

DULCE AMÉLIA DE BRITO NEVES
MARIA MERIANE VIEIRA ROCHA
PATRÍCIA SILVA
(Organizadoras)

CARTOGRAFIA
DA PESQUISA
E ENSINO
DA ARQUIVOLOGIA
NO BRASIL:
IV REPARO

**CARTOGRAFIA DA PESQUISA
E ENSINO DA ARQUIVOLOGIA
NO BRASIL: IV REPARQ**



**UNIVERSIDADE
FEDERAL DA PARAÍBA**

Reitora Margareth de Fátima Formiga Melo Diniz
Vice-Reitor Bernardina Maria J. Freire de Oliveira



**Editora
UFPB EDITORA DA UFPB**

Diretora Izabel França de Lima

Supervisão de Editoração Almir Correia de Vasconcellos Júnior
Supervisão de Produção José Augusto dos Santos Filho

COMISSÃO ORGANIZADORA DA IV REPARQ

Prof. Maria Meriane Vieira Rocha (UFPB)
Prof. Rosa Zuleide Lima de Brito (UFPB)
Prof. Esmeralda Porfírio de Sales (UEPB)

COMISSÃO CIENTÍFICA

Dulce Amélia de Brito Neves (Coord. - UFPB)
Bernardina M. J. Freire de Oliveira (UFPB)
Ana Célia Rodrigues (UFF)
Eva Cristina Leite da Silva (UFSC)
José Maria Jardim (UFRJ)
Josemar Henrique de Melo (UEPB)
Kátia Isabelli Melo de Souza (UNB)
Lucia Maria Velloso de Oliveira (FCRB)
Lúcia de Fátima Guerra Ferreira (UFPB)
Daniel Flores (UFSM)
Ivana Parrela (UFMG)
Maria Celina S. Mello e Silva (Museu de Astronomia)
Maria Leandra Bizello (UNESP/MARÍLIA)
Paulo Roberto Elian dos Santos (FIOCRUZ)
Renato Pinto Venâncio (UFMG)
Rosa Zuleide Lima de Brito (UFPB)

DULCE AMÉLIA DE BRITO NEVES
MARIA MERIANE VIEIRA ROCHA
PATRÍCIA SILVA
(Organizadoras)

CARTOGRAFIA DA PESQUISA
E ENSINO DA ARQUIVOLOGIA
NO BRASIL: IV REPARQ

Editora da UFPB
João Pessoa
2016

Copyright © 2016 - Dulce Amélia de Brito Neves et al.
Efetuado o Depósito Legal na Biblioteca Nacional,
conforme a Lei nº 10.994, de 14 de dezembro de 2004.

TODOS OS DIREITOS RESERVADOS À EDITORA DA UFPB

É proibida a reprodução total ou parcial, de qualquer forma ou por qualquer meio. A violação dos direitos autorais (Lei nº 9.610/1998) é crime estabelecido no artigo 184 do Código Penal.

O conteúdo desta publicação é de inteira responsabilidade do autor.

Impresso no Brasil. Printed in Brazil.

Projeto Gráfico	EDITORA DA UFPB
Editoração Eletrônica	Alexandre Câmara
Design de Capa	Alexandre Câmara
Ilustração de Capa	Alexandre Câmara
Revisão	Patrícia Silva e Maria Meriane Vieira Rocha
Normalização	Patrícia Silva e Kleisson Lainnon da Silva

Catálogo na fonte:

Biblioteca Central da Universidade Federal da Paraíba

C328 Cartografia da pesquisa e ensino da arquivologia no Brasil: IV REPARQ [recurso eletrônico] / Dulce Amélia de Brito Neves, Maria Meriane Vieira Rocha, Patrícia Silva, organizadoras.- João Pessoa: Editora da UFPB, 2015.
Recurso digital (6,5 MB)
Formato: ePDF
Requisito do Sistema: Adobe Acrobat Reader
Edição Bilingue
ISBN: 978-85-237-1227-3 (recurso eletrônico)
1. Arquivologia - ensino. 2. Cartografia - pesquisa e ensino - arquivologia. 3. Arquivologia - currículos. 4. Docentes - arquivologia - perfil. I. Neves, Dulce Amélia de Brito. II. Rocha, Maria Meriane Vieira. III. Silva, Patrícia.

CDU: 930.25:37

EDITORA DA UFPB Cidade Universitária, Campus I – s/n
João Pessoa – PB
CEP 58.051-970
editora.ufpb.br
editora@ufpb.edu.br
Fone: (83) 3216.7147

SUMÁRIO

CONFERÊNCIA DE ABERTURA	10
ENSEÑANZA E INVESTIGACIÓN DE LA ARCHIVÍSTICA EN ESPAÑA <i>Concepción Mendo Carmona</i>	11
PLENÁRIAS	48
A DIPLOMÁTICA CONTEMPORÂNEA: SUA INTERFACE COM O ENSINO E A PESQUISA EM ARQUIVOLOGIA <i>Ana Célia Rodrigues</i>	49
A PESQUISA EM ARQUIVOLOGIA: MÉTODOS, ESPECIFICIDADES E DIÁLOGOS <i>José Maria Jardim</i>	73
DESAFIOS CONTEMPORÂNEOS DOS CURRÍCULOS DE ARQUIVOLOGIA: A QUESTÃO DOS DOCUMENTOS ARQUIVÍSTICOS DIGITAIS E SUAS RELAÇÕES INTERDISCIPLINARES DA ARQUIVOLOGIA <i>Daniel Flores</i>	91
HARMONIZAÇÃO CURRICULAR: ANÁLISE DAS CONFIGURAÇÕES ACADÊMICO-INSTITUCIONAIS E DO PERFIL DOCENTE DOS CURSOS DE ARQUIVOLOGIA NO BRASIL <i>Welder Antônio Silva, Cíntia Aparecida Chagas Arreguy, Leandro Ribeiro Negreiros</i>	119
COMUNICAÇÕES ORAIS	179
A ENTRADA DE ARQUIVISTAS E TÉCNICOS DE ARQUIVOS NO PODER EXECUTIVO FEDERAL DO BRASIL NO PERÍODO DE 2004 A 2012: SUA RELAÇÃO COM OS EGRESSOS DOS CURSOS DE ARQUIVOLOGIA E O QUANTITATIVO DE MINISTÉRIOS EXISTENTES <i>Djalma Mandu de Brito</i>	180

