

ONTOLOGIA DE DEPENDÊNCIA TECNOLÓGICA DE DOCUMENTOS DIGITAIS: INSTRUMENTO DE APOIO À PRESERVAÇÃO DIGITAL

Eloi Juniti Yamaokaⁱ
Fernando Ostuni Gauthierⁱⁱ

Resumo: Este artigo apresenta o uso da ontologia para modelar a dependência tecnológica de objetos digitais. Na computação e na ciência da informação, ontologia é um artefato que permite a modelagem de conhecimento em algum domínio e possibilita especificar um vocabulário para fazer asserções que podem ser utilizadas por software. Para a que possa ser compreensível por humanos, o documento digital exige o uso de tecnologia. Essa dependência tecnológica requer um cuidadoso acompanhamento da obsolescência dessas tecnologias utilizadas para a criação e visualização do documento digital. O ambiente tecnológico de uma organização que faz uso intensivo da tecnologia da informação e comunicação pode ser constituído de milhares de componentes inter-relacionados e a tarefa de verificar o impacto da desativação de determinada tecnologia é muito trabalhosa. Visando reduzir os efeitos desse problema, foi desenvolvida a ontologia DeTec, com o uso da Web Ontology Language (OWL). Para a modelagem foi utilizada a plataforma Protégé da Stanford University). A ontologia DeTec é constituída de oito classes principais que permite a identificação de quais tecnologias um objeto digital tem dependência. Essa funcionalidade possibilita evitar a desativação equivocada de uma tecnologia e a consequente perda dos objetos dela dependente.

Palavras-chave: Preservação Digital. Ontologia. Formatos de Arquivos. Dependência Tecnológica.

ONTOLOGY OF TECHNOLOGICAL DEPENDENCE OF DIGITAL DOCUMENTS: THE DIGITAL PRESERVATION SUPPORTING TOOL

Abstract: *This article introduces the use of ontology for modeling the technological dependence of digital objects. In computing and information science, ontology is an artifact designed to enable the knowledge modeling in some area and lets you specify a vocabulary to make assertions that can be used by software. To be consumed by humans in an intelligible form, the digital document requires the use of technology. This technological dependency requires careful management to track the obsolescence of technologies used for creating and viewing of digital document. The technological environment of an organization that makes intensive use of information and communication technology consists of thousands of interrelated components and the task of verifying the impact of disabling certain technology is very laborious. For this issue, was developed the DeTec ontology, with the use of the Web Ontology Language (OWL). For the modeling issues was used the Protégé platform of Stanford University). The DeTec ontology is made up of eight main classes that enables the identification of the dependencies between a digital object and the related technologies. This feature can prevent mistaken deactivation of a technology and the consequent loss of her dependent objects.*

Keywords: *Digital Preservation. Ontology. File formats. Digital documents technological dependency.*



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/)

ⁱ Universidade Federal de Santa Catarina e Serviço Federal de Processamento de Dados - eloijy@yahoo.com.br ; eloi.yamaoka@serpro.gov.br

ⁱⁱ Universidade Federal de Santa Catarina – gauthier@egc.ufsc.br

1 INTRODUÇÃO

Com o desenvolvimento e popularização da tecnologia da informação e comunicação (TIC), a produção e a disseminação da informação tornaram mais intensas e a informação assumiu papel central, acarretando profundas transformações na sociedade. O surgimento dessa sociedade informatizada e permanentemente conectada afetou diretamente o processo de produção e disseminação da informação científica, segmento onde a informação e o conhecimento são os principais insumos e produtos.

Na sociedade digital, a preservação do conteúdo se tornou um desafio, não só pela fragilidade das mídias magnéticas e óticas, mas também pela obsolescência da tecnologia.

Uma diferença fundamental do documento em papel para o documento digital é que, no primeiro, é possível consumir a informação sem o uso de recursos adicionais, no documento digital é necessário uma tecnologia que consiga decodificar os bits em sinais inteligíveis a humanos. A falta dessa tecnologia pode levar a perda total do conteúdo.

A maioria dos produtos e serviços do mercado atual foi criada nos últimos cinco anos e há carência de métodos comprovados que garantam: a) que as informações serão preservadas; b) que essas informações serão acessadas usando tecnologias disponíveis; e c) que qualquer informação acessível seja autêntica e confiável (CHEN, 2001).

A necessidade de preservação sempre existiu, mas a preservação digital tem diferenças fundamentais em relação à preservação dos documentos em papel. Enquanto no papel, a preservação do suporte permita a preservação do conteúdo e com frequência ocorrerem perdas parciais, no caso do objeto digital, a preservação do suporte é somente uma pequena parte do processo e um objeto digital corrompido, leva frequentemente à perda total.

