidade é a possibilidade de um sistema adicionar usuários e recursos ao sistema. Significa ainda a possibilidade dos usuários e recursos estarem longes um dos outros e que seja fácil de gerenciar, mesmo que abranja muitos departamentos.

Para serem compartilhados, os recursos precisam de um *software* que ofereça uma interface de comunicação, indicando o recurso a ser acessado, manuseado e atualizado com segurança. Esse *software* é chamado de gerente de recurso e é responsável por um conjunto de métodos para acesso aos outros recursos do sistema.

Dessa forma, o usuário acessa os recursos utilizando o gerenciador de recurso e esse acesso pode ser dar de duas formas ou modelos: o modelo cliente-servidor e o modelo baseado em objeto.

No modelo cliente servidor, as tarefas e as cargas de trabalho são distribuídas entre os fornecedores de um recurso ou serviço, designados de servidores, e os que solicitam os serviços, definidos como clientes. O servidor é uma máquina que está executando um ou mais serviços que compartilham recursos com os clientes. Um cliente não compartilha seus recursos, mas solicita um conteúdo ou função do servidor.

Figura 7 - Modelo de Arquitetura Cliente-Servidor



Fonte: disponível em <https://www.emaze.com/Modelo-cliente-servidor.pptx>

No modelo baseado em objeto, cada recurso compartilhado é visto como um objeto, que deve ser identificado de forma única e deve transitar livre pela rede sem perder sua identidade. Quando um programa requer um serviço de algum recurso compartilhado, ele envia uma mensagem ao objeto correspondente.

A principal vantagem de um sistema distribuído em relação a um sistema centralizado é o baixo custo, isso porque o valor para implementação de um sistema distribuído geralmente é muito inferior ao valor de um sistema centralizado com um poder de processamento equivalente (TANENBAUM; VAN STEEN, 2008).

Outra vantagem de um sistema distribuído é a confiabilidade. Um sistema distribuído pode contar com algoritmos que detectam falhas e redirecionam as tarefas para uma outra máquina.

Outro fator de vantagem é a flexibilidade. Os recursos disponibilizados por uma máquina podem ser remanejados para outra, mantendo assim uma redundância de recursos e maior disponibilidade dos serviços.

Entretanto, há algumas desvantagens na utilização de um sistema distribuído, como a segurança. É muito difícil estabelecer um protocolo de segurança eficiente, porque todas as máquinas possuem acessos entre elas, então deve ser feita implementações que garantam a integridade dos dados.

Outro fator que pode ser considerado uma desvantagem é a viabilidade. Nem todos as tarefas que são executadas num modelo centralizado serão executadas da mesma forma em um sistema distribuído. É preciso que seja verificado se a tarefa pode ser fragmentada e distribuída no sistema.

2.6 Metodologia

Etimologicamente, a palavra método deriva do latim *methodus* e cujo significado é o "caminho ou via para realizar algo". Marconi e Lakatos (2003, p. 83) define método como o "conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo – conhecimento válidos e verdadeiros –, trançando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista".

Dessa forma, o método deve ser entendido numa perspectiva mais ampla, como sendo o caminho escolhido para chegar a um fim. Já o método científico pode ser definido como um conjunto de procedimentos intelectuais e técnicos adotados para se atingir o conhecimento (GIL, 2008).

Para o presente estudo foi utilizado o método dedutivo. Para Santos (2013) o método dedutivo tem suas proposições enfocadas na situação geral para explicar as particularidades.

O conhecimento humano caracteriza-se pela relação estabelecida entre o sujeito e o objeto, podendo-se dizer que esta é uma relação de apropriação. A complexidade do objeto a ser conhecido determina o nível de abrangência da apropriação. Assim, a apreensão simples da realidade cotidiana é um conhecimento popular ou empírico, enquanto o estudo aprofundado e metódico da realidade enquadra-se no conhecimento científico. O questionamento do mundo e do homem quanto à origem, liberdade ou destino, remete ao conhecimento filosófico.

Segundo Gil (2010, p. 17), pesquisa pode ser definida como:

(...) procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos. A pesquisa desenvolve-se por um processo constituído de várias fases, desde a formulação do problema até a apresentação e discussão dos resultados. Só se inicia uma pesquisa se existir uma pergunta, uma dúvida para a qual se quer buscar a resposta.

Pesquisar, portanto, é buscar ou procurar resposta para alguma coisa. As razões que levam à realização de uma pesquisa científica podem ser agrupadas em razões intelectuais (desejo de conhecer pela própria satisfação de conhecer) e razões práticas (desejo de conhecer com vistas a fazer algo de maneira mais eficaz). Para se fazer uma pesquisa científica, não basta o desejo do pesquisador em realizá-la; é fundamental ter o conhecimento do assunto a ser pesquisado.

A pesquisa é a atividade nuclear da Ciência. Ela possibilita uma aproximação e um entendimento da realidade a investigar. A pesquisa é um processo permanentemente inacabado. Processa-se por meio de aproximações sucessivas da realidade, fornecendo-nos subsídios para uma intervenção no real.

Para Severino (2014), a primeira diferenciação que se pode fazer num trabalho científico é aquela entre pesquisa quantitativa e pesquisa qualitativa. Segundo Richardson (2012) o método quantitativo caracteriza-se pelo emprego quantificado tanto nas modalidades de coleta de informações, quanto no tratamento delas por meio de técnicas estatísticas. Já o método qualitativo é mais adequado para entender a natureza de um fenômeno social. Dessa forma, optou-se pela pesquisa qualitativa, analisando o questionário respondido pelos policiais que atuam no local e pela descrição do ambiente do setor de comunicação.

De acordo com Santos (2013), a pesquisa pode ser classificada de duas maneiras: a primeira, com base nos objetivos pretendidos e a segunda se baseia nos procedimentos técnicos utilizados pelo pesquisador.

Quanto aos objetivos, esta pesquisa pode ser classificada como descritiva-explicativa. Observando somente o aspecto descritivo, este estudo não atenderia ao objetivo proposto nesta pesquisa, pois somente teria preocupação com as características do fenômeno. Dessa forma, unindo os dois tipos de pesquisas torna mais possível a obtenção do resultado.

A pesquisa descritiva tem como objetivo principal a descrição das características de determinada população ou fenômeno, utilizando de técnicas como coleta de dados, entrevistas, questionários e outros. Neste tipo de pesquisa, realizam-se o estudo, a análise, o registro e a interpretação dos fatos do mundo físico sem a

interferência do pesquisador. A finalidade é observar, registrar e analisar os fenômenos ou sistemas técnicos (GIL, 2008).

Já a pesquisa explicativa é aquela que tem como preocupação central, identificar os fatores que determinam ou que contribuem para a ocorrência dos fenômenos (GIL, 2008). Esse tipo de pesquisa registra fatos, analisa-os, interpreta-os e identifica suas causas. Essa prática visa definir modelos e estruturas e gerar hipóteses por força de dedução lógica.

