

XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 2– Organização e Representação do Conhecimento

USO E APLICAÇÃO DE ONTOLOGIAS EM REPOSITÓRIOS INSTITUCIONAIS ACADÊMICOS

USE AND APPLICATION OF ONTOLOGIES IN ACADEMIC INSTITUTIONAL REPOSITORIES

Adriana Carla Ribeiro dos Santos – Universidade Estadual Paulista “Júlio Mesquita Filho”(UNESP) Campus de Marília

Edberto Farneda – Universidade Estadual Paulista Paulista “Júlio Mesquita Filho”(UNESP) Campus de Marília

Maria Leandra Bizello – Universidade Estadual Paulista Paulista “Júlio Mesquita Filho”(UNESP) Campus de Marília

Modalidade: Trabalho Completo

Resumo: Os Repositórios Institucionais são ferramentas essenciais para a coleta, organização, preservação e divulgação da produção intelectual de universidades e comunidades acadêmicas, além de promover a preservação da informação. Porém, a eficácia dos repositórios institucionais depende de sistemas bem estruturados e semanticamente ricos, onde as ontologias se destacam como uma abordagem promissora. Este estudo visa criar um modelo de ontologia para o Repositório Institucional da Universidade Estadual Paulista, com o intuito de melhorar a organização, descrição e recuperação dos conteúdos informacionais. A pesquisa é exploratória e bibliográfica, seguida pela aplicação experimental dos dados coletados. A construção da ontologia foi baseada na metodologia de Noy e McGuinness (2001) e desenvolvida utilizando o *software Protégé*. Analisamos os tipos de materiais presentes no Repositório Institucional da Universidade Estadual Paulista e categorizamos esses materiais em seis grandes classes: Produções Científicas, Trabalhos Acadêmicos, Materiais de Eventos, Materiais Educacionais, Materiais Técnicos e Jurídicos, e Mídias e Comunicação. A ontologia criada inclui classes principais e suas subclasses, além de especificar propriedades de dados e objetos relevantes. A implementação de ontologias no Repositório Institucional da Universidade Estadual Paulista oferece uma estrutura flexível, suportando a integração de dados, desambiguação de termos e interoperabilidade entre diferentes sistemas.

Palavras-chave: Ontologias; Repositórios Institucionais; Repositório Institucional da Universidade Estadual Paulista.

Abstract: Institutional repositories are essential tools for collecting, organizing, preserving and disseminating the intellectual output of universities and academic communities, as well as promoting the preservation of information. However, the effectiveness of institutional repositories depends on well-structured and semantically rich systems, where ontologies stand out as a promising approach.

This study aims to create an ontology model for the São Paulo State University Institutional Repository in order to improve the organization, description and retrieval of information content. The research is exploratory and bibliographical, followed by experimental application of the data collected. The construction of the ontology was based on the methodology of Noy and McGuinness (2001) and developed using Protégé software. We analyzed the types of materials present in the São Paulo State University Institutional Repository and categorized them into six major classes: Scientific Productions, Academic Works, Event Materials, Educational Materials, Technical and Legal Materials, and Media and Communication. The ontology created includes main classes and their subclasses, as well as specifying relevant data and object properties. The implementation of ontologies in the São Paulo State University Institutional Repository offers a flexible structure, supporting data integration, term disambiguation and interoperability between different systems.

Keywords: Ontologies; Institutional Repositories; Universidade Estadual Paulista Institutional Repository.

1 INTRODUÇÃO

Os Repositórios Institucionais (RIs) são ferramentas que coletam e mantêm a produção intelectual de uma ou mais universidades ou comunidades (Crow, 2002), refere-se também aos conjuntos de serviços disponibilizados por uma universidade para os membros de sua comunidade, com o objetivo de gerenciar e compartilhar materiais digitais criados pela instituição e pelos membros da comunidade (Lynch, 2003).

Os RIs promovem o acesso aberto e facilitam a (re)utilização e preservação de informações. No entanto, a gestão eficiente e a recuperação informacional com precisão, dependem de sistemas bem estruturados e semanticamente ricos. Nesse contexto, o uso de ontologias se destaca como uma abordagem promissora para melhorar a organização, a descrição e a recuperação de conteúdos nos RIs.

Gruber (1993) descreve uma ontologia como uma especificação explícita e formal de uma conceituação compartilhada. Portanto, no campo da Ciência da Computação e da Ciência da Informação, ontologia é definida como uma representação formal de um conjunto de conceitos e seus relacionamentos dentro de um domínio.

As ontologias são utilizadas para modelar o conhecimento de forma estruturada e compreensível tanto por humanos quanto por máquinas. Elas fornecem uma base para a interoperabilidade semântica, permitindo que sistemas diferentes compreendam e utilizem a mesma informação de maneira consistente.

Uma das principais vantagens do uso de ontologias nos RIs é a capacidade de organizar e estruturar dados de forma lógica e coerente. As ontologias permitem a criação de uma

taxonomia hierárquica que facilita a classificação dos documentos, teses, artigos e outros materiais armazenados no repositório. Isto resulta em uma navegação mais intuitiva e uma melhor organização dos recursos, facilitando o acesso e a recuperação da informação.