DA GRADUAÇÃO AO MERCADO DE TRABALHO: UMA REFLEXÃO ACERCA DOS EGRESSOS DO CURSO DE ARQUIVOLOGIA DA UNIVERSIDADE FEDERAL DE SANTA MARIA <i>Tatiana Costa Rosa, Rosanara Pacheco Urbanetto</i>	198
EDITAIS DE CONCURSOS FEDERAIS PARA ARQUIVISTAS: UMA ANÁLISE DA RELAÇÃO ENTRE A FORMAÇÃO EM ARQUIVOLOGIA E CERTAMES PÚBLICOS <i>Josemar Henrique Melo, Américo Augusto Nogueira Vieira, Marcilio Toscano Franca Filho, Ademir Clemente, Cleber Ferreira Silva</i>	226
AVALIAÇÃO DE DESEMPENHO ACADÊMICO: APRENDIZAGENS E DESAFIOS <i>Tânia Barbosa Salles Gava, Luciana Itida Ferrari, Dulcinea Sarmento Rosemberg</i>	252
O CURSO DE ARQUIVOLOGIA E AS COMPETÊNCIAS DE SEUS ALUNOS: UFBA E UEL <i>Linete Bartalo, Jussara Borges</i>	274
GESTÃO DE DOCUMENTOS: UMA PROPOSTA DE CURSO DE CAPACITAÇÃO A DISTÂNCIA PARA O PODER EXECUTIVO FEDERAL <i>Djalma Mandu de Brito</i>	308
PARA ALÉM DOS MUROS DOS ARQUIVOS: GESTÃO DE DOCUMENTOS EM ARQUIVÍSTICA À LUZ DA TEORIA DOS SISTEMAS ABERTOS <i>Luiz Carlos da Silva, Jorge Santa Anna</i>	335
LEVANTAMENTO DOS ARQUIVOS PÚBLICOS MUNICIPAIS DO ESPÍRITO SANTO: ANALISANDO ASPECTOS DE INFRAESTRUTURA E DE GESTÃO <i>Luiz Carlos da Silva, Jorge Santa Anna</i>	362
GESTÃO DOCUMENTAL COMO FERRAMENTA DE EFICIÊNCIA NA GESTÃO PÚBLICA: ESTUDO DO ARQUIVO GERAL DA SECRETARIA MUNICIPAL DE SAÚDE DO MUNICÍPIO DE JOÃO PESSOA - PB <i>Jesiel Ferreira Gomes, Jucely Neves de Melo</i>	390

APLICAÇÃO DA NORMA INTERNACIONAL DE GESTÃO DE RISCOS: UM ESTUDO DE CASO NA SUPERINTENDÊNCIA DO PORTO DO RIO GRANDE <i>Luciana Souza de Brito, Roberta Pinto Medeiros</i>	417
CONSTRUÇÃO DE VOCABULÁRIO CONTROLADO VINCULADO A UM INSTRUMENTO DE CLASSIFICAÇÃO DE DOCUMENTOS DE ARQUIVO PARA FACILITAR O ACESSO À INFORMAÇÃO PÚBLICA <i>Erick Oliveira Alves de Souza, Talles Humberto Souza Moreira, Julia Araujo Donato, Renato Tarciso Barbosa de Sousa</i>	440
A FUNÇÃO AVALIAÇÃO DE DOCUMENTOS DE ARQUIVO NO PODER EXECUTIVO FEDERAL BRASILEIRO <i>Eliane Braga Oliveira, Maria Ivonete Gomes Nascimento</i>	461
O TRABALHO DE ARRANJO E DESCRIÇÃO DE ACERVOS ARQUIVÍSTICOS NO BRASIL <i>Eliezer Pires da Silva, Cintia da Silva Ribeiro</i>	482
APLICAÇÃO DO MARKETING NO ARQUIVO GERAL DA UNIVERSIDADE FEDERAL DA PARAÍBA <i>Juliana Soares da Fonseca, Maria Meriane Vieira Rocha, Ana Claudia Medeiros de Sousa</i>	503
REPOSITÓRIO ARQUIVÍSTICO DIGITAL: ANÁLISE DAS FERRAMENTAS PARA A PRESERVAÇÃO DO PATRIMÔNIO DOCUMENTAL EM MEIO DIGITAL <i>Mateus de Moura Rodrigues</i>	523
O REPOSITÓRIO DIGITAL COMO UM RECURSO PARA ACESSO E PRESERVAÇÃO DO DIÁRIO DE CLASSE <i>Sérgio Renato Lampert</i>	551
PRESERVAÇÃO DIGITAL DOS VIDEOGAMES: PRIMEIRAS APROXIMAÇÕES COM A ARQUIVOLOGIA <i>Roberto Lopes dos Santos Junior</i>	578
COMPORTAMENTO INFORMACIONAL DOS USUÁRIOS DO SISTEMA DE ARQUIVOS DA UEL – SAUEL <i>Linete Bartalo, Wilmara Rodrigues Calderon, Ivone Guerreiro DiChiara, Neiva Aranda Lopes Butarello</i>	608