Este trabalho aborda um dos problemas da preservação digital que é a obsolescência da tecnologia que pode ocasionar a eliminação indesejada de grandes quantidades de documentos digitais produzidas em tecnologias legadas. Aborda o problema no contexto da comunicação eletrônica e da constante transformação do universo digital. Introduce conceitos de formato de arquivo e por fim, apresenta a ontologia DeTec.

2 O CONTEXTO: A COMUNICAÇÃO ELETRÔNICA

A comunicação eletrônica alcançou significativo avanço e consolidação, introduzindo novo paradigma na comunicação científica. Entre as características desse novo paradigma estão: o incremento da comunicação entre pesquisadores, a desintermediação, a velocidade de disseminação e a redução das limitações expostas pelas dimensões espacial e temporal. A quantidade de repositórios digitais aumenta impulsionada pela facilidade e baixo custo de implantação.

O movimento *Open Access* tem importante relação com essas transformações. A desintermediação permite o autodepósito da produção científica pelos autores e a quebra das limitações temporais e espaciais permite o depósito independente da distância geográfica entre autor e o repositório e em qualquer dia e horário.

A facilidade e o baixo custo levaram ao crescimento global da quantidade de repositórios. Diretórios de repositórios como o *Directory of Open Access Repositories* (OpenDOAR)³ da Universidade de Nottingham e o *Registry of Open Access Repositories* (ROAR)⁴ da Universidade de Southampton apresentam evidências desse crescimento, especialmente nos últimos anos:

- no início de 2006 o OpenDOAR registrava em seu banco de dados apenas 105 repositórios e em janeiro de 2012 mais de 2.100, um crescimento de 1900% em seis anos;
- os registros do ROAR por sua vez, mostra que do total de 2.709 repositórios, 77,9% (2110 repositórios, foram cadastrados nos últimos seis anos.

Com a introdução da TIC em praticamente todos os segmentos da sociedade, a percepção é que as transformações ocorrem com maior frequência. Enquanto as grandes mudanças na comunicação impressa ocorriam em períodos de até um século, a comunicação eletrônica sofre importantes mudanças em uma década (MEADOWS, 2000, p.27).

³ www.opendoar.org

⁴ roar.eprints.org

3 O UNIVERSO DIGITAL E A METAMORFOSE DA TECNOLOGIA

O universo digital sofre de incessante metamorfose tecnológica, com constante desenvolvimento e introdução de novas tecnologias. Gordon Moore em 1965 previu que o número de transistores em um microprocessador dobraria a cada ano (MOORE, 1965). Essa previsão ficou conhecida como a Lei de Moore que é citada, debatida e revista nos últimos 40 anos. Entre 1986 e 2007 estima-se que a capacidade computacional cresceu à taxa de 58% ao ano e a capacidade de telecomunicação, 28% ao ano (HILBERT; LÓPEZ, 2011).

As novidades da TIC trazem facilidades e recursos adicionais aos usuários, porém é também uma ameaça às informações produzidas e armazenadas em tecnologias antecessoras. A obsolescência tecnológica pode ser tanto de hardware quanto de software. Quando o hardware torna-se obsoleto e de difícil manutenção e acesso, o conteúdo das mídias que dependem dessa tecnologia pode ficar inacessível, mesmo que conservado em boas condições.

O disco flexível, conhecido também como disquete, foi introduzido pela IBM em 1971⁵, tornou-se o principal meio de armazenamento de informação para pequenos sistemas. Durante os cerca de 40 anos de vida, teve incremento em velocidade e capacidade de armazenamento e redução de tamanho. O disco flexível tornou-se obsoleto e o anúncio oficial da interrupção da sua produção pela Sony em 2010⁶, foi marcado como o fim do disquete. O acesso às informações armazenadas nesses meios está limitado, pois os atuais computadores não dispõem mais dos dispositivos de gravação e leituras dessas mídias.

De forma semelhante, softwares tornam-se obsoletos, seja pela descontinuidade ou pela incompatibilidade com novas gerações de sistemas operacionais. Nesse caso, mesmo que o documento digital tenha sido copiado para novos suportes, é a capacidade de leitura e apresentação que se perde.

Um documento produzido em um determinado processador de texto, frequentemente fica dependente desse software, que por sua vez, depende do ambiente operacional (sistema operacional, softwares, componentes adicionais, fontes, etc.).

O universo digital é constituído por uma diversidade de tipos de objetos digitais como texto, fotos, diagramas, filmes, áudio e programas executáveis.

Além do constante aumento de resolução, os objetos digitais são associados e incorporados resultando em objetos digitais mais complexos, a exemplo de um livro

⁵ IBM Archives http://www-03.ibm.com/ibm/history/exhibits/storage/storage_chrono20.html

⁶ PCWorld http://www.pcworld.com/businesscenter/article/194959/sony_to_end_floppy_disk_production.html

multimídia, que tem a incorporação de texto, som, filmes e programas que interage com o leitor.

