

DOCUMENTOS DIGITAIS EDITÁVEIS, PADRÕES ABERTOS E INTEROPERABILIDADE DE DADOS: UM LEVANTAMENTO NOS SÍTIOS DE UNIVERSIDADES PÚBLICAS EM SANTA CATARINA

Ivan Paes José
Moisés Lima Dutra

Resumo: Este trabalho faz uma análise dos documentos digitais editáveis disponíveis nos sítios da Universidade Federal de Santa Catarina (UFSC) e da Universidade do Estado de Santa Catarina (UDESC), com o objetivo de verificar se estes estão em conformidade com as orientações do Governo Brasileiro, que sugere o uso do formato aberto *Open Document Format* (ODF) para a criação de documentos digitais editáveis em toda a esfera pública nacional. Realizou-se igualmente uma avaliação de interoperabilidade, que consistiu em se analisar se as informações contidas nos documentos coletados dos referidos sítios foram recuperadas sem nenhum tipo de perda, independentemente de extensão e *softwares* de escritório utilizados para acessar os documentos. Ao final, a pesquisa constatou que nenhuma das instituições está adotando as orientações propostas pelo Governo Brasileiro. Quanto à interoperabilidade dos dados, verificou-se que ocorreram perdas significativas nas informações dos documentos, com alguns deles, inclusive, tendo a sua leitura completamente impossibilitada.

Palavras-Chave: Documentos digitais editáveis. Interoperabilidade. Padrão ODF. Extensões Proprietárias. Extensões não-proprietárias.

1 INTRODUÇÃO

Assim como a imprensa foi responsável pelo aumento e disseminação do registro em papel no século XVIII, a Internet foi a responsável pelo aumento em escala exponencial na produção e disseminação do documento digital no fim do século XX. Devido a esse crescimento na produção e uso de documentos digitais, os profissionais da área de Ciência da Informação, já em meados dos anos 1960, alertavam que a produção científica exclusiva do meio digital se perderia, pois não havia a devida preocupação com a criação desses documentos e nem estratégias de preservação.

Grandes esforços têm sido empregados na melhoria das técnicas de preservação, transmissão, durabilidade e longevidade da informação para os usuários, de maneira a garantir, a preservação do conhecimento da humanidade.

Neste contexto, a recuperação da informação é um ponto central na preservação digital de documentos, pois, apesar de todos os avanços das técnicas para recuperar esta informação a partir de documentos digitais, este ainda é um problema com o qual nos deparamos frequentemente.

Esta pesquisa faz uma análise dos documentos digitais editáveis disponíveis nos sítios da Universidade Federal de Santa Catarina (UFSC) e Universidade do Estado de Santa Catarina (UDESC), para verificar se os mesmos estão em conformidade com as orientações do Governo Brasileiro, que sugere o uso de formato aberto¹. Procuramos, de forma específica:

- Quantificar os documentos editáveis que estão disponíveis para os usuários dos domínios ufsc.br e udesc.br;
- Efetuar um levantamento sobre as extensões de documentos existentes nos referidos sítios, apontando igualmente quais extensões são proprietárias e quais são não-proprietárias, categorizando os documentos editáveis por *softwares* de escritório; e
- Testar a interoperabilidade dos documentos com *softwares* de escritório proprietários e não-proprietários.

¹ Disponível em:

<http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra?codteor=547009&filename=PL+3070/2008>. Acesso em: 17/12/2013

2 REFERENCIAL TEÓRICO

Segundo Thomaz (2005, p. 11),

Apesar de o tema já ter sido abordado pela primeira vez em 1964, considera-se o marco inicial das discussões sobre os desafios do documento eletrônico para os arquivos, o relatório preparado pelo professor Robert Henri Bautier para o *International Council on Archives - ICA*, apresentado na *Conférence Internationale de la Table Ronde des Archives - CITRA*, realizada em Bonn, Alemanha, em 1971.