Quanto aos procedimentos técnicos, esta pesquisa classifica-se como pesquisa bibliográfica e pesquisa de campo.

Segundo Gil (2008) a pesquisa bibliográfica é desenvolvida a partir de material já elaborado, constituído principalmente de livros e artigos científicos. A principal vantagem da pesquisa bibliográfica é permitir ao pesquisador a cobertura de uma gama de fenômenos mais ampla do que aquela que poderia pesquisa diretamente.

A pesquisa de campo é aquela utilizada com o objetivo de conseguir informações e/ou conhecimentos acerca de um problema, para qual se procura uma resposta. Consiste na observação de fatos ou fenômenos tal como ocorrem espontaneamente (MARKONI e LACATOS, 2003). A importância da pesquisa de campo neste estudo está voltada para a análise das características dos fatos e fenômenos que norteiam a atividade do Centro de Comunicação Social da PMDF.

Há várias técnicas e métodos de coleta de dados que podem ser empregados a fim de obter informações sobre o objeto em estudo. Nesta pesquisa foi utilizada a técnica de questionário para entrevistar a equipe do Centro de Comunicação Social. Segundo Marconi e Lakatos (2003), o questionário é um instrumento de coleta de dados constituído por uma série de perguntas que devem ser respondidas sem a presença do entrevistador.

2.6.2 Instrumentos de Pesquisa

Foram utilizados nesta pesquisa a entrevista semiestruturada, que buscou informações sobre o gerenciamento das informações no CCS. A partir da entrevista e das conversas informais pretendeu-se estabelecer o atual funcionamento do setor, baseado no referencial teórico e na percepção dos agentes que lá trabalham na

atualidade propondo ações que possam melhorar o gerenciamento e acesso à informação.

Segundo Santos (2013), a entrevista é um excelente instrumento de pesquisa e é largamente usada no mundo das organizações, com múltiplas finalidades. A entrevistas visa buscar informações que tragam a complexidade de ação e funcionamento do objeto.

Neste caso, o intuito da entrevista foi analisar o funcionamento do CCS e as ações por seus integrantes, abrindo oportunidade para o informante delinear tanto o potencial do setor, quanto as ações necessárias para um melhor funcionamento.

Outro instrumento muito importante nesta pesquisa foi o levantamento do funcionamento, localização e condições de trabalho da secretaria de comunicação PMDF. A partir de um roteiro buscou se o levantamento de informações sobre o CCS, seu ambiente físico e funcionamento.

Os itens do roteiro seguem:

- 1- Condições físicas;
- 2- Equipamentos;
- 3- Pessoal;
- 4- Como se dá a recepção da informação até sua disseminação;
- 5- Cadeia de comando e controle.

A eficácia da entrevista passa pela competência e preparo do entrevistador. Cada entrevista é um caso particular. Elas variam de acordo com o objetivo e, em qualquer situação, o entrevistador necessita ter bastante compreensão da natureza humana.

Roteiro ou formulário é usado quando o pesquisador pretende do informante respostas mais amplas e com maiores detalhes. É usado como roteiro para as entrevistas e se aproxima muito do questionário aberto. Para este e os demais instrumentos da pesquisa, deve preceder a elaboração, a organização de um plano para ordenar as perguntas e dar sequencia lógica ao conjunto de elementos a serem coletados. O formulário, como os demais instrumentos deve ter um início, um meio e um fim, interligados e interagidos (SANTOS, 2013 p. 260)

Outro instrumento que deve contemplar todas as pesquisas é a pesquisa exploratória. Neste trabalho, a pesquisa exploratória, através de visita a sessão e conversa informal com os policiais que atuam neste departamento, oportunizou tanto definir o objeto de pesquisa quanto a metodologia.

2.7 Análise dos Dados e Proposta de GED

Conforme os objetivos pretendidos com esta pesquisa, as perguntas dos questionários foram divididas em vários temas como: conhecimentos existentes, procedimentos administrativos em relação ao arquivo, utilização de ferramentas eletrônicas para gerenciamento da informação, dentre outras. As perguntas foram colocadas no questionário de forma que os temas estivessem diversificados, para evitar que um tema de uma pergunta determinasse a resposta da outra pergunta.

A coleta das informações dos questionários se deu no período de 16 a 23 de novembro de 2015, e todos respondidos pela seção do Centro de Comunicação Social da PMDF, que se localiza no Quartel do Comando-Geral e ocupa aproximadamente metade do primeiro pavimento do prédio.

Verificou-se pela entrevista que os equipamentos de imagem, impressoras, computadores não são adequados para o trabalho de gerenciamento eletrônico da informação. Os entrevistados afirmaram que o ideal para implantação de um sistema eletrônico seria, primeiramente, o investimento em tecnologia, a partir de sistemas que comportem o trabalho com imagens e equipamentos que façam a impressão com a qualidade necessária.

Verificou-se ainda que há momentos que o operador se utiliza de recursos próprios para o desempenho de sua função, tendo que adquirir insumos (máquina fotográfica, gravador de áudio, filmadora) ou seja, equipamentos a fim de melhorar as condições de trabalho e, por conseguinte, no gerenciamento das informações.

Foi descrito também a existência de uma equipe técnica especializada, como jornalistas e pessoas capacitadas para o trabalho de assessoria de imprensa, recebimento e gerenciamento da informação. Entretanto, durante a pesquisa verificou-se a necessidade de capacitação visando o conhecimento e experiência de mercado, como agências de publicidade, jornais impressos ou televisivos. O que ampliaria a ação do CCS e otimizaria o acesso a informação por outros meios, haja vista que no momento o principal canal de comunicação é o site: <<http://www.pmdf.df.gov.br/site/>>

Foi aferido que a divulgação da notícia se dá, atualmente, pela Divisão de Assessoria de Imprensa (DAI). Eles recebem as informações das unidades ou equipes de policiais de serviço, produzem as matérias e inserem no site.

Há, portanto, uma grande preocupação quanto à acessibilidade dos documentos por parte dos entrevistados. Como não há armazenamento das informações, a pesquisa por notícias publicadas tem que ser feita no próprio portal, o que dificulta a recuperação da informação.

A equipe cria *releases* e tem conhecimentos em computação gráfica associam sempre imagem e texto com a personalidade que a informação necessita e o compromisso com a cultura policial, valorizando as ações da corporação no âmbito do DF. Parcerias e atendimento ao público foram incorporadas ao site e informações de ocorrências e roubos a carro. A equipe prima pela atualização das informações embora seja limitado o atendimento em virtude da ausência de arquivamento e acesso rápido a informação.

Foram confirmados, ainda, pelas entrevistas, a existência de alguns problemas que são determinados pelo gerenciamento inadequado de documentos, ou seja, não há processos claros de elaboração, revisão e descarte.

A partir das informações coletadas e do estudo do referencial teórico, base desta pesquisa, observou-se a possibilidade de implantação de um sistema gerenciamento eletrônico da informação, utilizando compartilhamento de recursos por sistemas distribuídos.