USOS E USUÁRIOS DE ARQUIVO: UMA BREVE RELEXÃO <i>Tiago Braga da Silva, Junia Gomes da Costa Guimarães e Silva</i>	634
ELEMENTOS TEMÁTICOS DA PESQUISA EM ARQUIVOLOGIA CONTEMPLADOS NO PERIÓDICO CIENTÍFICO PONTO DE ACESSO <i>Kátia de Oliveira Rodrigues, Sérgio Franklin, Eliete Lima</i>	657
A LEI DE ACESSO À INFORMAÇÃO PÚBLICA COMO OBJETO DE PESQUISA NA ARQUIVOLOGIA: A ANÁLISE DE ARTIGOS CIENTÍFICOS <i>Lorrane Cristina Passos Sezinando</i>	680
ARQUIVOS COMO MECANISMOS DE DIFUSÃO DE CONHECIMENTOS PARA A APRENDIZAGEM E INOVAÇÃO EM ORGANISMOS PRODUTORES DE SAÚDE <i>Francisco José Aragão Pedroza Cunha, Gillian Leandro de Queiroga Lima, Louise Anunciação Fonseca de Oliveira</i>	701
DISPERSÃO DA LITERATURA CIENTÍFICA EM ARQUIVOLOGIA: UM ESTUDO DAS REVISTAS INDEXADAS PELA WEB OF SCIENCE (WOS) <i>Rita de Cássia Portela Silva, Maria Luisa Lascurain Sánchez</i>	724
A CONSTRUÇÃO COLETIVA DOS ARQUIVOS PESSOAIS: UM OLHAR SOBRE O ARQUIVO PESSOAL DE DOM ADRIANO MANDARINO HYPÓLITO <i>João Marcus Figueiredo Assis, Bruno Ferreira Leite</i>	745
ENTRE A ARQUIVOLOGIA E A HISTÓRIA - PROJETO DE REORGANIZAÇÃO DO ARQUIVO DA SANTA CASA DE MISERICÓRDIA DA PARAÍBA <i>Josemar Henrique Melo, Nereida Soares Martins da Silva</i>	764
RELATÓRIO FINAL: DELIBERAÇÕES, RECOMENDAÇÕES E MOÇÕES	788
AGRADECIMENTOS	792

CONSTRUÇÃO DE VOCABULÁRIO CONTROLADO VINCULADO A UM INSTRUMENTO DE CLASSIFICAÇÃO DE DOCUMENTOS DE ARQUIVO PARA FACILITAR O ACESSO À INFORMAÇÃO PÚBLICA

*Erick Oliveira Alves de Souza
Universidade de Brasília
eosouza.fun@gmail.com*

*Talles Humberto Souza Moreira
Universidade de Brasília
talleshumberto@gmail.com*

*Julia Araújo Donato
Universidade de Brasília
julia.araujo.donato@hotmail.com*

*Renato Tarciso Barbosa de Sousa
Universidade de Brasília
talleshumberto@gmail.com*

1 INTRODUÇÃO

A produção crescente de documentos e informações parece ser um fenômeno contemporâneo. O arquivista Bruno Delmas, em uma projeção feita na década de 1980, chega a estimar que metade da massa documental existente no mundo, naquela época, tinha sido acumulada nos últimos 37 anos (DELMAS

apud SOUSA, 2015, p.15).

Essa explosão implicou em problemas na organização dos documentos proporcionados por uma falta de gestão que gerou um grande entrave para a recuperação e o acesso aos documentos e, assim, à informação. E, com o passar dos anos, esse número de documentos produzidos e recebidos cresceu em proporções ainda mais surpreendentes. Os meios tecnológicos de produção e reprodução de documentos facilitam o desenho desse cenário (SOUSA, 2015, p.16).

Esse volume enorme de documentos e informações exige, cada vez mais, um instrumental mais sofisticado. Desconfia-se da capacidade do plano de classificação de documentos e informações de dar conta de atender as necessidades de acesso aos documentos e informações nele contidas.

Além disso, as alterações ocorridas nas formas de acessar à informação, proporcionadas pelos motores de busca com o advento da internet, e as dificuldades de operacionalização da classificação dos documentos de arquivo nos ambientes de trabalho e nos protocolos limitam o papel da classificação no acesso aos documentos de arquivo (SOUSA, 2015, p.21). Portanto, a classificação não pode responder mais sozinha pela importante e atualizada tarefa de busca à informação contida nos documentos de arquivo. (SOUSA, 2015, p. 21).

Esta pesquisa parte do pressuposto que para tornar a busca mais eficiente aos documentos de arquivo é necessário combinar o plano de classificação de documentos e informações com um vocabulário controlado.

Aproveitamos os resultados de uma pesquisa de iniciação científica desenvolvida, em 2009, na Universidade de Brasília, intitulada Construção de vocabulário controlado para identificação do conteúdo informacional dos documentos acumulados pela atividade-meio da Administração Pública

Federal (SOUSA, MESQUITA, MARTINS, 2010), que tinha como objetivo a montagem de um protótipo com uma lista de termos com maior ocorrência para serem incluídos nas unidades de classificação. Na verdade, o estudo em tela é uma continuidade dessa pesquisa anterior.

2 OBJETIVOS

Percebe-se, então, a exigência de uma maior sofisticação dos instrumentos de recuperação da informação. O código ou plano de classificação, hoje em dia, é um dos únicos instrumentos utilizados para tal fim. Dessa forma, a classificação não está sendo suficiente para o grande volume documental, pois privilegia somente a função que originou os documentos, ou seja, um determinado conjunto de documentos. Então, como melhorar o acesso à informação pública?

A elaboração de um vocabulário controlado integrado a uma estrutura de classificação, utilizando como base a legislação que originou os documentos, pode ser a solução para a recuperação da informação eficiente. O objetivo da pesquisa, então, é verificar a possibilidade da construção de um vocabulário controlado a partir da mineração de dados do marco regulatório (leis e decretos) das grandes funções das atividades-meio da Administração Pública Federal, isto é, a gestão dos recursos humanos, gestão dos recursos financeiros, gestão dos recursos materiais e gestão dos recursos informacionais.