O relatório apresentava as seguintes conclusões:

1. Crescimento no uso das técnicas de computação pelos arquivos no final dos anos 1960;
2. Crescimento da quantidade de pesquisadores, utilizando os computadores para processar seus dados no período de 1965-1971, levando ao aumento da demanda por documentos eletrônicos na administração pública; e
3. Eliminação de documentos em suporte digital por parte de algumas instituições no mesmo período sem a intervenção de arquivistas em função de três fatores básicos: 1) falta de conhecimento sobre computação; 2) escassez de recursos para preservação do acervo e 3) obstáculos legais para o recolhimento desse tipo de documento (BAUTIER, 1971 *apud* THOMAZ, 2005, p. 12).

Após mais de 40 anos da publicação do referido relatório, ainda se enfrenta o problema exposto na terceira conclusão. Conforme Innarelli (2011, p. 83),

Atualmente, por conta da ignorância, de interesses ou negligência com a documentação, os profissionais da Tecnologia da Informação e Comunicação atuaram como responsáveis pela gestão e preservação da documentação digital, pois os profissionais da Ciência da Informação e suas áreas afins desconheciam, não só a importância das novas tecnologias para a gestão documental, mas a existência da documentação digital, neste meio tempo, os documentos digitais foram geridos por profissionais que não têm competência para entender a teoria em relação aos documentos gerados e gerenciados pelos sistemas de informação que desenvolveram e desenvolvem.

Esse problema no Brasil, segundo Innarelli (2011), ainda ocorre, pois os grupos que vêm discutindo acerca da preservação digital não têm a visibilidade, o investimento e a importância que deveriam ter.

Nossa sociedade preserva sua cultura através da transmissão do conhecimento e de sua consequente apropriação, sendo que essa apropriação, quando registrada e transmitida, gera um novo estado de conhecimento, ciclo que garante nossa evolução sócio-cultural. (INNARELLI, 2011, p. 73)

Márdero Arellano complementa, dizendo que:

A natureza dos documentos digitais está permitindo ampla produção e disseminação de informação no mundo atual. É fato que na era da informação digital se está dando muita ênfase à geração e/ou aquisição de material digital, em vez de manter a preservação e o acesso a longo prazo aos acervos eletrônicos existentes (MÁRDERO ARELLANO, 2008, p. 44).

Percebe-se reiteradamente que a recuperação da informação permanece sendo um tópico central na questão na preservação digital de documentos.

[...] as fotos de satélites da NASA dos anos 70, que documentavam as condições da floresta amazônica brasileira estão irremediavelmente perdidas, assim como dados do censo americano de 1960 (BORGHOFF et al, 2003 *apud* ALMEIDA, CENDÓN, SOUZA, 2012, p. 128).

Cita-se igualmente os problemas enfrentados pelas equipes de resgate no Tsunami de 2004, no Sudeste Asiático:

A operação de auxílio às vítimas do tsunami no final de 2004 foi marcada por todo tipo de

dificuldade. Percalços seriam inevitáveis em uma tragédia dessa proporção, que deixou um rastro de 200 000 mortes no Sudeste Asiático. Mas as equipes enviadas por mais de 60 países depararam com um problema inesperado: os relatórios e as planilhas trocados entre elas muitas vezes não abriam ou sofriam perdas de configuração que comprometiam seu entendimento. Como se não bastasse a dimensão da tragédia, os trabalhos de resgate e apoio às vítimas tiveram de enfrentar as complicações da tecnologia. (CESAR, 2007).

Na contramão da tendência mundial, a Biblioteca da Virgínia, nos Estados Unidos da América, tomou uma decisão surpreendente para arquivar seus documentos públicos.

Devido ao fato de inexistirem normas nacionais dos EUA concernentes aos meios nos quais os registros eletrônicos serão mantidos e considerando-se a volatilidade deste tipo de meio, registros em formatos eletrônicos não serão aceitos para armazenamento nesta instituição. Se ou quando os padrões de permanência de mídia eletrônica se tornarem uma realidade, esta política será revista em conformidade. Até esse momento, registros eletrônicos avaliados como de arquivo permanente devem ser mantidos na agência ou localidade convertidos em microfilme, papel, ou outro meio aceitável que atendam às normas nacionais para a permanente retenção de arquivo antes de serem transferidos para o Arquivo (VIRGINIA STATE LAWS ON OPTICAL IMAGES, 1999, tradução nossa).