O contínuo crescimento do tamanho do universo digital com a produção de novos objetos e as sucessivas cópias realizadas por usuários aumenta exponencialmente a complexidade e o custo da preservação digital.

Pode-se ter uma ideia do tamanho do universo digital nas estatísticas do YouTube, o popular website de compartilhamento de vídeos. Na sua página de estatística, o YouTube divulga que a cada minuto 60 horas de vídeos são carregadas, isso é, em um mês são publicados mais horas de vídeos do que a produção de 60 anos das três maiores emissoras dos EUA⁷.

4 FORMATO DE ARQUIVO

Um formato define a forma como o arquivo (o objeto digital) está codificado e a informação organizada. Um formato é identificado pela extensão do nome do arquivo, como o txt, pdf, rtf, html, etc. A extensão possibilita o sistema identificar que programa está definido para leitura e apresentação daquele arquivo.

Muitos formatos são definidos como padrão pela International Organization for Standardization (ISO), por exemplo, os formatos:

- PDF 1.7 - ISO 32000-1:2008
- PDF/A-1 - ISO 19005-1:2005
- PDF/A-2 - ISO 19005-2:2011
- ODF - ISO 26300:2006
- OpenXML - ISO 29500:2008
- TIFF - ISO 12639:2004

Em qualquer abordagem de preservação digital, a questão dos formatos de arquivos é um dos principais pontos (BARVE, 2007) e várias iniciativas e projetos de pesquisa surgem com essa preocupação.

O PRONOM⁸ do Arquivo Nacional do Reino Unido mantém um sistema de informação na web com informações técnicas sobre formatos de arquivos e sobre os softwares

⁷ http://www.youtube.com/t/press_statistics

⁸ <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

que o suporta. Adicionalmente, o PRONOM disponibiliza gratuitamente uma ferramenta para apoiar a identificação dos formatos de arquivos, denominado *Digital Record Object identification* (DROID⁹) .

A Universidade de Harvard esteve à frente de projeto como o JSTOR/Harvard Object Validation Environment (JHOVE¹⁰) e o Global Digital Format Registry (GDFR¹¹). O JHOVE foi desenvolvido em parceria com o JSTOR¹², a biblioteca digital que agrega publicações acadêmicas a fim de preservar e prover o acesso.

O JHOVE é uma ferramenta com funções de identificação, validação e caracterização de objetos digitais, isto é, para responder três principais questões: 1) Qual o formato do objeto digital? 2) O objeto digital é do formato suposto? 3) Quais as propriedades mais importantes do formato identificado?

O GDFR é um projeto da Biblioteca da Universidade de Harvard com o escopo de prover um serviço distribuído para armazenar, descobrir e fornecer informações de representação de formatos digitais. O GDFR foi desenvolvido numa arquitetura conhecida como *peer-to-peer* que possibilita o compartilhamento de recursos tecnológicos dos computadores ligado à sua rede.

Uma parceria entre os projetos GDFR e o PRONOM busca a construção do Unified Digital Format Registry (UDFR¹³).

A Biblioteca do Congresso Americano mantém um website¹⁴ com informações técnicas sobre formatos relevantes para o acervo digital da biblioteca, incluindo fatores de sustentabilidade que possam afetar a preservabilidade de longo prazo de um determinado formato.

Na preservação digital o conceito de formato de preservação tem grande importância.

O Digital Preservation Task Group do CENDI, um grupo formado por gestores de informação científica e técnica do governo americano, define que formatos de preservação são aqueles com melhor possibilidade de preservação, incluindo a capacidade de capturar o material em um arquivo, exibir e disseminar a informação agora e no futuro (CENDI DIGITAL PRESERVATION TASK GROUP, 2007).

Na proposta de Christensen (2004) para o armazenamento de longo prazo, o formato: 1) deve ser simples para descrever, compreender e implementar; 2) não deve depender de

⁹ <http://droid.sourceforge.net/>

¹⁰ <http://hul.harvard.edu/jhove/>

¹¹ <http://www.gdfr.info/>

¹² <http://www.jstor.org/>

¹³ <http://www.udfr.org/>

¹⁴ <http://www.digitalpreservation.gov/formats/index.shtml>

hardware específico; 3) não deve depender de sistemas operacionais específicos; 4) não deve depender de software proprietário; e 5) ser tolerante a falhas.

5 A PRESERVAÇÃO DE OBJETOS DIGITAIS

Todo processo de preservação implica no conhecimento do objeto a ser preservado, como a sua composição, características e fragilidades. Um ponto comum dos processos de preservação é manter o objeto inalterado para mantê-lo mais próximo possível do estado original.