O espaço empírico da pesquisa é a Administração Pública Federal, que conta, inclusive, com um instrumento de classificação para os documentos das atividades-meio, elaborado pelo Conselho Nacional de Arquivos, por meio de sua Câmara Técnica de Classificação. Esse instrumento foi publicado, originalmente, em 1996, mas desde esse ano vem sofrendo alterações. A escolha do Código de Classificação, do

Conselho Nacional de Arquivo (Conarq), foi feita por que a Administração Pública Federal é um dos principais cenários da Arquivologia no Brasil. Além disso, há poucos códigos de classificação publicados no país com a mesma abrangência. Esse código trabalha com as atividades-meio da Administração Pública Federal, e consegue atingir outros órgãos públicos estaduais e distritais e muitas vezes empresas privadas. Ademais, a Administração Pública Federal, por ser composta por órgãos importantes e responsáveis por muitas decisões, trabalha como o número grande de atividades, e isso, conseqüentemente, gera um volume documental elevado. Tudo isso justifica a escolha pela Administração Pública Federal como universo desta pesquisa.

Busca-se, portanto, a construção do vocabulário controlado, a partir da mineração de dados feita nas leis e decretos que regulamentam as grandes funções das atividades-meio, vinculando os termos às unidades de classificação do Código de Classificação de Documentos de Arquivo, elaborado pelo Conselho Nacional de Arquivos.

Entende-se que essa vinculação entre termos e unidades de classificação aumente a eficiência no acesso às informações contidas nos documentos de arquivo.

O vocabulário controlado poderá ser utilizado, também, para melhorar a busca nos sistemas informatizados de protocolo. O registro e o controle da tramitação dos documentos, segundo Sousa (1997, p. 35-36), é uma atividade, hoje, desenvolvida por unidades específicas, que figuram nas estruturas organizacionais com a denominação de “protocolo”, “protocolo e arquivo”, “comunicação administrativa”, “documentação e comunicação administrativa” etc. Estes setores, na maior parte dos casos, são responsáveis pelo registro e distribuição das correspondências produzidas e recebidas pelos órgãos, bem

como pela protocolização dos processos e sua tramitação.

Na tradição administrativa brasileira, apesar da variedade de expressões, esse setor ou atividade é reconhecido como protocolo. O dicionário afirma: protocolo é etimologicamente uma palavra derivada do grego *protókollon*, isto é, “a primeira folha colada aos rolos de papiro, e na qual se escrevia um resumo do conteúdo do manuscrito”. O Dicionário Brasileiro de Terminologia Arquivística entende como o “setor encarregado do recebimento, registro, distribuição e tramitação de documentos”. Essas palavras compartilham traços que indicam identificação e controle. Neste caso, protocolo é a prova da entrada ou saída da informação e de sua localização na organização.

A inoperância histórica desse setor ou atividade deu a ele uma conotação pejorativa dentro das organizações. De uma unidade estratégica para uma política de gestão dos arquivos transformou-se em um simples entreposto para o recebimento e expedição das informações e um depósito de uma pequena parcela dos documentos que teve sua tramitação encerrada. As portas de entrada, com os novos recursos tecnológicos disponíveis (fac-símile, correio eletrônico) e a ausência de procedimentos reconhecidos pela organização, ampliaram-se sem que houvesse uma igual expansão das atividades de controle dos protocolos. Os setores de trabalho recebem ou expedem documentos arquivísticos por fac-símile ou por e-mail que passam a margem de qualquer tipo de controle ou registro. Não podemos esquecer, ainda, daqueles que entram e saem pelas mãos dos funcionários. Percebe-se, então, que nem todos são registrados nos protocolos. Muitos tramitam sem qualquer tipo de controle. Normalmente, apenas os processos recebem um número, que se constitui na chave de busca dele. Essas atividades geram uma quantidade muito grande de fichas, livros e formulários. Não há, salvo raras exceções, relação entre essas

atividades e aquelas executadas nas outras seções dos órgãos, no que se refere à classificação dos documentos arquivísticos.

O uso cada vez mais frequente da Informática possibilitou, em muitas organizações públicas e privadas, a substituição do registro e controle manual da tramitação por sistemas automatizados. É possível encontrá-los disponibilizados em redes locais e remotas. Esses sistemas têm sido desenvolvidos por profissionais de Informática. A estrutura das bases de dados, criadas para este fim, reproduz os mesmos campos definidos nas fichas, formulários e livros de protocolo. (SOUSA, 1997, p. 36)

Um dos principais problemas encontrados nos sistemas informatizados é o preenchimento do campo intitulado Assunto ou Descrição do Assunto, que tem como objetivo permitir uma identificação do conteúdo informacional do documento. Nesse campo, deve ser descrita a razão de ser do registro documental, qual o motivo de sua criação.

Trata-se de um campo chave para o processo de busca das informações. Sabemos, entretanto, que a maior parte das atividades que são realizadas diariamente em qualquer organização é rotineira. Portanto, um mesmo conteúdo informacional pode ser encontrado em inúmeros registros documentais. Independente da espécie documental. Por exemplo, a atividade de compra de material produz, geralmente, os mesmos documentos, ou melhor, o mesmo conteúdo informacional fixado em espécies diferentes. A rotina tem início com uma solicitação feita por uma correspondência (memorando ou ofício), uma pesquisa de preço junto a fornecedores realizada por meio de correspondências, a resposta dos possíveis fornecedores e a efetivação da compra (nota fiscal, autorização de pagamento etc.). O conteúdo informacional, portanto, é somente um: compra ou aquisição de material.

O campo Assunto é preenchido a partir dos termos

existentes no cabeçalho do documento. Quando isso não existe, os próprios funcionários do setor de protocolo encarregam-se de preenchê-lo.