3 PROPOSTA DE ARQUITETURA

Com o objetivo de validar a proposta descrita neste capítulo e de gerar parâmetros de avaliação para cada módulo apresentado, foi necessária uma proposta de arquitetura que demonstrasse algumas métricas de funcionamento da arquitetura sugerida. Além disso, foi realizado um estudo de caso através de um questionário para que pudesse verificar a necessidade da utilização de sistemas de GED e assim validar tal arquitetura e a proposta do trabalho e que assim fosse necessário que o mesmo consiga simular mecanismos de *OCR* que possa coletar informações de fichas de processos de trabalho.

Desta forma, será tratada neste capítulo, a arquitetura lógica com descrição dos principais módulos propostos para implementação de um serviço de *GED* com técnicas de *OCR* com os diagramas de casos de uso, bem como a implementação e alguns resultados encontrados.

Após a descrição da arquitetura lógica, será mostrada uma proposta de ambiente para o desenvolvimento de um protótipo, mostrando as ferramentas utilizadas. Além disso, podemos descrever um cenário de testes.

3.1 Representação da Arquitetura

Esta seção apresenta a estrutura de componentes da arquitetura proposta neste trabalho para o desenvolvimento de um *OCR*. A arquitetura proposta foi dividida em seis módulos distintos. Estes são interligados da seguinte maneira: um módulo de interface com o usuário (*GUI*), um módulo principal, um de pré-processamento, um de segmentação das imagens, um para extração de características e por fim, um classificador. A tabela 2 mostra todos os módulos com suas respectivas funções.

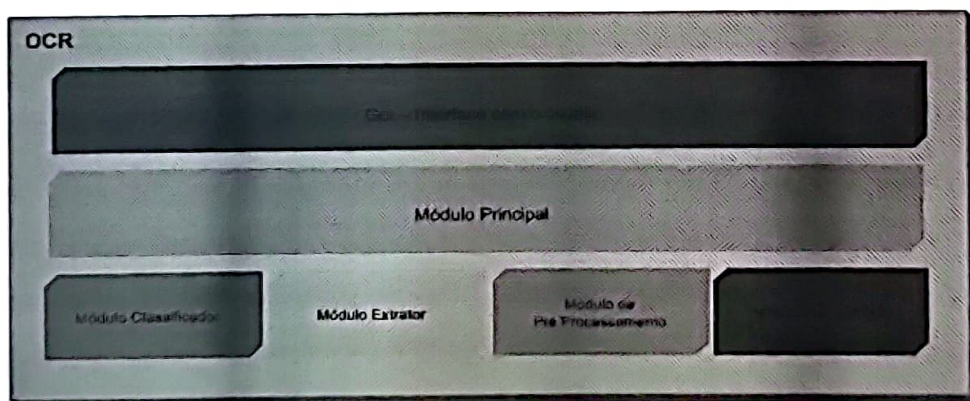
Tabela 2 - Descrição dos módulos apresentado na arquitetura

Nome	Função
Módulo de interface com o usuário – GUI	Apresenta a interface gráfica com usuário.
Módulo principal	Interação com os demais módulos realizando um papel de gerente.
Módulo de pré-processamento	Tratamento da imagem
Módulo de segmentação	Recebe imagens já processadas e realiza a segmentação em palavras.
Módulo de extração de características	Avalia e extrai características relevantes existentes de uma imagem.
Módulo de classificação	Treina a rede neural, classificando e apresentando os resultados para o módulo principal.

Fonte: (MENDONÇA, 2008)

A figura 8 trata-se de uma arquitetura em módulos que ilustra os módulos (subsistemas) de um sistema de OCR. Para a apresentação desse diagrama (definição da arquitetura) utilizou-se alguns padrões de engenharia de software definido como *Design Pattern* (padrão de projeto) *Facade* e *Singleton*. Sendo assim, cada módulo deve implementar uma classe *facade*, com a finalidade de simplificar o uso e facilitar o entendimento de cada módulo, além de reduzir a dependência entre eles. O objetivo de uso de padrões nesta arquitetura é fornecer modelos reusáveis para o desenvolvimento de um sistema de OCR. A interface que será apresentada por cada *facade* será explicada abaixo.

Figura 8 - Arquitetura em módulos para um sistema de OCR



Fonte: Mendonça (2008, p. 75)

3.1.1 Módulo de Interface Gráfica

Esse módulo tem o objetivo de promover uma interface gráfica com o usuário, apresentando resultados extraídos de outros módulos. Além disso, é através deste módulo que se tem a possibilidade de realizar algumas manipulações na imagem como: visualização da imagem, recortes, buscas, adicionar padrões de treinamento, para que possa assim classificar o documento no sistema de *GED*. Sendo assim o mesmo irá possuir como parâmetro de entrada o caminho das imagens a ser tratada e como saída a visualização da imagem selecionada.

Em relação à troca de informações existente nesse módulo, o mesmo se comunica somente com o módulo principal onde este se comunica com os demais módulos, que veremos com maiores detalhes adiante.

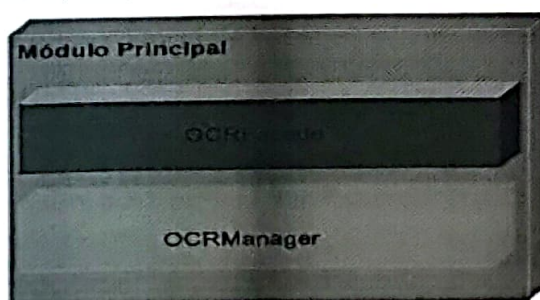
Os serviços de um OCR podem ser fornecidos aos usuários via um protocolo específico, como por exemplo: desktop ou até mesmo utilizando uma interface que implemente serviços web (Web Services).

3.1.2 Módulo principal

Este módulo tem como funcionalidade a interação dos mecanismos propostos, possuindo uma característica de gerência dos demais módulos, realizando um papel fundamental para a arquitetura apresentada. A comunicação com este módulo é feita através da interface *OcrFacade*.

As tarefas do módulo principal são executadas pela classe *OcrManager*, que por sua vez repassa as funções executadas para a interface *OcrFacade* que tem a função de interagir com os demais módulos para atender às diversas solicitações.

Figura 9 - Subdivisão do módulo principal



3.1.3 Módulo de pré-processamento

Este módulo tem como funcionalidade realizar tratamentos nas imagens. A comunicação com este módulo é feita através da interface `ProcessadorFacade`.

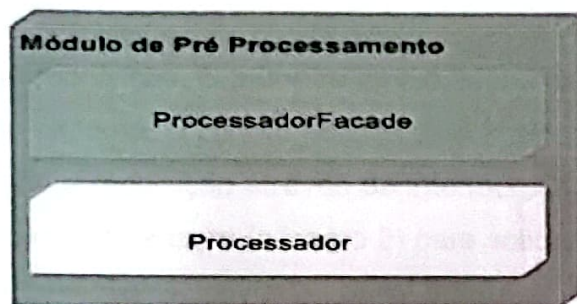
As tarefas do módulo de pré-processamento são executadas pelo processador, conforme demonstração a seguir.