Na preservação digital, por um lado, a pretensão é mantê-los em estado autêntico, por outro lado, é necessário permitir que o objeto digital seja acessado usando a tecnologia disponível no momento, o que inevitavelmente implica em alguma alteração – é o paradoxo da preservação digital (CHEN, 2001; THIBODEAU, 2002).

A resposta à pergunta “o que é objeto digital” tem mudando junto com a evolução da TIC, passando de simples sequência de caracteres alfanuméricos para complexos objetos multimídias com incorporação de sons e vídeos tridimensionais.

Um documento digital como um relatório, produzido em um processador de texto, com incorporação de fotografias e diagramas é constituído por vários objetos digitais. O relatório pode ter sido gravado no formato RTF, com fotos e diagramas incorporados no formato JPG.

Todos os objetos digitais herdam suas propriedades de três classes: objeto físico, objeto lógico e objeto conceitual. Como objeto físico, o objeto digital é simplesmente uma inscrição de sinais sobre uma mídia; um objeto lógico é um objeto que é reconhecido e processado por software; e um objeto conceitual é o objeto como é reconhecido e entendido por uma pessoa, ou em alguns casos reconhecido e processado por uma aplicação de computador (THIBODEAU, 2002).

Essa definição de objeto digital em três níveis tem duas importantes consequências para o processo de preservação digital. O primeiro, é que um objeto conceitual pode ter diferentes codificações, e o segundo, é que diferentes codificações pode preservar as características essenciais do objeto conceitual (THIBODEAU, 2002). Por exemplo, um determinado relatório que gravado no formato RTF e no formato PDF, possui somente um

objeto conceitual e dois objetos lógicos, pois em cada formato a representação lógica é diferente.

A preservação digital implica na preservação da habilidade de reproduzir os objetos digitais com o uso de software para processar as inscrições armazenadas. Para que possa ser consumido de forma inteligível por humanos, o documento digital exige o uso de tecnologia (FERREIRA, 2006). Portanto, manter preservado somente objeto físico e lógico não garante a obtenção do objeto conceitual. É necessário também preservar o software para traduzir o objeto lógico, ou converter o objeto para um formato atualizado.

6 A DEPENDÊNCIA TECNOLÓGICA DO ACERVO DIGITAL

Atualmente os usuários de TIC têm dependência de uma miscelânea de componentes tecnológicos, produzidos por inúmeros fornecedores, tornando o ambiente extremamente complexo. Os documentos digitais produzidos e recebidos em uma instituição são diretamente afetados por essa complexidade.

Uma biblioteca que disponibiliza o seu acervo digital aos usuários com infraestrutura própria de TIC, necessita adquirir e instalar, não só os equipamentos (servidores, estações, dispositivos de rede), mas uma miscelânea de software para garantir o acesso com desempenho e segurança no nível aceitável.

Esse acervo digital pode ser armazenado em vários computadores, em diferentes mídias óticas e magnéticas com variados tipos e em diferentes formatos de arquivos. Gerir todos os componentes tecnológicos para garantir o acesso ao acervo ao longo do tempo é complexo e trabalhoso.

A operação e manutenção do ambiente tecnológico tornam-se mais caro, proporcionalmente ao porte da infraestrutura montada, e com a diversidade de tecnologia que deve ser mantida para garantir o acesso ao acervo. A formação de um legado de formatos de arquivos, leva à necessidade de manter um legado de tecnologia e a simples migração do acervo para uma nova tecnologia pode acarretar em problemas de autenticidade, considerando que todos os processos de migração geram perdas e possibilitam adulteração do objeto, mesmo sem a intenção, que podem ser percebidos muitos anos depois, dificultando ou até impossibilitando a sua recuperação.

A decisão de adquirir e internalizar uma nova tecnologia, geralmente ocorre com mais tranquilidade e entusiasmo, considerando que a nova tecnologia irá proporcionar novas possibilidades e acesso a serviços e objetos que incorporam novos recursos. Por outro lado, a decisão de descontinuar uma tecnologia e a migração do conteúdo legado para outra tecnologia é uma decisão difícil, considerando os riscos inerentes a essa decisão.

Visando trazer contribuição no processo de busca de solução para o problema do mapeamento e gestão da dependência tecnológica, foi concebida a ontologia DeTec, com o uso do editor de ontologia Protégé¹⁵, na versão do Protégé-OWL 3.4.8 (*Build 629*).

7 ONTOLOGIA DeTec

Na computação e na ciência da informação, ontologia é um artefato projetado para permitir a modelagem de conhecimento em algum domínio e permite especificar um vocabulário para fazer asserções que podem ser utilizadas por software. Uma ontologia define um conjunto de primitivas representacionais, que são as classes, atributos e relacionamentos (GRUBER, 2009). Os indivíduos são as instâncias de classes que representam os objetos no domínio (HORRIDGE et al., 2007). Uma base de conhecimento é a ontologia com um conjunto de indivíduos (NOY; MCGUINNESS, 2001).