Com uma atitude como a da Biblioteca da Virgínia, a manutenção do seu acervo dependerá de investimento muito alto, ou seja, não se tem mais o problema de recuperar as informações nos documentos digitais, porém existem vários outros problemas, como afirma Arellano:

O suporte físico da informação, o papel e a superfície metálica magnetizada se desintegram ou podem se tornar irrecuperáveis. Existem, ademais, os efeitos da temperatura, umidade, nível de poluição do ar e das ameaças biológicas; os danos provocados pelo uso indevido e o uso regular, as catástrofes naturais e a obsolescência tecnológica. (ARELLANO, 2004, p.15).

Diante deste cenário, o nosso olhar se volta para o Brasil, onde procuramos verificar se estamos tendo a preocupação de preservar nossos documentos de forma que no futuro estes possam ser utilizados sem nenhum problema.

3 METODOLOGIA

Caracterizou-se como uma pesquisa exploratório-descritiva no que tange aos objetivos e qualitativa em relação ao enfoque do problema. Exploratória porque segundo, Cervo e Bervian (1996, p. 49), “tem por objetivo familiarizar-se com o fenômeno ou obter nova percepção do mesmo e descobrir novas ideias”. Descritiva na medida em que procurou descrever as características (GIL, 2002, p. 42) dos documentos editáveis disponíveis nos domínios UFSC e UDESC, tendo como suporte as técnicas estatísticas, para a organização dos resultados em tabelas e gráficos com o intuito de resumir, descrever e apresentar os dados. Foi construída de forma qualitativa e quantitativa, sendo que a qualitativa tem o intuito de privilegiar as particularidades e descrever os significados e descobertas, e a quantitativa o de fornecer dados na forma numérica para a análise proposta.

Com base nos objetivos propostos, a pesquisa foi realizada por meio do acesso aos sítios da UFSC e UDESC, com o intuito de averiguar de que forma essas instituições estão liberando seus documentos editáveis para os usuários que irão à busca desses documentos.

Para a análise do conteúdo foram utilizadas as técnicas de Bardin (1977): i) Pré-análise; ii) Exploração do material; e iii) Tratamento dos resultados obtidos e interpretação.

Na pré-análise foi utilizado o buscador Google (www.google.com.br) como ferramenta para coleta de dados nos domínios udesc.br e ufsc.br, por meio da utilização dos operadores *ext* (busca arquivos específicos) e *site* (busca apenas em um *website* ou *domain* pré-determinado). Neste procedimento foram recuperados 133 documentos com a extensão *.odt* no domínio ufsc.br. A partir dos resultados da pré-análise, foi realizada a separação dos documentos digitais pela sua extensão proprietária ou não-proprietária. Nos mesmos moldes, este procedimento foi realizado nos sítios dos cursos de Biblioteconomia da UFSC e UDESC para identificar quais extensões estão sendo utilizadas.

Foi avaliada a interoperabilidade dos documentos digitais, por meio da abertura dos mesmos das suítes de escritório Microsoft Office 2010 e LibreOffice 3.6, com o intuito de averiguar o comportamento dos documentos. Para escolha da amostra foi utilizada a amostragem aleatória simples (GIL, 2002), todos os documentos coletados dos quatro sítios em uma pasta, no computador, e de forma aleatória foram selecionados 50 documentos entre proprietários e não-proprietários, e o sistema operacional utilizado foi o Windows 7 Ultimate 64 bits Service Pack 1.

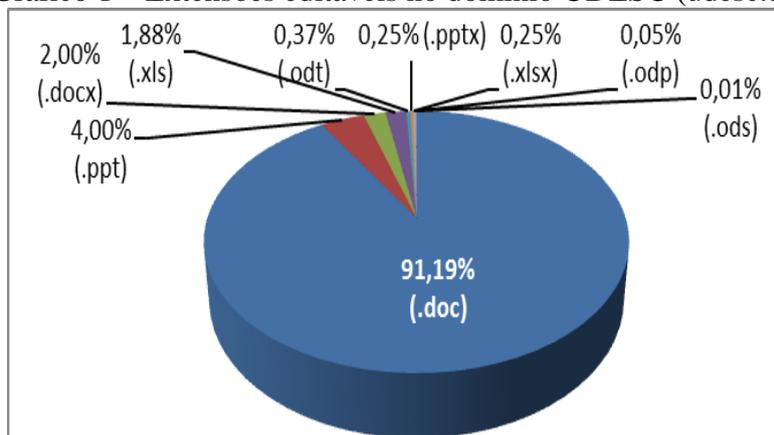
Após a abertura dos documentos digitais pelos *softwares* de escritório, analisou-se a disponibilização da informação, ou seja, se existiu perda de informação do documento, se ocorreu algum erro no momento da abertura e se algum *software* de escritório não conseguiu abrir algumas das extensões encontradas. Os resultados obtidos na pesquisa são apresentados e discutidos a seguir.