A análise do conteúdo desse campo nos registros de sistemas de protocolo informatizados indicou uma total falta de padrão para o preenchimento do mesmo. É impossível imaginar que uma organização, por mais complexa que ela seja, tenha uma variedade tão expressiva de assuntos. Os ganhos de eficiência no acesso à informação perdem-se num emaranhado de termos vazios, fragmentados, sem conexão e sem uma vinculação muito clara com o conteúdo informacional dos documentos. É preciso lembrar que os recursos informáticos não permitem ainda uma interpretação que resolva a imprecisão do preenchimento do campo Assunto. A lógica é binária.

Pretende-se colaborar, com esta pesquisa, também, para a solução do problema do preenchimento do campo Assunto, dos sistemas informatizados de gestão de documentos.

3 FUNDAMENTAÇÃO TEÓRICA

A pesquisa está fundamentada no entendimento de que a informação contida nos documentos de arquivo é uma informação **interna**, produzida por pessoas (físicas ou jurídicas) no desenvolvimento de suas atividades, de forma necessária e inevitável; é uma informação **previsível**, pois é fruto de processos estabelecidos (procedimentos administrativos ou processos de negócios). É uma informação **regulada** em sua criação, uso e conservação. A criação está regulada por normas legais e/ou procedimentos internos. A utilização (tramitação, acesso, informação, obtenção de cópias) está sancionada por normas legais e/ou por normativa interna. A conservação (destinação

final) é regulada por normas. (CRUZ MUNDET, 2006).

O documento arquivístico é um artefato humano com pressupostos e características específicas. O ambiente e o conteúdo são delimitados e definidos pelo sujeito acumulador, que pode ser uma pessoa física ou jurídica. Então quando falamos de arquivo, estamos nos referindo a um conjunto finito de documentos acumulados, que tem suas fronteiras demarcadas pela missão do criador, no caso das instituições, e pela área de atuação, no caso das pessoas físicas. Ao contrário daqueles encontrados em bibliotecas, por exemplo, os documentos arquivísticos não constituem um conjunto formado em vista de uma finalidade específica: eles representam, mais que tudo, o produto da atividade do sujeito criador.

Entender o modo como as instituições se estruturam e como executam suas funções e atividades é compreender como os documentos são acumulados. Ele é resultado de um ato desenvolvido e, na maioria dos casos, cotidianamente repetido. A gênese se dá quando a organização tem algo a cumprir, a provar, a determinar. Surge naturalmente como resultado das ações desenvolvidas pelo sujeito criador. Após o registro das informações em suportes (papel, mídia magnética, microfilme, películas fotográficas, películas cinematográficas etc.), é necessário mantê-los pelos valores administrativos, técnicos, legais, fiscais, probatórios, culturais e históricos que possam conter.

À medida que os documentos vão sendo acumulados, estabelecem relações entre si. Eles estão unidos por um elo criado no momento em que são produzidos e recebidos, determinado pela razão de sua elaboração e que é necessário à própria existência e a capacidade de cumprir seu objetivo. Eles são um conjunto indivisível de relações intelectuais, onde o “todo é maior que a soma de suas partes”.

Se o documento é o resultado da atividade de uma pessoa física ou jurídica, podemos falar do caráter orgânico desse registro. A organicidade é revelada pelo inter-relacionamento e pelo contexto de existência e de criação. Entretanto, nem todos os documentos orgânicos são de caráter arquivístico, pois essa qualificação é limitada em termos de suportes (convencionais ou eletrônicos). Por exemplo, é comum encontrar, principalmente nas indústrias, informações orgânicas tridimensionais que não são arquivísticas. O suporte, nesse caso, não permite o reconhecimento desse documento como de caráter arquivístico, apesar de entendermos que as características físicas não sejam os atributos mais seguros para definição do caráter arquivístico de um documento orgânico.

Sendo assim, a recuperação da informação para o acesso à informação pública pode ser padronizada, já que os documentos possuem, normalmente, sempre a mesma estrutura. Acredita-se, então, que a eficiência dessa recuperação se dê pelo controle de vocabulário aliado a uma estrutura de classificação que possa atender a critérios necessários para a preservação do vínculo arquivístico.

Defendemos, nesta pesquisa, o uso do conceito de classificação para representar a atividade intelectual de construção de instrumentos para organização dos documentos, independentemente da idade à qual eles pertençam. A confusão terminológica entre dois termos (arranjo e classificação) não parece salutar ao desenvolvimento da Arquivística, pois expõe uma quebra entre arquivos correntes e permanentes, que no nosso entendimento não existe. Trata-se apenas de fases de um mesmo processo. É evidente que o tipo de uso que se faz dos conjuntos documentais altera-se com as idades, ou melhor, novos usos vão sendo agregados, mas essa é uma questão a ser resolvida por uma outra função arquivística: a descrição.

No âmbito desta pesquisa, utilizaremos o termo classificação para identificar a ação intelectual de construir esquemas para agrupar os documentos a partir de princípios estabelecidos. A ordenação como a forma de disposição dos tipos documentais dentro das divisões estabelecidas no esquema de classificação. O arquivamento como a ação física de colocar os documentos em pastas ou caixas orientada pelo esquema de classificação e pela ordenação definida.

Para vocabulário controlado, utilizaremos a definição de Kobashi (2008, p.1), que o entende como:

uma LINGUAGEM ARTIFICIAL constituída de termos organizados em estrutura relacional. Um vocabulário controlado é elaborado para padronizar e facilitar a entrada e a saída de dados em um sistema de informações. Tais atributos promovem maior precisão e eficácia na comunicação entre os usuários e o sistema de informações.

A partir dessa problemática, um vocabulário controlado é utilizado para indexar documentos. Indexar é caracterizar conteúdos de documentos por meio dos descritores de um vocabulário controlado (KOBASHI, 2008, p. 2). Dessa forma, a informação contida no próprio documento e a informação contextual podem ser alcançadas, isto é, o “vínculo arquivístico” – de acordo com a compreensão de Luciana Duranti - é recuperado.