A classe processador irá tratar as imagens realizando um trabalho de regularização. Sendo assim, o módulo de pré-processamento deve possuir funções para o tratamento das imagens, tais como: conversão da imagem para uma escala de cinza, retirar ruídos, realçar contorno e transformar a imagem em binário, com a finalidade de auxiliar na extração de características e o auxílio na segmentação.

- **Converter para escala de cinza:** Essa funcionalidade tem como objetivo converter uma imagem RGB para níveis de cinza, que possibilita uma análise estatística da distribuição de tons de cinza, que por sua vez pode ser utilizada para extração das características da imagem.
- **Retirar ruídos:** uma imagem digitalizada apresenta algumas irregularidades no qual chamamos de ruídos, que dificultam a classificação e o reconhecimento da imagem. Sendo assim essa funcionalidade tem por objetivo eliminar o máximo possível de ruídos existentes na imagem.
- **Realçar contorno:** Essa funcionalidade tem como objetivo realçar os contornos da palavra para que possa começar a identificar uma característica da escrita. Além disso, é importante a utilização de técnicas específicas para realização desse método.
- **Imagem binária:** Essa funcionalidade tem a função converter uma imagem para binário com objetivo facilitar a extração características nela existente.

Na figura 10 temos a representação da subdivisão do módulo de pré-processamento. Sendo assim, o módulo aqui apresentado tem como funcionalidade o tratamento das imagens, retirando o máximo de imperfeições encontradas nas mesmas e criando características para um melhor reconhecimento para que assim possa alimentar a base de dados do sistema GED.

Figura 10 - Subdivisão do módulo de pré-processamento

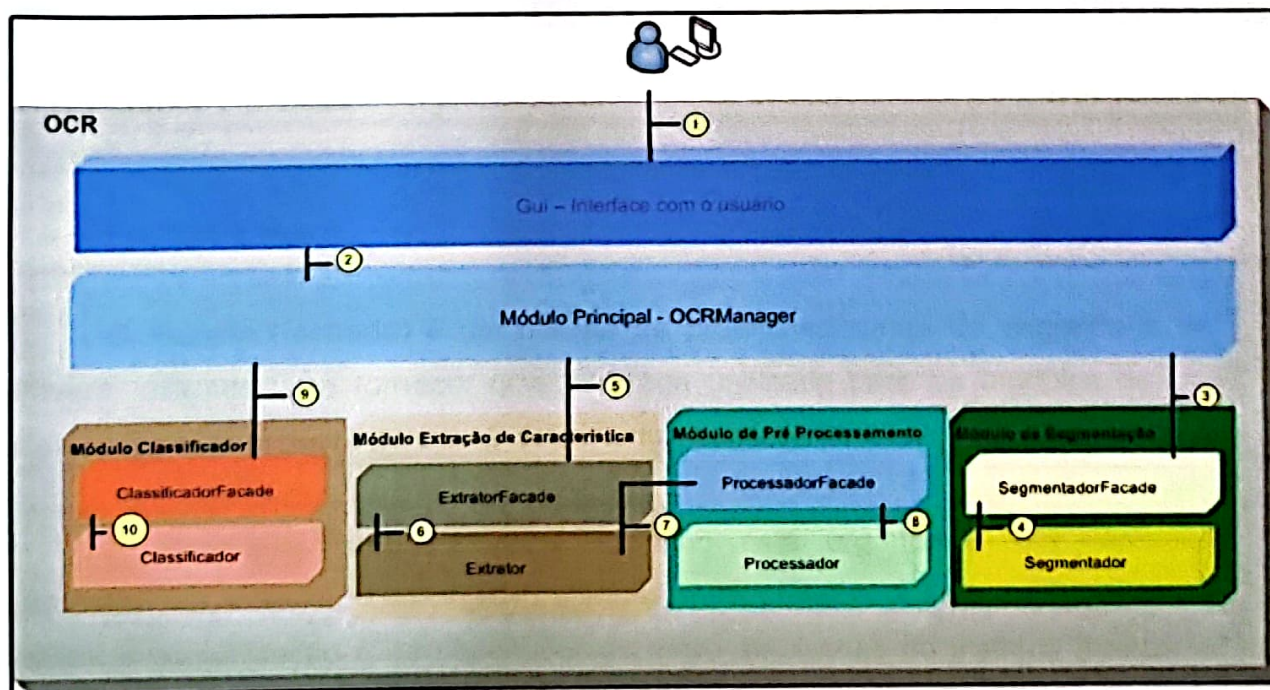


Fonte: Mendonça (2008)

3.1.4 Inter-Relacionamento entre os módulos

Esta seção apresenta os inter-relacionamentos entre os módulos componentes da arquitetura proposta. Os inter-relacionamentos entre os componentes da arquitetura evidenciam a funcionalidade do módulo principal onde mostra a comunicação entre os módulos do sistema que serão criados a partir desta arquitetura. A identificação destes inter-relacionamentos facilita o entendimento do comportamento do sistema gerado.

Figura 11 - Detalhamento dos inter-relacionamentos da arquitetura



Fonte: Mendonça (2008 p. 66)

Os exemplos da Figura 11 evidenciam os inter-relacionamentos entre os módulos da arquitetura de um OCR. Podem ser vistos os passos executados por um OCR para responder uma requisição. O primeiro passo, identificado pelo número 1, é a requisição do cliente ao sistema. O cliente solicita uma imagem, dada seu local ou sua URL. O sistema recebe a requisição através da interface GUI. Esta, por sua vez, comunica-se com o módulo OcrManager (número 2) para solicitar as requisições.

Para isso, a requisição é enviada para interface SegmentadorFacade (número 3). Este envia a para a classe (subsistema) segmentador (número 4) que realiza a segmentação da imagem e retorna o resultado para o OcrManager, que envia para módulo Extração de característica fazendo a comunicação com a interface ExtratorFacade (número 5) que envia para a classe (subsistema) extrator, responsável por extrair características na imagem (número 6). Que por sua vez envia para o módulo de pré-processamento junto a interface processadorFacade (número 7) que envia para classe processador (número 8) e retornar o resultado ao módulo principal.

Por fim, após a imagem já extraída, as características são enviadas para o módulo classificador (número 9), na qual as funções são recebidas pela interface ClassificadorFacade (número 10) que por sua vez repassa as funções para a classe classificador que é armazenada em uma base de dados de arquivos indexados para que assim possa ser utilizada um sistema de gerenciamento de documentos - GED.