4 RESULTADOS

Apresentamos inicialmente o *corpus* da pesquisa, realizada nos sítios das instituições UFSC e UDESC. A partir da utilização dos operadores de busca *ext* e *site* do buscador Google (www.google.com.br) nos domínios UDESC (udesc.br) e UFSC (ufsc.br) obteve-se os seguintes resultados.

O Gráfico 1 mostra a distribuição das extensões encontradas no domínio da UDESC, para 14.695 documentos recuperados.

Gráfico 1 – Extensões editáveis no domínio UDESC (udesc.br)

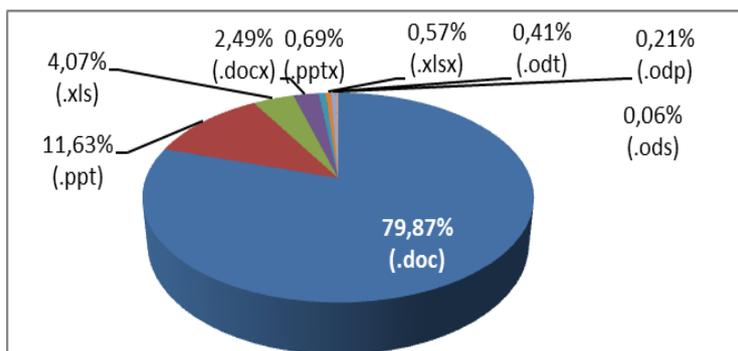


Fonte – Dados da pesquisa.

Observamos que 99,57% (14.632) das extensões são proprietárias e 0,43% (63) são extensões não-proprietárias.

Os resultados da UFSC (Gráfico 2) não são muitos diferentes dos obtidos na UDESC, embora exista uma quantidade superior de documentos disponível aos usuários da UFSC (32.677).

Gráfico 2 – Extensões editáveis no domínio UFSC (ufsc.br)

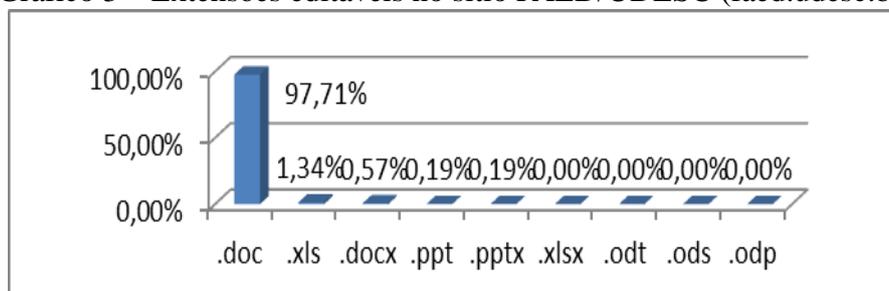


Fonte – Dados da pesquisa.

Apesar disso, a quantidade de documentos com extensões proprietária e não-proprietária se assemelham, e aqui também é observada uma enorme diferença entre as duas, 99,32% e 0,68%, respectivamente.

No Gráfico 3 são apresentados os resultados para os sítios dos cursos de Biblioteconomia.