A construção do vocabulário controlado como forma de identificar as palavras necessárias pode ser feita por meio de uma mineração de textos dos marcos regulatórios das grandes funções das atividades-meio (recursos humanos, recursos materiais, recursos financeiros e recursos informacionais). Os termos substantivos minerados serão vinculados às unidades de Classificação do Código de Classificação do CONARQ, pois esse foi elaborado baseado nas funções e atividades administrativas

necessárias aos órgãos para realizar suas missões. Desse modo, deve-se fazer uma pesquisa na legislação referentes aos assuntos de cada subclasse do campo 000 do código.

Os benefícios da mineração de textos, para Ambrósio e Morais (2007, p.1), estão relacionados à busca de informações específicas em documentos, à análise qualitativa e quantitativa de grandes volumes de textos, e a melhor compreensão do conteúdo disponível em documentos textuais. Assim, os termos controlados associados as unidades de classificação da estrutura funcional do código de classificação podem permitir o acesso rápido e eficiente à informação demandada.

Gonçalves (2012) definiu essa atividade como:

Mineração de texto é uma subárea da mineração de dados interessada no desenvolvimento de técnicas e processos para a descoberta automática de conhecimento valioso a partir de coleções de documentos texto. Esse processo utiliza algoritmos capazes de analisar coleções de documentos texto com o objetivo de extrair conhecimento.

Já Aranha e Passos (2006, p.1) conceituam a mineração da seguinte forma:

Mineração de textos consiste em extrair regularidades, padrões ou tendências de grandes volumes de textos em linguagem natural, normalmente, para objetivos específicos. Inspirado pelo data mining ou mineração de dados, que procura descobrir padrões emergentes de banco de dados estruturados, a mineração de textos pretende extrair conhecimentos úteis de dados não estruturados ou semi-estruturados.

Segundo Ambrósio e Morais (2007, p.6), ao utilizar a mineração de textos,

um usuário não solicita exatamente uma busca, mas sim uma análise de um documento. Entretanto, este não recupera o conhecimento em si. É importante que o resultado da consulta seja analisado e contextualizado para posterior descoberta de conhecimento.

Na prática, a mineração de textos define um processo que auxilia na descoberta de conhecimento inovador a partir de documentos textuais, que pode ser utilizado em diversas áreas do conhecimento. Dessa forma, analisa-se a mineração de textos ou text mining como uma forma prática de recuperação da informação.

4 METODOLOGIA

Parte 1 - Revisão da literatura

O ponto de partida da pesquisa foi a revisão da literatura correspondente ao assunto com o intuito de delimitar o tema. Os textos trabalhados foram: “Como elaborar vocabulário controlado para aplicação em arquivos” da SMIT, Johanna Wilhelmina e KOBASHI, Nair Yumiko e “Construção de vocabulário controlado para identificação do conteúdo informacional dos documentos acumulados pela atividade-meio da Administração Pública Federal”, do SOUSA, Renato Tarciso Barbosa, MESQUITA, Heloisa Carvalho e MARTINS, Larissa Marques.

Depois, houve a leitura do livro “Indexação e Resumos – Teoria e prática”, de F.W Lancaster, objetivando compreender as bases e regras da indexação, aprofundar ideias como as de descritores e sinônimas para que sejam acrescentadas na pesquisa.

Parte 2 – Pesquisa na legislação

A parte prática da pesquisa teve início com a identificação e seleção dos atos normativos correspondentes às classes

030 (Material) e 040 (Patrimônio) do Código de Classificação de Documentos de Arquivo: atividade-meio, elaborado pelo CONARQ. O objetivo dessa parte da metodologia foi buscar um entendimento maior sobre as atividades que produzem os documentos relacionados a essas classes.

As subclasses utilizadas, inicialmente, foram a de Material (030) e a de Patrimônio (040). Todas relacionadas aos documentos relativos às normas, regulamentações, diretrizes, procedimentos, estudos e/ou decisões de caráter geral específicos das funções atividade e atividades dessas subclasses.

Todos os textos legais encontrados foram retirados do sítio do Palácio do Planalto, mais especificamente da Subchefia de Assunto Jurídicos da Casa Civil. A primeira identificação era feita por meio do cabeçalho da lei, o qual descrevia, de forma objetiva e direta, o seu conteúdo geral. Feito isso, para confirmar se realmente havia associação entre a legislação e a subclasse, a pesquisa do termo específico expresso no grupo da subclasse era realizada no corpo da legislação encontrada. Abaixo, apresentamos uma pequena amostra dos resultados dessa etapa.

Relação da legislação

030 MATERIAL

030.1 CADASTRO DE FORNECEDORES

- Lei 8.666/1993

031 ESPECIFICAÇÃO. PADRONIZAÇÃO. CODIFICAÇÃO PREVISÃO. CATÁLOGO. IDENTIFICAÇÃO CLASSIFICAÇÃO

- Decreto 99.658/1990
- Decreto 6.087/2007

032 REQUISIÇÃO E CONTROLE DE SERVIÇOS REPROGRÁFICOS (INCLUSIVE ASSINATURAS AUTORIZADAS E REPRODUÇÕES DE FORMULÁRIO)

- Decreto 2.271/1997

033 AQUISIÇÃO (inclusive licitações)

- Lei 10.520/2002
- Lei 11.079/2004
- Lei 8.666/1993
- Instrução Normativa 205/SEDAP/1988