3.1.5 Padrões utilizados

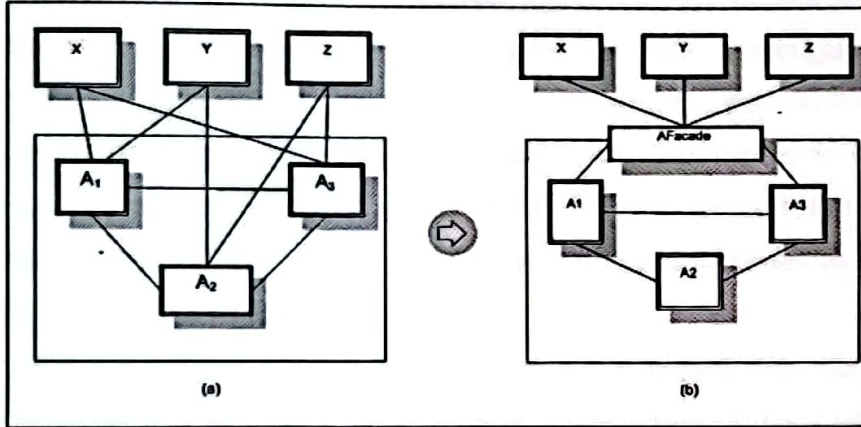
3.1.5.1 Padrões *Facade*

O *Facade* (fachada) é um padrão de projeto estrutural de engenharia de software, utilizado para fornecer uma interface unificada para os módulos de um sistema. O objetivo deste padrão é definir uma interface de alto nível para facilitar o uso das funcionalidades do módulo.

A arquitetura proposta neste trabalho está organizada em camadas compostas por módulos. Um objetivo comum entre projetos orientados a objetos é diminuir a comunicação e as dependências entre as classes do sistema (reduzir o acoplamento), de forma a facilitar a manutenção do sistema. Suponhamos a situação de alto acoplamento mostrada pela figura 12, onde no modelo (a) os módulos X, Y e

Z acessam diretamente as classes A1, A2 e A3 do módulo A. no modelo (b) é possível perceber a redução do acoplamento através do uso do padrão.

Figura 12 - Uso do padrão de projeto *Facade*



Fonte: Mendonça (2008)

3.1.5.2 Padrões *Singleton*

O padrão *Singleton* é um padrão de projeto de criação de objetos. O objetivo é garantir uma instância única e acessível de forma global e uniforme para toda classe que implemente este padrão. Neste contexto, este padrão foi usado para tornar a interface de cada módulo da arquitetura única e acessível de forma global.

Outra forma de garantir o acesso global é o uso de uma variável global. Porém a variável global não garante a unicidade. A solução proposta pelo padrão *Singleton* é tornar a própria classe responsável pela sua única instância. A classe que implementa o padrão *Singleton* garante o acesso à sua instância e ainda intercepta as requisições para criação de novos objetos, garantindo que nenhuma outra instância seja criada.

3.2 Detalhamento do objeto

Para avaliar a organização da arquitetura do sistema de GED foi elaborado um guia de serviços para que assim validasse o sistema. Onde a primeira parte será receber os documentos, Higienizar, Recuperar, Organizar, Classificar, Digitalizar, Registrar e Integrar Metadados dos documentos constantes, gerando arquivos em formato PNG (*Portable Network Graphics*) em cada sessão para o qual for emitida as

ordens de serviço, pois os documentos não poderão ser retirados das instalações da PMDF pela(s) Ordem(ns) de Serviço(s).

A Empresa será responsável pela digitalização e organização dos documentos elencados no objeto desse termo de referência, bem como, recursos materiais, recursos humanos e de logística para execução do processo de digitalização. Também deverá executar atividades de correção de eventuais equívocos detectados e resultantes da fiscalização, no momento do encerramento de cada grupo de pastas funcionais. Após 10 dias do início dos serviços na entidade beneficiada, a mesma deverá realizar a transferência de tecnologia para essa entidade.

Cabe esclarecer que, mesmo com a existência de equipamentos de digitalização dentro das unidades da PMDF, os equipamentos a serem utilizados para a realização dos serviços são de encargo da empresa prestadora desse e devem subsidiar também o processo de transferência de tecnologia.

Assim, cada grupo de pastas deverá receber uma identificação única durante todo o processo, e conterá atributos que facilitem sua identificação e recuperação, tais como número do romaneio, identificação do servidor, quantidade de documentos nas pastas encaminhadas nesse romaneio, responsáveis pela entrega e recepção, entre outros, garantindo assim a integridade do processo como um todo, e criando um primeiro nível de informação de controle, auditoria e fiscalização do processo.

O instrumento de medição será o número de páginas oriundas das folhas para o trabalho a ser realizado pela Empresa. Esse processo de contabilização dos serviços irá observar se, das folhas encaminhadas para formação do GED, quais destas contém informações em frente e verso, o que irá estabelecer o número de páginas e suas características. Contar-se-á uma página por folha simples e duas páginas por folha frente e verso que contenham escrita, devendo ser desconsiderada a página que tenha a expressão "EM BRANCO", carimbada ou em manuscrito, as com risco transversal bem como as páginas totalmente em branco.

4 DESENVOLVIMENTO DO PROTÓTIPO

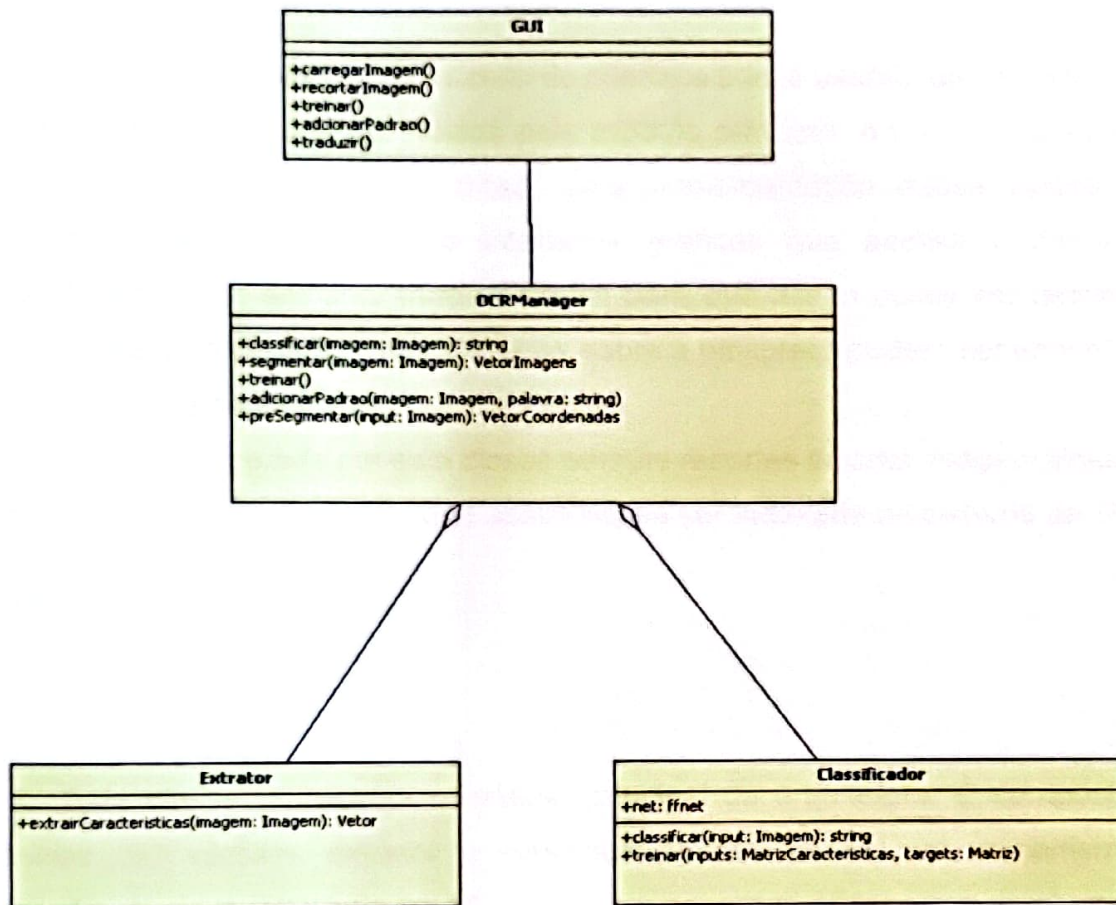
Para desenvolvimento do protótipo aqui proposto, seguiu-se o modelo de arquitetura apresentada no item 3.1, utilizando todas as ferramentas descritas no item 3.1.5, de maneira que o mesmo tem como característica simular o funcionamento de um OCR para arquivamento de documento em sistemas de GED.