O Protégé é uma plataforma de código aberto, desenvolvida no *Stanford Center for Biomedical Informatics Research* da *Stanford University School of Medicine*. É um conjunto de ferramentas para a construção de modelos de domínio e aplicações baseada em conhecimento com ontologias, com o suporte à criação, visualização e manipulação de ontologias¹⁶.

A versão Protégé-OWL, permite a construção de ontologias na linguagem OWL – *Web Ontology Language*, uma recomendação do consórcio W3C¹⁷.

Ontologias OWL poder ser categorizadas em três sublinguagens: OWL-Lite, a menos expressiva; a OWL Full, a mais expressiva e a OWL-DL com expressividade entre a Lite e a Full (HORRIDGE et al., 2007).

Pela possibilidade da realização de inferências, a opção para a ontologia DeTec foi a OWL-DL.

¹⁵ <http://protege.stanford.edu/>

¹⁶ <http://protege.stanford.edu/overview/>

¹⁷ <http://www.w3.org/TR/owl-overview/>

Para a construção da DeTec, foram seguidos os passos tradicionalmente utilizados para construção de ontologias. O domínio a ser modelado já estava definido, portanto, o processo seguido foi:

- identificação das classes e a hierarquização;
- identificação e definição das propriedades das classes;
- estabelecimento das restrições;
- inclusão de instâncias; e
- checagem da consistência da ontologia.

A dependência tecnológica do objeto digital ficou explicitada em oito classes principais, apresentados na figura 1, na visão da janela de classes do Protégé.

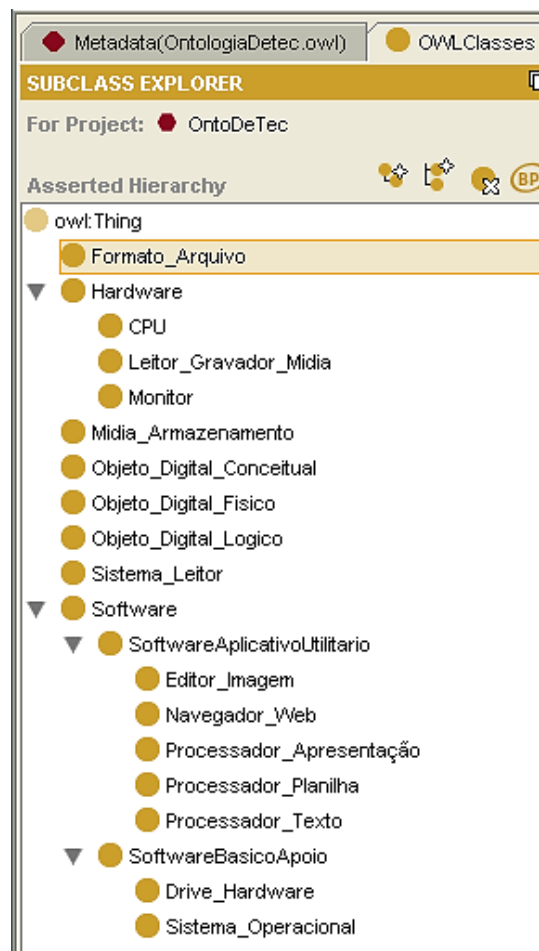


Figura1 – Classes da ontologia Detec

O objeto digital foi representado em três classes, adotando os conceitos propostos por Thibodeau (2005).

A classe Sistema_Leitor é um elemento conceitual, é o agregador de componentes necessários para interpretar o objeto digital lógico para apresentação do objeto digital conceitual.

O Sistema_Leitor permite mapear o hardware, o software, o formato de arquivo e a mídia de armazenamento, logo, um determinado objeto digital deverá necessariamente se relacionar com um sistema leitor através do formato do arquivo.