Para tanto, elaborar uma ontologia para um repositório institucional de uma universidade envolve a definição de um conjunto estruturado de conceitos e relações que descrevem os diversos tipos de recursos e atividades acadêmicas. A ontologia precisa ser detalhada o suficiente para representar adequadamente os dados do repositório, mas também flexível para permitir a expansão conforme as necessidades evoluem.

O Repositório Institucional (RI) da Universidade Estadual Paulista (UNESP) é uma plataforma digital destinada a coletar, armazenar, preservar, organizar e disseminar a produção intelectual e científica da comunidade acadêmica da UNESP. Este repositório serve como um instrumento essencial para promover o acesso aberto ao conhecimento produzido na instituição, facilitando a visibilidade e o impacto das pesquisas desenvolvidas, conforme, descrito em sua página: <https://repositorio.unesp.br/home>.

O objetivo deste artigo é criar um modelo de ontologia para o Repositório Institucional da Universidade Estadual Paulista. Para atingir o objetivo proposto, realizamos uma pesquisa exploratória com levantamento bibliográfico sobre o tema em questão, seguido pela aplicação experimental do material coletado. A ontologia inclui classes, propriedades de dados e propriedades de objetos. O modelo de ontologia é definido com base em classes principais e suas subclasses, além de especificar propriedades relevantes.

2 O USO DE ONTOLOGIA NOS REPOSITÓRIOS INSTITUCIONAIS

O termo "ontologia" tem suas raízes na Metafísica, que é a "Filosofia Primeira" de Aristóteles, dedicada ao estudo essencial do ser. Embora Aristóteles nunca tenha empregado explicitamente a expressão "ontologia", a concepção de uma ciência do Ser está implicitamente presente em sua obra "Categorias", integrante do *Organon*, uma compilação de textos relacionados à lógica (Ferneda, 2013).

Gruber (1993) apresenta uma definição de ontologias:

Uma ontologia é uma especificação explícita de uma conceitualização. O termo é emprestado da filosofia, onde uma Ontologia é uma descrição sistemática da Existência. Para sistemas de Inteligência Artificial, o que "existe" é o que pode ser representado. Quando o conhecimento de um domínio é representado em um formalismo declarativo, esse conjunto de objetos que podem ser representados é chamado universo de discurso. Esse

conjunto de objetos, e as relações descritíveis entre eles, são refletidos no vocabulário representativo com o qual um programa baseado no conhecimento representa o conhecimento. Assim, no contexto da IA, podemos descrever a ontologia de um programa definindo um conjunto de termos representativos. Nessa ontologia, definições associam os nomes de entidades no universo do discurso (ex. classes, relações, funções ou outros objetos) com um texto legível por humanos que descreva o que esses nomes significam, e axiomas formais que forcem a interpretação e bom uso desses termos. Formalmente, uma ontologia é a declaração de uma teoria lógica (Gruber, 1993, p. 2-3).

Guarino (1998) acrescenta que uma ontologia é uma maneira explícita e formal de conceitualizar os termos e restrições dentro de um domínio de interesse. O quadro seguinte, apresenta definições de ontologias:

Quadro 01 – Definição de ontologia

Autor(es)	Definição de ontologia
Maedche e Staab (2001)	Servem como esquemas de metadados que fornecem vocabulários de conceitos com semânticas explicitamente definidas e processáveis por máquinas.
Hovy (2002)	É um conjunto de termos associados com definições em linguagem natural que utiliza relações formais e é relativo a algum domínio de interesse.
Daconta <i>et al.</i> (2003)	Ontologia define termos e conceitos comuns na representação e descrição de determinada área de conhecimento e, assim, padroniza o significado.
Almeida e Bax (2003)	Aperfeiçoam os mecanismos para recuperar a informação pela organização do conteúdo de fontes de dados em um domínio específico, contribuindo para a disposição do conhecimento.
Ramalho (2006)	É um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas por programas computacionais.
Castro (2008)	Ela é o resultado da junção de dois termos gregos <i>onta</i> (entes) e <i>logos</i> (teoria, discurso, palavra). Ontologia significa, portanto, teoria dos entes. “Ente” está aí representando todas as coisas sobre as quais se pode dizer que são – ou que a ontologia é a teoria do ser enquanto tal.
Almeida (2013)	Ferramentas que buscam modelar um domínio para a produção automática de inferências [...], facilitando, portanto, a instrução às máquinas sobre a esfera semântica dos conceitos por trás dos termos.
Lima e Maculan (2017)	É considerada uma linguagem mais semanticamente rica em relação aos outros SOC, pois, além de não haver limites para as variações das relações, todas as relações entre termos podem ser explicitadas numa ontologia, inclusive as relações de equivalência e associativas.