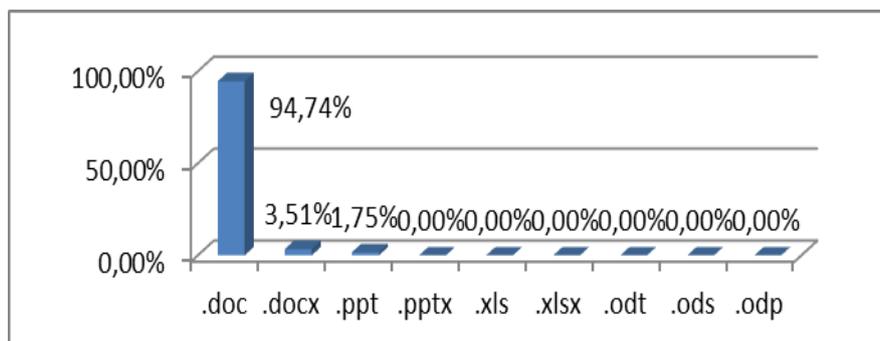
Gráfico 3 – Extensões editáveis no sítio FAED/UEDESC (faed.udesc.br)



Fonte – Dados da pesquisa.

Na UDESC, este sítio está situado dentro da Faculdade de Educação, Centro de Ciências Humanas e da Educação, cujo endereço eletrônico é faed.udesc.br. Na UFSC, o sítio equivalente está dentro do departamento de Ciência da Informação (CIN), cujo endereço é dptcin.paginas.ufsc.br (Gráfico 4).

Gráfico 4 – Extensões editáveis no sítio CIN/UFSC (dptcin.paginas.ufsc.br)

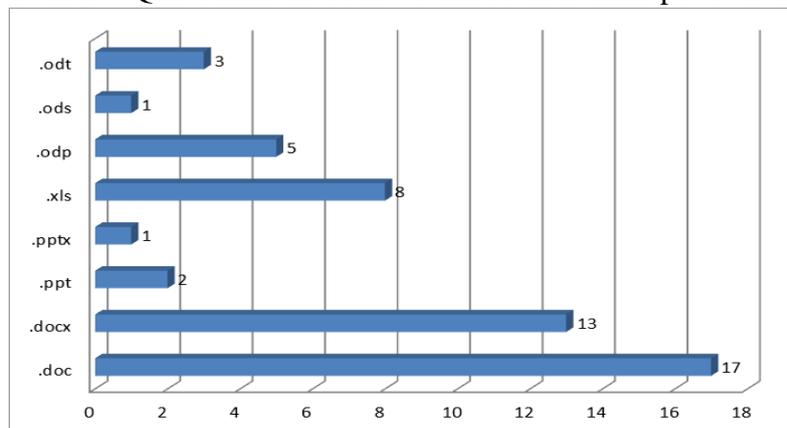


Fonte – Dados da pesquisa.

Observa-se que o sítio da FAED/UEDESC não possui extensão não-proprietária, sendo as extensões proprietárias 100% dos documentos dos 524 documentos recuperados. O departamento de Ciência da Informação da UFSC igualmente não possui extensão não-proprietária, para 57 documentos recuperados neste sítio.

A seguir, foram separados 50 documentos aleatoriamente dos sítios: ufsc.br, udesc.br, faed.udesc.br e dptcin.paginas.ufsc.br (Gráfico 5).

Gráfico 5 – Quantidade de documentos selecionados por extensões

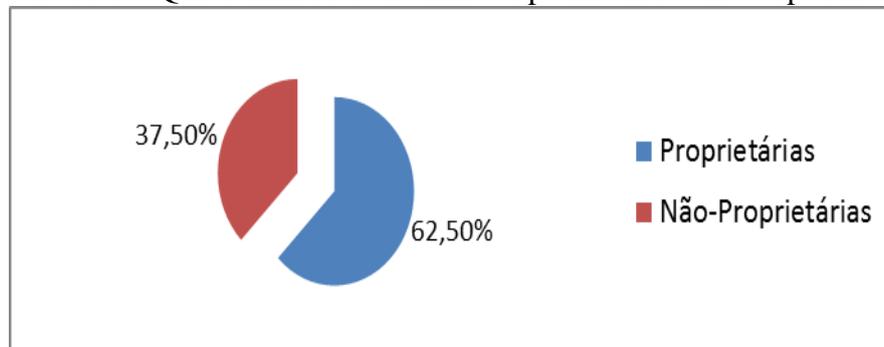


Fonte – Dados da pesquisa.

Estes documentos foram utilizados na avaliação da interoperabilidade, na qual todos os documentos foram abertos pelos softwares de escritório da Microsoft Office 2010 e Libreoffice, executados no sistema operacional Microsoft Windows 7 Ultimate SP 1 64 bits.