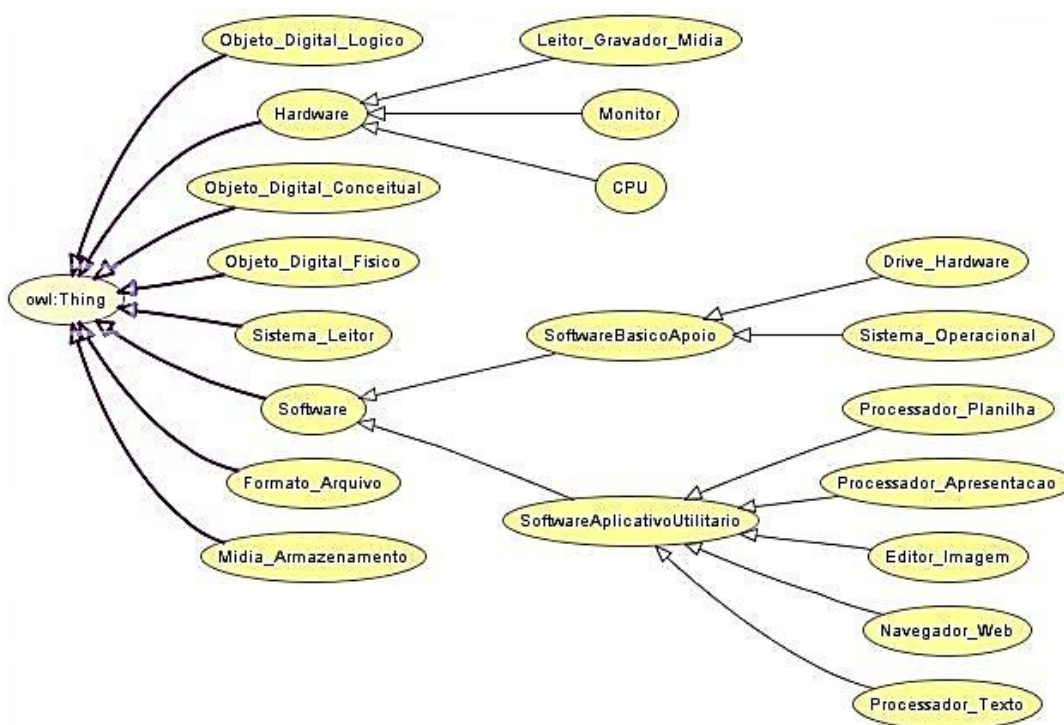


Figura2 – Ontologia DeTec visualizado pelo OWLViz

A figura 2 apresenta a visualização da taxonomia da DeTec utilizando o módulo de extensão OWLViz.

Com a definição das propriedades, isto é, das relações entre as classes, e das restrições, ficou evidente a importância do objeto digital lógico no modelo.

A ontologia assume as seguintes definições:

1. um objeto digital conceitual está representado em um ou mais objetos digitais lógicos;
2. um objeto digital lógico está armazenado em um (obrigatoriamente) ou mais objeto digital físico;
3. um objeto digital lógico tem necessariamente um formato de arquivo.

As, classes, propriedades e restrições podem ser vista na figura 3, num diagrama da ontologia DeTec.