A divisão da arquitetura em módulos deve-se as características individuais de cada um, ou seja, um módulo que verifica as imagens, um que extrai as caracterizas entre outros, sendo cada um para uma função específica. Entretanto para representação dessa arquitetura foi proposto um protótipo que apresenta as funcionalidades desde a classificação do documento até a parte de armazenamento do mesmo.

A seguir veremos o diagrama de classes proposto para um possível protótipo a ser utilizado pelo Centro de Comunicação Social da PMDF, identificando as funcionalidades de cada módulo desenvolvido e onde reside a sua inteligência. Além disso, critérios para a tomada de decisão e mecanismos gerais de funcionamento.

O diagrama de classes representado pela figura 13 segue a arquitetura apresentada no item 3.1, porém sem a presença de alguns módulos, onde os mesmos não foram implementados.

Figura 13 - Diagrama de classe do protótipo



Fonte: Elaborado pelo autor

4.1.1 Pontos Iniciais

A figura 13 apresenta o diagrama de classe com os métodos e seus parâmetros de entrada e saída. Além disso, esta seção apresenta os inter-relacionamentos entre os módulos componentes da arquitetura proposta. Onde os componentes evidenciam a comunicação entre os módulos do sistema que será criado a partir desta arquitetura. A identificação destes inter-relacionamentos facilita o entendimento do comportamento do sistema gerado.

Dessa forma serão apresentadas a seguir as funcionalidades e características dos métodos de cada classe aqui apresentado.

4.1.2 Classe GUI

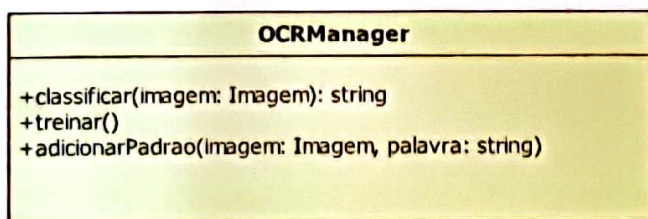
Esta classe representa o módulo de interface com o usuário que é responsável por atender as operações acessadas pelo módulo principal, o qual é representado pela classe `OcrManager`. Além disso, para implementação dessa classe, esta biblioteca permite construção de interfaces gráficas que analisa o documento colocando o mesmo em uma forma editável para que assim possa ser gerenciado através do sistema de *GED*. Mais detalhes sobre a biblioteca podem ser encontrados no item 3.3 deste capítulo.

A interface gerada por esta classe permite recortes de uma imagem simulando o módulo de segmentação para que assim possa ser indexada no sistema de *GED*.

4.1.3 Classe `OcrManager`

Esta classe representa o módulo principal da arquitetura. É composta por instâncias das classes `extrator` e `classificador`, através de um mecanismo de agregação, como mostra a figura 14.

Figura 14 - Classe do módulo principal



Fonte: Elaborado pelo autor

Na figura 14 temos a representação da interface da classe `OcrManager`, detalhando todos os métodos utilizados. Conforme descrito a seguir.

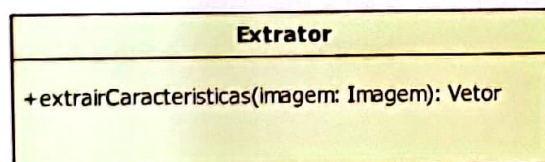
- ***classificar***: Este método recebe como parâmetro uma imagem e retorna uma *string* que representa o texto contido na imagem. Na implementação deste método existe uma interação com a classe `extrator` e `classificador`.

- **adicionarPadrao:** Este método recebe como entrada dois parâmetros, classifica o documento para sua indexação futura e assim armazena na base de dados do sistema de *GED*.
- **treinar:** Este método tem como objetivo treinar a base dados associando a imagem digitalizada ao arquivo indexado no banco de dados. Na sua implementação é repassado o banco de padrões para instância da classe classificadora, onde ocorre o treinamento do serviço de indexação.

4.1.4 Classe Extrator

Esta classe representa o módulo de extração de característica, possui um único método chamado `extrairCaracterística`, que recebe como parâmetro uma imagem e retorna um vetor de característica para que assim possa ser indexado e armazenado na base de dados.

Figura 15 - Classe do módulo extrator

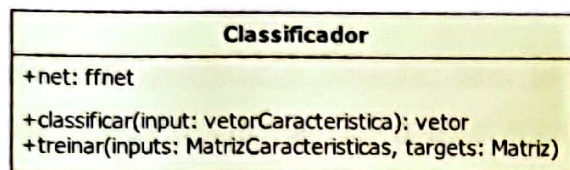


Fonte: Elaborado pelo autor

4.1.5 Classe Classificador

Esta classe representa o módulo classificador e possui como atributo uma instância de base de dados de documentos indexados, classificando assim o documento digitalizado que até o momento é caracterizado como uma imagem em seguida é extraído informações através de um sistema de *OCR* na qual é armazenado em uma base de dados e assim classificada.

Figura 16 - Classe do módulo Classificador



Fonte: Elaborado pelo autor

- **classificar:** Este método recebe como parâmetro um vetor de características e retorna um vetor numérico, onde cada número corresponde ao código ASCII da cada letra do alfabeto. Os parâmetros são passados para a instância do sistema de OCR na qual possa ser indexado e armazenado na base de dados do sistema de GED.
- **treinar:** Este método recebe como parâmetro uma matriz de características e uma matriz numérica (*targets*), onde cada número corresponde um código ASCII. Cada linha da matriz de característica possui uma linha correspondente da matriz numérica. Estes parâmetros também são passados para a instância base de dados para que possa ser classificado.