O Gráfico 6 mostra a relação de extensões obtida a partir da avaliação da interoperabilidade.

Gráfico 6 – Quantidade de extensões Proprietárias e Não-Proprietárias

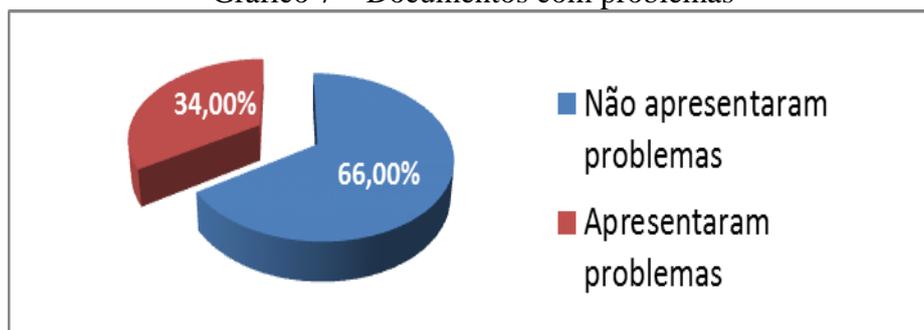


Fonte – Dados da pesquisa.

Foram identificadas 37,50% de extensões não-proprietárias e 62,50% de proprietárias.

O Gráfico 7 apresenta dados relativos à pesquisa sobre interoperabilidade entre o software proprietário *Microsoft Office 2010* e o software livre *LibreOffice 3.6*.

Gráfico 7 – Documentos com problemas

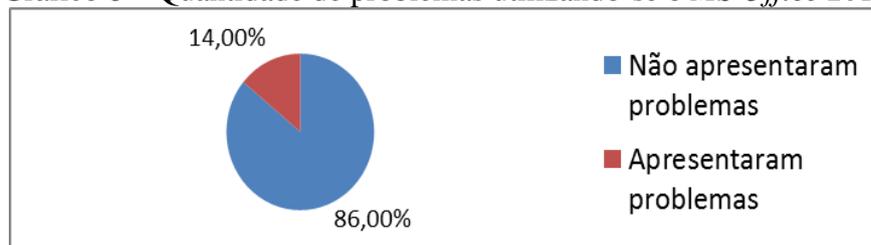


Fonte – Dados da pesquisa.

O Gráfico 7 traz a relação de documentos que apresentaram algum tipo de problema, não apenas de abertura/acesso, mas todos os problemas de formatação, animações que não funcionaram da forma esperada, imagens que não foram carregadas, fundo sem a cor desejada e textos apagados. No total, 34% desses documentos apresentaram os problemas acima relatados. Esses dados se referem aos documentos abertos nos dois *softwares* de escritório utilizados na pesquisa. Isso não significa que os documentos apresentaram problemas nos dois programas ao mesmo tempo. Por exemplo, em alguns casos ocorreram problemas com um arquivo no *Microsoft Office 2010* e nenhum no *LibreOffice* e vice-versa.

O Gráfico 8 mostra a quantidade de problemas encontrados exclusivamente durante a utilização do *software* de escritório *Microsoft Office 2010*. Os dados considerados englobam todos os tipos de problemas apresentados.

Gráfico 8 – Quantidade de problemas utilizando-se o *MS Office 2010*

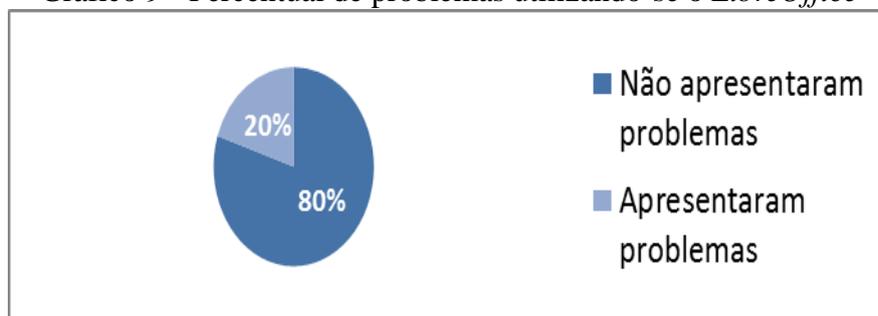


Fonte – Dados da pesquisa.