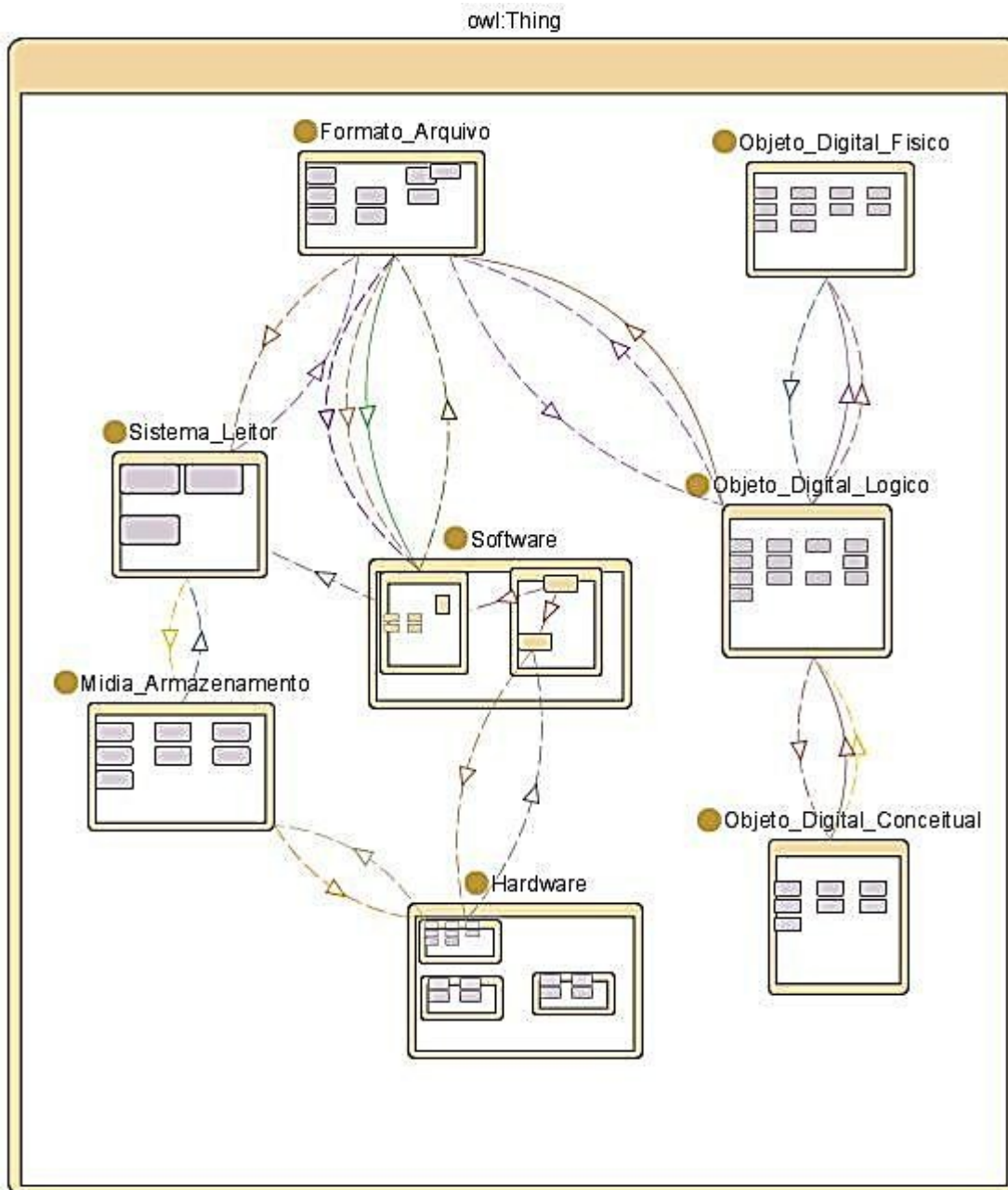


Figura 3 – Ontologia DeTec

O diagrama foi gerado através do Jambalaya, que é um módulo de extensão do Protégé, desenvolvida pela universidade de Victoria.

É visível na figura 3 que as maiores ocorrências de relacionamentos ocorrem em três classes: formato de arquivo, objeto digital lógico e software. A quantidade maior de relacionamentos é um indicativo da importância dessas classes no modelo. O formato de arquivo especifica a codificação utilizada para o objeto lógico, além de permitir a identificação do software a ser utilizado para a apresentação do documento digital (objeto digital conceitual). O objeto digital lógico detêm as informações de codificação e formatação

do documento, o que permite a sequência de bits do objeto físico ser recuperado como um objeto conceitual inteligível por humanos.

A ontologia DeTec permite mapear e responder a questões como:

- de qual hardware e software um determinado documento digital é dependente;
- de qual (is) software(s) um determinado hardware é dependente;
- de quais hardwares e softwares uma mídia de armazenamento é dependente;
- numa descontinuidade de tecnologia, qual o impacto e quais mídias de armazenamento e formatos de arquivos requerem migração;
- de qual software outro software é dependente; etc.

A desativação de um componente tecnológico sem um correto diagnóstico é um grande risco e a ontologia DeTec possibilita mapear e responder o impacto da desativação de algum componente de tecnologia numa organização, reduzindo o risco de perdas de importantes conhecimentos armazenados em documentos digitais.

Quando um componente tecnológico necessita ser desativado, a organização pode identificar os objetos digitais que serão afetados, podendo planejar a migração desses objetos para formatos atualizados, suportados pelo ambiente tecnológico.

Ter mapeado quais hardwares e softwares devem ser mantidos e quais podem ser desativados é também um caminho para a racionalização do complexo ambiente de TIC que as organizações possuem, reduzindo custos e riscos.

8 CONSIDERAÇÕES FINAIS

O tema preservação digital tem despertado interesse de pesquisadores da comunidade científica e empresarial na busca de solução para evitar fenômeno já batizado de *digital dark age* por Kuny (1997) ou a amnésia coletiva (ROMHARDT, 1997).

A preservação da informação é um dos assuntos mais importantes para a história, cultura e economia da humanidade, assim como para o desenvolvimento da civilização (LEE et al., 2002).

O grande desafio do futuro não estará relacionado ao como armazenar a informação que deverá ser mantida, mas principalmente relacionado à redução do custo de armazenamento, à redução do risco e perda do conteúdo e à extração de todo valor do conteúdo salvo (GANTZ; REINSEL, 2010).

Segundo McGarry (1999) a 'memória da sociedade' foi tradicionalmente relacionada às bibliotecas e ultimamente o conceito tem sido transferido para as bases de dados eletrônicas. Na atualidade, a memória não está mais restrita ao conceito de 'registro de eventos passados', o desenvolvimento da Gestão do Conhecimento e a evolução do conceito do que é conhecimento e a necessidade de constante criação de novos conhecimentos permite ratificar a importância da preservação dos registros de informação e conhecimento.

Na ciência, o conhecimento explicitado e registrado, tem grande importância, sem o qual a comunicação com a comunidade fica praticamente nula.

Embora com frequência a literatura relacionada à preservação digital dê ênfase à preservação do documento digital, isto é, do objeto, ficou evidente que o foco não deve ser somente na preservação do objeto, de forma análoga a preservação de um exemplar de um livro impresso.