5 CONCLUSÃO

Atualmente, instituições públicas e privadas têm produzido e acumulado documentos eletrônicos de caráter arquivístico em grande volume e velocidade. Esse é um fenômeno que se acentuou a partir da década de 80, com a disseminação das tecnologias da informação, e isso representa um grande desafio porque, se não houver um tratamento adequado para esses documentos, haverá riscos de perda de provas ou perda da história da PMDF.

O grande volume de papel produzido nas organizações demanda espaços para arquivamento, e a demanda por disponibilização de informação digital também acompanha esta evolução.

Com a popularidade da tecnologia computacional, na década de 90, muda-se a forma de registro e comunicação da informação nas instituições públicas e privadas. Os documentos criados nessas organizações passaram a ser feitos no formato digital, e isso acarretou uma série de consequências, como a dificuldade de localização dos documentos e/ou perda dos mesmos, entre outros problemas.

A informação hoje é acessada rapidamente, a tecnologia da informação possibilita que sistemas sejam alimentados gerando a possibilidade de armazenamento e disseminação da informação com eficiência e rapidez.

O resultado final deste trabalho foi um conjunto de mecanismos a serem implementadas em um sistema de Gerenciamento Eletrônico de Documento – GED utilizando técnicas de OCR para digitalização dos documentos e extração de palavras para indexações. Os mecanismos que foram demonstrados através de uma proposta de arquitetura foram dividido em módulos, onde cada módulo representa uma funcionalidade a ser implementada.

A apresentação da arquitetura proposta foi através de uma metodologia de engenharia de software baseada em diagramas modulares, de forma que pudesse facilitar o reuso e o incremento de funcionalidades adjacentes existentes em cada módulo. Além disso, foi realizada uma análise de técnicas para reconhecimento, extração de características para indexação de palavras chaves para armazenamento na base de dados do sistema de GED, na qual essas técnicas foram aplicadas nos módulos apresentados.

As tecnologias de OCR, de armazenamento em base de dados, segmentação de palavras foram essenciais para conclusão da arquitetura proposta, uma vez que essas tecnologias são bastante usadas em extração de documentos. Sendo assim, fez-se uso dessas tecnologias para o desenvolvimento de um protótipo que permita a validação da arquitetura proposta.

A maior vantagem para a PMDF seria a possibilidade de manter um registro com todas as notícias e informativos, servindo tanto como valor histórico para a instituição, como ferramenta de pesquisa para policiais e cidadãos civis.

Outra vantagem seria a adequação da informação aos sete critérios de controle sugeridos pelo *Information Technology Governance Institute* como requisito para guarda e segurança da informação em uma instituição, são eles: efetividade, eficiência, confidencialidade, integridade, disponibilidade, conformidade e confiabilidade (ITGI, 2007).

A implantação de um GED no CCS deve levar em conta a necessidade do compartilhamento das informações com as outras unidades da PMDF. Um documento criado e armazenado pelo CCS deve estar disponível para todas as unidades da PMDF, sendo respeitados todos os critérios de usuário e permissão de acesso.

A partir do sistema em GED a PMDF pode atender a população de forma mais eficiente e rápida. A informação poderá auxiliar os cidadãos em diversos segmentos e ações. A proposta que aqui segue compreende o material físico e humano existente e a potencialidade natural do setor, de maneira que é viável e importante a utilização do GED no atual momento.

Dessa forma, com base no exposto durante o trabalho, o pesquisador entende que a hipótese foi confirmada e a implantação de um sistema GED pode vir a contribuir significativamente para a forma como são geridas as publicações de ocorrências e os informativos feitos pelo Centro de Comunicação Social da PMDF. O pesquisador entende ainda que o Centro de Comunicação Social da Polícia Militar do Distrito Federal detém todas as condições necessárias para a tomada de decisão de implantação de um sistema GED, voltado para veiculação de notícias institucionais, de forma a atender com mais eficiência o público interno e externo.

5.1.1 Trabalhos futuros

Em uma primeira análise, o objetivo de trabalhos futuros seria desenvolver o protótipo para a implementação direta de todos os módulos propostos na arquitetura.

Em seguida, avaliar a possibilidade da arquitetura de reconhecer documentos manuscritos com a utilização de técnicas de OCR.

A proposta que aqui se segue detém uma possibilidade de implementação no CCS, extensível a toda PMDF, melhorando qualidade do serviço de divulgação de notícia, considerando o trabalho atual da equipe do CCS, que hoje detém a missão de trabalhar com meios próprios e muita disposição para efetuar o brilhante trabalho que por hora atende à demanda atual.

REFERÊNCIAS

- AVENDON, Don M. **GED de A a Z: tudo sobre gerenciamento eletrônico de documentos**. São Paulo: Cenadem, 2002.
- BAHIA, Juarez. **Introdução à Comunicação Empresarial**. São Paulo: Mauad, 1995. 67 p.
- BALDAM, Roquemar de Lima; VALLE, Rogerio; CAVALCANTI, Marcos. **GED: Gerenciamento Eletrônico de Documentos**. 2. ed. São Paulo: Érica, 2004.
- BRASIL. Lei nº 8159, de 08 de janeiro de 1981. **Dispõe Sobre A Política Nacional de Arquivos Públicos e Privados e Dá Outras Providências**. Brasília, Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L8159.htm>. Acesso em: 23 nov. 2015.
- BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, Disponível em: <http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm>. Acesso em: 23 nov. 2015.
- BRETAS, Marco Luiz. A Polícia Carioca no Império. **Revista Estudos Históricos**, Rio de Janeiro, v. 12, n. 22, p.219-234, nov. 1998. Disponível em: <http://www.pm.al.gov.br/apm/downloads/bc_policial/pol_05.pdf>. Acesso em: 27 set. 2015.
- CABRAL, Ana Maria. **TECNOLOGIA DIGITAL EM BIBLIOTECAS E ARQUIVOS. Transiforção**, Campinas, v. 14, n. 2, p.167-177, 2002. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/1499/1473>>. Acesso em: 30 out. 2015.
- CANTU, Carlos H. **Firebird essencial**. Rio de Janeiro: Ciência Moderna, 2005. 308p.
- CONSELHO NACIONAL DE ARQUIVOS. **E-ARQ BRASIL: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos**. Rio de Janeiro: Mce Gráfica e Editora Ltda, 1011. 137 p. Disponível em: <<http://www.siga.arquivonacional.gov.br/media/earqbrasil2011.pdf>>. Acesso em: 25 nov. 2015.
- COULOURIS, George; DOLLIMORE, Jean; KINDBERG, Tim. **Sistemas Distribuídos: conceitos e projeto**. 4a ed. Porto Alegre: Bookman, 2007.
- DATE, C. J.. **Introdução à Sistema de Banco de Dados**. 1. ed. Rio de Janeiro: Campos, 1984. 513 p.
- DELTA (São Paulo). **Delta Gestão de Documentos**. Disponível em: <<http://deltarp.com.br/servicos/digitalizacao-de-documentos-papel/>>. Acesso em: 23 nov. 2015.