É importante destacar que, dos 14% de problemas obtidos com a utilização do *MS Office 2010*, 4% se referem exclusivamente aos documentos que sequer conseguiram ser abertos por este software.

O Gráfico 9 apresenta o percentual de problemas encontrados utilizando-se o *software* de escritório *LibreOffice*.

Gráfico 9 - Percentual de problemas utilizando-se o *LibreOffice*



Fonte – Dados da pesquisa.

Observa-se que este programa apresenta uma quantidade maior de problemas, com relação ao *MS Office*. Compreende-se este fato por meio da informação de que a quantidade de documentos com extensões proprietárias é bem superior a de documentos com extensões não-proprietárias, que são o foco do *LibreOffice*.

Apesar disso, todos os documentos foram abertos por essa suíte de escritório, ou seja, todos os dados ficaram legíveis e foram possíveis de serem interpretados. Diferentemente do *Microsoft Office 2010*, que apresentou problemas na abertura de 4% dos documentos, e indisponibilizou boa parte dos dados originais.

Fazendo uma relação com as recomendações do governo brasileiro, foi possível observar que as instituições pesquisadas possuem documentos em formato aberto, mais propriamente dito o ODF, mas que a quantidade destes ainda é pequena nos domínios das duas instituições, não chegando a 1% do total. A situação é mais crítica quando voltamos nossos olhares para os sites que hospedam os cursos de Biblioteconomia, pois nestes não foi recuperado nenhum documento em formato aberto. Diante disso, detectou-se uma inobservância tanto pela UFSC quanto pela UDESC das diretrizes propostas pelo governo para a utilização de documentos digitais de padrões abertos em instituições públicas.

5 CONCLUSÕES

A pesquisa mostrou que as instituições UFSC e UDESC estão disponibilizando diversos documentos editáveis em seus sites. Porém, analisando-se diretamente nos sites que hospedam os cursos de Biblioteconomia das referidas instituições, não foi observado o mesmo cenário que se apresentou nas pesquisas dos domínios. Dos 14.695 documentos editáveis recuperados no domínio da udesc.br, 3,57% pertencem ao sítio FAED/UDESC, enquanto que dos 32.677 documentos editáveis recuperados no domínio ufsc.br menos de 1% (0,17%) pertencem ao sítio CIN/UFSC.

Este trabalho mostrou que, na sua maioria, as extensões proprietárias estão presentes em mais de 98% dos documentos das duas instituições, sendo as extensões não-proprietárias muito pouco utilizadas. Esse dado é preocupante, pois significa que a maioria das informações armazenadas dependem de uma empresa para serem interpretadas de forma que se tenha 100% de garantia de que o documento que as contém será recuperado sem perda de dados.

Pôde-se avaliar também que a interoperabilidade dos documentos dos referidos sites é possível e está sendo feita, apesar de que alguns documentos que utilizam extensões não-proprietárias não foram abertos no *software* de escritório da *Microsoft Corporation*. Porém, todos os documentos foram abertos pelo *software* de escritório *LibreOffice*. No entanto, alguns documentos com extensões proprietárias perderam algumas de suas características, que entretanto, apesar de tudo, impossibilitaram por completo a recuperação das informações destes documentos.

Pode-se concluir que ambas as instituições não estão em conformidade com as diretrizes do

Governo Brasileiro para criação de documentos editáveis, através da utilização de padrões abertos, pois, conforme se verificou, muito pouco ainda é utilizado deste tipo de extensão. Diante disso, se faz necessário um alerta a essa questão nas universidades analisadas e nas instituições de ensino públicas em geral, pois deixar quase todo o seu conhecimento e trabalho “refém” de uma empresa talvez não seja a melhor estratégia de preservação a ser adotada.

REFERÊNCIAS

ALMEIDA, Maurício Barcellos; CENDÓN, Beatriz Valadares; SOUZA, Renato Rocha. Metodologia para implantação de programas de preservação de documentos digitais a longo prazo. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 17, n. 34, p. 103-130, 2012. Disponível em: < <http://www.brapci.ufpr.br/download.php?dd0=20369>>. Acesso em: 17/12/2013.