No tocante à dependência tecnológica, as frequentes afirmações que o documento digital que tem dependência de hardware e software, deve ser cuidadosamente interpretado. Na ontologia DeTec foi possível perceber que:

- é o formato que tem dependência do software; e
- é a mídia de armazenamento que tem dependência do hardware.

Essa diferenciação conceitual pode ter importante impacto na concepção e desenvolvimento de soluções de preservação digital.

A preservação digital implica no desenvolvimento de modelos e arquiteturas que permitam a preservação da forma e do conteúdo do acervo digital, sem perder a autenticidade.

A ontologia DeTec está limitada a um modelo que permite mapear a relação entre os objetos digitais e os componentes tecnológicos de uma organização, não constituindo o escopo do trabalho um modelo que contemple todos os processos necessários para garantir a preservação de documentos digitais.

A aplicação da DeTec pode permitir a antecipação, auxiliando no planejamento de desativações e migrações de objetos digitais para novos formatos, evitando assim destruição involuntária de conteúdos.

Num trabalho futuro, a Detec deverá evoluir, introduzindo conceitos ainda não contemplados no modelo e ser validado, num estudo de caso que permita a inclusão de indivíduos reais na ontologia, transformando-a numa base de conhecimento no conceito de Noy e McGuinness (2001).

REFERÊNCIAS

BARVE, S. File Formats in Digital Preservation. In: International Conference on Digital Libraries, 2007, Bangalore. **Anais...** Bangalore:2007. p. 239-248.

CENDI DIGITAL PRESERVATION TASK GROUP. **Formats for digital preservation: A review of alternatives and issues.** 2007.

CHEN, S. S. The paradox of digital preservation. **Computer**, v. 34, n. 3, p. 24-28, 2001. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=910890>>. Acesso em: 13 maio 2011.

CHRISTENSEN, S. S. **Archival Data Format Requirements.** Copenhagen: The Royal Library. 2004. Trabalho não publicado. Disponível em:< http://netarkivet.dk/wp-content/uploads/Archival_format_requirements-2004.pdf>. Acesso em: 30 maio 2012.

FERREIRA, M. **Introdução à Preservação Digital: Conceitos, estratégias e actuais consensos.** Guimarães: Escola de Engenharia da Universidade do Minho, 2006.

GRUBER, T. **Ontology.** Disponível em: < <http://tomgruber.org/writing/ontology-definition-2007.htm>>. Acesso em: 05 abr 2012.

HILBERT, M.; LÓPEZ, P. The world's technological capacity to store, communicate, and compute information. **Science**, v. 332, n. 6025, p. 60-65. New York. 2011. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21310967>>. Acesso em: 01 março 2012.

HORRIDGE, M. et al. **A Practical Guide To BUilding OWL Ontologies Using Protégé 4 and CO-ODE Tools. 1.1.** Manchester: The University Of Manchester, 2007.

KUNY, T. A Digital Dark Ages: Challenges in the Preservation of Electronic Information. In: Annual IFLA General Conference; 63rd, 1997, Copenhagen. **Conference Programme and Proceedings.** Copenhagen: IFLA. 1997. Disponível em: <<http://archive.ifla.org/IV/ifla63/63kuny1.pdf>>. Acesso em: 10 julho 2010.

LEE, K. et al. The State of the Art and Practice in Digital Preservation. **Journal Of Research Of The National Institute Of Standards And Technology**, v. 107, n. 1, p. 93-106, 2002. Disponível em: < <http://introking3.com/002ftp/33/ase2/j71lee.pdf>>. Acesso em: 02 abril 2012.

MEADOWS, A. J. Avaliando o desenvolvimento da comunicação eletrônica. In: **Comunicação Científica.** Brasília: UnB, 2000. p.23-34.

MOORE, G. E. Cramming more components onto integrated circuits. **Electronics**, v. 38, n. 8, 1965. Disponível em: <ftp://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf>. Acesso em: 01 maio 2012.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101 : A Guide to Creating Your First Ontology.**Stanford: Stanford University. 2001. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101.pdf>. Acesso em: 13 maio 2009.

ROMHARDT, K. **Processes of Knowledge Preservation**: Away from a Technology Dominated Approach. Zürich: University of Geneva, 1997. Disponível em: <<http://www.dfki.uni-kl.de/~aabecker/Final/Romhardt/romhardt.html>>. Acesso em: 12 dezembro 2009.

THIBODEAU, K. Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In: The State of Digital Preservation: An International Perspective, 2002, Washington. **Proceedings...** Washington: CLIR and Library of Congress. 2002. Disponível em :< <http://www.clir.org/pubs/reports/pub107>>. Acesso em: 02 dezembro 2009.