- EIKVIL, Line. **OCR Optical Character Recognition**. 1993. Disponível em: <<https://www.nr.no/~eikvil/OCR.pdf>>. Acesso em: 30 nov. 2015.
- ELMASRI, Ramez; NAVATHE, Shamkant B.. **Fundamentals of Database System**. Califórnia: Benajmin/cummings Publishing Company, 1989. 802 p.
- ELMASRI, Ramez; NAVATHE, Shamkant B.. **Sistema de Banco de Dados**. 6. ed. São Paulo: Pearson Education do Brasil, 2011. 788 p.
- EMMERICH, Wolfgang; **Engineering Distributed Objects**. 1. Ed. England: John Wiley & Sons, 2000. 390p.
- FANTINI, Sérgio Rubens. **Aplicação do gerenciamento eletrônico de documentos: estudo de caso de escolhas de soluções**. 2001. 116 f. Tese (Doutorado) - Curso de Engenharia, Universidade Federal de Santa Catarina, Florianópolis, 2001. Disponível em: <<https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/79392/179200.pdf?sequence=1&isAllowed=y>>. Acesso em: 09 nov. 2015.
- FERNÁNDEZ, Félix Sagredo; ARROYO, José María Izquierdo. Reflexiones sobre "documento" palabra/objeto. **Boletín Millares Carlo**. Las Palmas de Gran Canaria, p. 161-198. dez. 1982. Disponível em: <<http://dialnet.unirioja.es/descarga/articulo/1448715.pdf>>. Acesso em: 09 nov. 2015.
- INFORMATION TECHNOLOGY GOVERNANCE INSTITUTE. **COBIT 4.1: Framework COBIT - Control Objectives for Information and related Technology**. 4.1 ed. Rolling Meadows: Isaca, 2007. 212 p. Disponível em: <http://analistati.com/wp-content/uploads/2010/01/cobit41_portuguese.pdf>. Acesso em: 25 nov. 2015. KOCH,
- MANTAS, John. An Overview of Character Recognition Methodologies. **Pattern Recognition**. Athens, p. 420-430. 22 nov. 1986.
- MARCONI, Marina de Andrade, LAKATOS, Eva Maria. **Metodologia do Trabalho Científico**. 2014, Editora Atlas.
- MENDONÇA, Fábio Lúcio Lopes de. **Proposta de Arquitetura de um Sistema com Base em OCR Neuronal Para Resgate e Indexação de Escritas Paleográficas do Século XVI ao XIX**. 2008. 118 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, 2008. Disponível em: <http://repositorio.unb.br/bitstream/10482/1157/1/DISSERTACAO_2008_FabioLucioLMendonca.pdf>. Acesso em: 08 dez. 2015.
- RHOADS, James B. **La Funcion de la gestion de documentos y archivos en los sistemas nacionales de información: un estudio del Ramp**. Paris: UNESCO, 1989. Disponível em: <<http://unesdoc.unesco.org/images/0008/000847/084735so.pdf>> Acesso em: 09 nov. 2015.
- RONDINELLI, Rosely Curi. **Gerenciamento Arquivístico de Documentos Eletrônicos**. 4. ed. Rio de Janeiro: Fgv, 2005. 183 p.

SANTOS, Izequias Estevam dos. **Manual de Métodos e Técnicas de Pesquisa Científica**. 8. ed. Niterói: Impetus, 2013.

SANTOS, João Almeida, FILHO, Domingos Parra, **Metodologia Científica**. 2011, Editora Cengage Learning.

SETZER, Valdemar W.; SILVA, Flávio Soares Corrêa da. **Bancos de Dados: Aprenda o que são; melhore seu conhecimento; construa o seu**. São Paulo: Edgar Blucher, 2005. 380 p.

SEVERINO, Antonio Joaquim. **Metodologia do Trabalho Científico**. 2013, 23ª Edição, Editora Cortez

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S.. **Sistema de Banco de Dados**. 3. ed. São Paulo: Pearson, 2004. 778 p.

SILVA, Gilvan Gomes da. **A LÓGICA DA POLÍCIA MILITAR DO DISTRITO FEDERAL NA CONSTRUÇÃO DO SUSPEITO**. 2009. 187 f. Tese (Doutorado) - Curso de Sociologia, Universidade de Brasília, Brasília, 2009. Disponível em:

SOIBELMAN, Leib. **Enciclopédia do Advogado**. 5ª ed. rev. Rio de Janeiro: Thex, 1994

SPRAGUE JUNIOR, Ralph H. **Electronic Document Management: Challenges and Opportunities for Information Systems Managers**. 1995. Disponível em: <<http://sprague.shidler.hawaii.edu/MISQ/MISQfina.htm>>. Acesso em: 27 out. 2015.

STALLINGS, William. **Cryptography and Network Security: Principles and Practice**. 5. ed. New York: Pearson, 2010.

TANENBAUM, Andre S.. **Sistemas Distribuídos: princípios e paradigmas**. 2. ed. São Paulo: Pearson Education do Brasil, 2008. 412 p.

TAKAI, Osvaldo Kotaro; ITALIANO, Isabel Cristina; FERREIRA, João Eduardo. **INTRODUÇÃO A BANCO DE DADOS**. Sao Paulo: Dcc-ime, 2005. 124 p.

Walter W. **Gerenciamento eletrônico de documentos: conceitos tecnologias e considerações gerais**. São Paulo: CENADEM, 1997.



**GOVERNO DO DISTRITO FEDERAL
POLÍCIA MILITAR DO DISTRITO FEDERAL
DEPARTAMENTO DE EDUCAÇÃO E CULTURA
INSTITUTO SUPERIOR DE CIÊNCIAS POLICIAIS**



ANEXO A – ENTREVISTA AOS INTEGRANTES DO DEPARTAMENTO DE GESTÃO DA INFORMAÇÃO.

Esta entrevista é parte dos instrumentos de coleta de dados que integram a pesquisa: PROPOSTA DE ARQUITETURA PARA VIABILIDADE DE IMPLANTAÇÃO DE UM SISTEMA DE GERENCIAMENTO ELETRÔNICO DE DOCUMENTOS PARA O CENTRO DE COMUNICAÇÃO SOCIAL DA PMDF.

- 1- Trabalha com a informação a quanto tempo?
1-5 anos () 6 a 10 anos () mais de 11 anos ().
- 2- Atualmente o sistema de informação utilizado pela PMDF é informatizado?
Como funciona na prática do dia a dia?
Sim () não ()
- _____
- _____
- _____
- _____
- _____
- 3- A informatização é um elemento da modernidade. Na sua opinião qual a importância de implantação de um sistema de informação totalmente informatizado?
- () rapidez na informação.
() acesso a informação facilitado.
() gerenciamento da informação.
() disponibilidade de acesso a sociedade.
() confiabilidade da informação.
() eficiência na recepção de dados.
() facilidade na indexação.
- 4- Em relação ao armazenamento e disseminação, atualmente como funciona no dia a dia, o arquivo das informações recebidas e a disseminação das mesmas?
() existe () inexistente
- _____
- _____
- _____
- _____
- _____