033.1 MATERIAL PERMANENTE

- Lei 8.666/1993

033.11 COMPRA (inclusive compra por importação)

- Lei 8.666/1993
- Instrução Normativa 205/SEDAP/1988

033.12 ALUGUEL. COMODATO. LEASING

- Lei 8.666/1993

033.13 EMPRÉSTIMO. CESSÃO. DOAÇÃO. PERMUTA

- Lei 8.666/1993
- Decreto 99.658/1990
- Instrução Normativa 205/SEDAP/1988

033.2 MATERIAL DE CONSUMO

- Lei 8.666/1993

033.21 COMPRA

- Lei 8.666/1993
- Instrução Normativa 205/SEDAP/1988

033.22 CESSÃO. DOAÇÃO. PERMUTA

- Lei 8.666/1993
- Decreto 99.658/1990
- Instrução Normativa 205/SEDAP/1988

033.23 CONFECÇÃO DE IMPRESSOS

034 MOVIMENTAÇÃO DE MATERIAL (permanente e de consumo)

- Decreto 99.658/1990
- Instrução Normativa 205/SEDAP/1988

Parte 3 – Mineração de textos

Após a pesquisa, identificação e associação da legislação referente às funções e atividades de cada subclasse, foi feita uma busca por Mineradores de Texto que melhor atendessem as necessidades da pesquisa.

Essa praticidade é vista, pois um software vai selecionar as palavras mais utilizadas em um determinado texto. O usuário só precisa definir alguns limites e/ou critérios de busca para o algoritmo de mineração. A mineração é um processo que envolve etapas como a preparação de dados, busca por padrões e avaliação do conhecimento (AMBRÓSIO e MORAIS, 2007, p.4).

Nessa procura por um minerador de textos, foram selecionadas três opções para serem utilizadas. A primeira opção, Text Mining Suite, aparentemente era apropriada, porém não foi possível o acesso, pois não se conseguiu a chave de segurança adequada para executar o software. A segunda, RapidMiner, continha muitas ferramentas que eram desnecessárias para o trabalho. E a terceira, Sobek, foi o escolhido como melhor opção. Esse software é uma ferramenta que foi elaborada por estudantes da Universidade Federal do Rio Grande do

Sul (UFRGS) para ser aplicado em funções educacionais e, portanto, é gratuito.

O Sobek foi selecionado, porque se encaixa nas características de um minerador de textos. Possui um campo destinado à inserção dos textos que irão ser minerados, algumas opções para a limitação e, assim padronização dos termos e, quando a mineração é realizada, os termos minerados são ligados as frases originais para, posteriormente, haver melhor avaliação do conhecimento. Primeiramente, ocorreu uma mineração na Lei 8.666, de 1993, para avaliar se realmente o Sobek era a melhor opção.

Por ser muito grande, a lei foi dividida em capítulos para ser minerada. Então, copiou-se os capítulos e os colocou na interface do minerador. Escolheu-se a opção de cinquenta conceitos que devem ser extraídos na média, uma frequência mínima de cinco repetições que precisam ocorrer para as palavras serem extraídas e uma lista de Stop Words² que se encontra em anexo. Pode-se fazer diversas combinações dessas limitações até que se chegue aos resultados esperados. Posteriormente, o botão “Extrair Grafo” foi selecionado.

Assim, foi gerada a relação das palavras usadas no texto, suas frequências e sua ligação com o contexto original. Este conhecimento serve para a verificar se o conhecimento extraído pode ser útil para o usuário final (AMBRÓSIO e MORAIS, 2007, p.4). Depois, o mesmo procedimento foi realizado com as demais leis, decretos e normas encontrados para as classes 030 e 040. Esses conceitos minerados são a representação das atividades da Administração Pública.

Parte 4 – Contextualização da mineração de textos

O próximo passo foi o estudo sobre a sinonímia dos termos que especificam as atividades da Administração Pública. Os

termos encontrados na mineração são a base da pesquisa, todavia, eles precisam ser juntados a outros termos, esses representando atividades.

Desse modo, foi utilizada uma lista, que apresentamos abaixo, contendo 59 termos que especificam as atividades. No entanto, essa lista não havia sido revisada, no que diz respeito à sinonímia. Por isso buscou-se os significados de todos os termos que foram analisados, e em caso de termos sinônimos um deles seria excluído. Nesse caso, por decisão da equipe, o termo mantido seria o termo mais formal. Ao fim dessa etapa, apenas a palavra “envio” foi excluída por não ser utilizada na legislação e por poder ser substituída por uma mais apropriada, como “encaminhamento”.

Segundo Ambrósio e Moraes (2007, p.6), ao utilizar a mineração de textos,

um usuário não solicita exatamente uma busca, mas sim uma análise de um documento. Entretanto, este não recupera o conhecimento em si. É importante que o resultado da consulta seja analisado e contextualizado para posterior descoberta de conhecimento.

Na prática, a mineração de textos define um processo que auxilia na descoberta de conhecimento inovador a partir de documentos textuais, que pode ser utilizado em diversas áreas do conhecimento.

A seguir, ocorreu a contextualização dos termos encontrados na legislação em suas respectivas classes e subclasses de Material e Patrimônio. Depois de todo o processo de mineração dos textos, viu-se a necessidade de analisar cada palavra, para saber se, de fato, é pertinente àquela legislação. Ou seja, os termos foram revistos e, como já estavam relacionados a cada grupo e subgrupo, foram contextualizados dentro da legislação específica, aquela legislação selecionada em etapa anterior.

Além disso, os termos substantivos, quando possível, foram passados para o singular e para o masculino, obedecendo as regras de indexação, de acordo com o Quadro 1.

Quadro 1 - Lista dos Termos que Especificam as Atividades e Lista de Stop Words

LISTA DOS TERMOS QUE ESPECIFICAM AS ATIVIDADES		
Abertura Afastamento Agradecimento Alteração Apresentação Aprovação Aquisição Armazenamento Atendimento Autorização Avaliação Classificação Comunicação Concessão Controle Convite Convocação Demissão Designação Determinação	Devolução Dispensa Divulgação Distribuição Elaboração Eleição Encaminhamento Entrega Envio Especificação Execução Exoneração Fatura Financiamento Fornecimento Identificação Inclusão Indeferimento Indicação Informação	Inscrição Instrução Levantamento Movimentação Orientação Pagamento Participação Prestação Prorrogação Realização Recolhimento Recomendação Reembolso Registro Serviço Solicitação Suspensão Visita Vistoria
LISTA DE STOP WORDS		
a à agora ainda além alguém algum alguma	algumas alguns ampla amplas amplo amplos ante antes	ao aos após aquela aquelas aquele aqueles aquilo

Fonte: autoria própria.