ARELLANO, Miguel Angel. **Preservação de documentos digitais**. Ci. Inf. [online]. 2004, vol.33, n.2, pp. 15-27. Disponível em: < <http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf> >. Acesso em: 17/12/2013.

BARDIN, Laurence. **Análise de conteúdo**. Lisboa: Edições 70, 1977.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino. **Metodologia Científica**. 4. ed. São Paulo: Makron Books, 1996. 209 p.

CESAR, Ricardo. O esperanto dos arquivos. In: **Grande Decisões**. Revista Exame. Edição 889. 2007. Disponível em: <<http://exame.abril.com.br/revista-exame/edicoes/0889/noticias/o-esperanto-dos-arquivos-m0125023>>. Acesso em: 17/12/2013.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.

INNARELLI, Humberto Celeste. PRESERVAÇÃO DIGITAL: A INFLUÊNCIA DA GESTÃO DOS DOCUMENTOS DIGITAIS NA PRESERVAÇÃO DA INFORMAÇÃO E DA CULTURA. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 8, n. 2, p.72-87, 2011. Semestral. Disponível em: < <http://www.sbu.unicamp.br/seer/ojs/index.php/rbci/article/view/487>>. Acesso em: 17/12/2013.

MÁRDERO ARELLANO, Miguel Ángel. **Critérios para a preservação digital da informação científica**. 2008. 354 f. Tese (Doutorado em Ciência da Informação)–UnB, Brasília, 2008. Disponível em: < http://eprints.rclis.org/12649/1/Tese_Miguel_Angel_Márdero_Arellano.pdf >. Acesso em: 17/12/2013.

THOMAZ, Katia P. Gestão e preservação de documentos eletrônicos de arquivo: Revisão de literatura - Parte 1 / Electronic record: problem evolution – Part 1. **Arquivística.net**, Rio de Janeiro, v. 1, n. 2, p. 8-30, jul./dez. 2005.

VIRGINIA STATE LAWS ON OPTICAL IMAGES. **The Librery of Virginia** [Internet]. Virginia - EUA: The Librery of Virginia; 1999 Disponível em: <<http://www.archiveindex.com/laws/law-va.htm>>. Acesso em: 17/12/2013.

EDITABLE DIGITAL DOCUMENTS, OPEN STANDARDS, AND DATA INTEROPERABILITY: A SURVEY ON SANTA CATARINA'S PUBLIC UNIVERSITY WEBSITES

Abstract: *This work undertakes a survey on the editable digital documents available on the websites hosted by the Federal University of Santa Catarina (UFSC) and by the State University of Santa Catarina (UDESC), aiming to check whether the editable digital documents are in accordance or not with the guidelines of the Brazilian Federal Government, which suggests the use of Open Document Format (ODF) for creating editable digital documents throughout the national public sector. Furthermore, a survey on interoperability was carried out, which consisted of analyzing the consistence of the documents that have been retrieved in both websites, concerning any loss of information, regardless of the files' extensions and the used software applications. Finally, we can consider that both UFSC and UDESC are not adopting the guidelines proposed by the Brazilian Federal Government. Regarding the interoperability, we have found out that there were significant losses of information in the documents, and in some cases they were even completely unable to be opened and read.*

Keywords: *Editable digital documents. Interoperability. ODF. Proprietary extensions. Non-proprietary extensions.*

Ivan Paes José

Email: ivan.paes@gmail.com

Moisés Lima Dutra

Professor Adjunto da Universidade Federal de Santa Catarina, Departamento de Ciência da Informação. Doutor em Computação pela Universidade de Lyon 1, França (2009). Mestre em Engenharia Elétrica (2005) e Bacharel em Computação (1998) pela Universidade Federal de Santa Catarina. Suas atuais linhas de pesquisa são Web Semântica, Ontologias e Ferramentas Colaborativas. Está vinculado ao grupo de pesquisa ITI-RG (Inteligência, Tecnologia e Informação - Research Group).

Email: moises@cin.ufsc.br

<p>Recebido em: 18-12-2013 Aceito em: 25-02-2014</p>
--