5 CONSIDERAÇÕES GERAIS

A ideia que moveu a construção deste trabalho foi a de tornar mais eficiente o processo de busca aos documentos de arquivo e às informações neles contidas. Partiu-se, portanto, da noção de que essa busca pode ser mais eficiente aliando a estrutura de um instrumento de classificação com termos definidos por meio da mineração nos textos legais, que regulamentam as grandes funções das atividades-meio da Administração Pública Federal. Entendemos, também, que a vinculação dos termos com as unidades de classificação pode potencializar os sistemas informatizados de gestão de documentos de arquivo.

Um dos principais objetivos da pesquisa foi de demonstrar a necessidade da Arquivística, com seu instrumental teórico-metodológico, de construir soluções para atender as demandas de acesso aos documentos de arquivo e as informações nele contidas em todas as suas possibilidades: na busca pelo documento em seu contexto de produção e pelo documento em sua individualidade.

Entendemos como desdobramento dessa pesquisa a possibilidade da operação de classificação poder utilizar-se da lista de termos controlados e ser utilizada por essa mesma lista. Ao classificar um documento em um sistema informatizado de gestão de documentos arquivísticos, o sistema pode oferecer ao usuário uma lista de termos controlados vinculados àquela unidade de classificação. E o sentido inverso também é possível. Ao identificar o conteúdo informacional do documento em uma lista de termos controlados, o sistema pode dar ao usuário possibilidades de classificação daquele conteúdo. Dessa forma, imagina-se que os erros identificados no preenchimento desses campos possam ser minorados com o uso dessa metodologia proposta no artigo em tela e que foi fruto de uma pesquisa científica.

REFERÊNCIAS

- CRUZ MUNDET, José Ramón. La gestión de documentos en las organizaciones. Madri : Pirámide, 2006.
- ARANHA, Christian; PASSOS, Emmanuek. A tecnologia de Mineração de Textos. RESE- Revista Eletrônica de Sistemas de Informação, n. 2, 2006. Disponível em: <189.16.45.2/ojs/index.php/reinfo/article/download/171/66> Acesso em: 30 mar. 2012.
- ARQUIVO NACIONAL. Dicionário Brasileiro de Terminologia Arquivística. Rio de Janeiro, Arquivo Nacional, 2005.
- BARTALO, Linete. Gestão em Arquivologia: abordagens múltiplas. In BARTALO, Linete; MORENO Nádina Aparecida. Londrina, EDUEL, 2008.
- BELLOTTO, Heloísa Liberalli. Arquivos permanentes: Tratamento documental. 4. ed. Rio de Janeiro, FGV, 2006.
- BRASIL. Lei nº 8.159, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Diário Oficial da União, Brasília, 09 jan. 1991, seção 1, p. 455.
- CONSELHO INTERNACIONAL DE ARQUIVOS. Multilingual Archival Terminology. Disponível em: <<http://www.ciscra.org/mat/>>. Acesso em: 13 abr. 2015.
- CONSELHO NACIONAL DE ARQUIVOS, Classificação, Temporalidade e Destinação de Documentos de Arquivo relativos às Atividades- Meio da Administração Pública. Rio de Janeiro, Conselho Nacional de Arquivos, 2001.
- GOLÇALVES, Eduardo Corrêa. Mineração de Texto: Conceitos e Aplicações Práticas. 2012. Disponível em: <http://www.devmedia.com.br/websys.5/webreader.asp?cat=2&artigo=4991&revista=sqlmagazine_105#a-4991> Acesso em: 30 mar. 2015.

KOBASHI, Nair Yumiko. Vocabulário Controlado: estrutura e utilização. Escola Nacional de Administração Pública. Disponível em: <http://www2.enap.gov.br/rede_escolas/arquivos/vocabulario_controlado.pdf>. Acesso em: 29 abr. 2015.

MORAIS, E. A. M; AMBRÓSIO, A. P. L. Mineração de Textos. Goiânia, Instituto de Informática, Universidade Federal de Goiás, 2007.

RODRIGUES, Ana Márcia Lutterbach. A teoria dos arquivos e a gestão de documentos. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 1, 2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362006000100009>. Acesso em: 26 mar. 2015.

SOUSA, Renato Tarciso Barbosa de. Alguns apontamentos sobre a classificação de documentos de arquivo. *Brazilian Journal of Information Science: Research Trends.*, Marília, v. 8, nº 1/2, 2014. Disponível em: <<http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/4246/3085>>. Acesso em: 30 mar. 2015.

SOUSA, Renato Tarciso Barbosa de. Os arquivos montados nos setores de trabalho e as massas documentais acumuladas na administração pública brasileira: uma tentativa de explicação. *Revista de Biblioteconomia de Brasília*, Brasília, v. 21, n. 1, jan./jun. 1997, p. 31-50.

SOUSA, Renato Tarciso Barbosa de Sousa, MESQUITA, Heloísa Carvalho, MARTINS, Larissa Marques Martins. Construção de vocabulário controlado para identificação do conteúdo informacional dos documentos acumulados pela atividade-meio da Administração Pública Federal. *Arquivo e Administração*, Rio de Janeiro, v. 9, n. 1, jan./jun. 2010.

SOUSA, Renato Tarciso Barbosa de; ARAÚJO JÚNIOR, Rogério Henrique de. A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística. *Ciência da Informação*, [S.l.], v. 42, n. 1, jan. 2013. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/2268>>. Acesso em: 29 abr. 2015.