Fonte: Elaborado pelos autores (2024)

As definições de ontologia apresentadas por diferentes autores no quadro 1 apresentam alguns pontos em comum, mas também apresentam algumas diferenças em ênfase e abordagem:

Pontos em Comum:

1. Representação de Conceitos: Todas as definições concordam que ontologias são utilizadas para representar e descrever conceitos em um domínio específico.
2. Padronização Semântica: Há um consenso de que ontologias padronizam o significado dos termos dentro de um domínio, facilitando a interoperabilidade e compreensão comum.
3. Suporte a Processamento por Máquinas: Muitas definições mencionam que ontologias são esquemas de metadados ou artefatos tecnológicos que podem ser processados por programas computacionais, permitindo inferências automáticas.

Diferenças Notáveis:

1. Ênfase na Formalização: Autores como Ramalho (2006) destacam a importância da formalização em linguagem lógica e formal para possibilitar inferências automáticas, enquanto outros autores como Castro (2008) discutem mais sobre a teoria dos entes e a origem filosófica do termo.
2. Amplitude de Relações: Algumas definições, como a de Lima e Maculan (2017), enfatizam a capacidade das ontologias de explicitar uma ampla gama de relações entre os termos, incluindo equivalências e associações, diferenciando-se pela riqueza semântica oferecida.
3. Finalidade Prática: Autores como Almeida e Bax (2003) mencionam especificamente a melhoria na recuperação da informação e organização de dados como um dos principais objetivos das ontologias, enquanto outros autores podem enfatizar mais o aspecto teórico ou filosófico.

Essas diferenças refletem nas diferentes abordagens teóricas e práticas para a construção e utilização de ontologias, cada uma com suas ênfases e contextos específicos.

O quadro seguinte, trata da importância das ontologias nos Repositórios Institucionais, evidenciando seu papel estratégico em diversas áreas fundamentais para a organização e recuperação da informação:

**Quadro 2 - Importância do uso e aplicação de ontologias
nos Repositórios Institucionais**

Uso e Aplicação	Caracterização	Funcionalidade
-----------------	----------------	----------------

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Esquema de metadados	Os metadados consistente na padronização, melhoria, organização e recuperação dos dados.	criar esquemas de metadados, a fim de descrever os recursos do repositório institucional, como artigos, dissertações teses, etc.
Classificação e indexação	Uma busca mais precisa e relevante para facilitar a localização de informações pelos usuários.	classificar e indexar conteúdos, na navegação e a busca no repositório institucional.
Integração de dados	A interoperabilidade tem a capacidade de reunir informações dispersas.	integrar dados de diferentes fontes, cria um repositório institucional unificado, com informações de outras instituições ou departamentos.
Visualização e Exploração de Dados	A visualização clara das relações entre dados, facilita a exploração e descoberta de novos dados.	Oferecer interfaces intuitivas, como mapas de conhecimento e gráficos de relacionamento.
Suporte à Pesquisa	A pesquisa interdisciplinar é considerada como novas conexões entre áreas de conhecimento.	Oferecer correlações e relações entre diferentes áreas do conhecimento.
Desambiguação de Termos	A redução de ambiguidades e aumento da precisão na recuperação de informações.	Desambiguar termos, assegurar que o contexto correto seja aplicado.
Interoperabilidade	Dados mais acessíveis e reutilizáveis entre diversas plataformas e instituições.	Facilitar a interoperabilidade entre diferentes sistemas e repositórios institucionais.
Manutenção e Atualização	Simplificação da gestão do repositório, assegurando a continuidade e a relevância dos dados armazenados.	Facilitar a atualização e manutenção do repositório, permitindo a incorporação de novas informações de forma consistente.

Fonte: elaborado pelos autores (2024)

Esses pontos evidenciam como a aplicação de ontologias nos Repositórios Institucionais aprimora a organização, a recuperação e a exploração de dados, além de favorecer a integração de informações e o incentivo à pesquisa interdisciplinar. De modo geral, o uso e aplicação de ontologias podem ser usadas como ferramentas educacionais para estudantes e pesquisadores em um determinado campo de conhecimento.

2.2 Metodologia

Para a fundamentação teórica desta pesquisa, realizamos primeiramente, um levantamento bibliográfico nas seguintes bases de dados: a Base de Dados de Periódicos em

Ciência da Informação (BRAPCI), a Base de Dados de Teses e Dissertações (BDTD), a Base do Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (BENANCIB), a *Scientific Electronic Library Online* (SciELO) e a *ScienceDirect*.

Utilizamos como critérios de seleção, o material publicado no período entre 2001 e 2023, em língua portuguesa e inglesa. As palavras-chave utilizadas foram: ontologias e repositório institucional. Os textos foram utilizados para elaboração de quadros explicativos da definição e do uso e aplicação das ontologias nos repositórios institucionais. Buscamos também, na literatura da Organização do Conhecimento os fundamentos teóricos de Gruber (1993) e Guarino (1998) referentes à ontologia.

Para representar a construção de uma ontologia analisamos os dados do RI da UNESP, em seguida, utilizamos os princípios da metodologia *Ontology Development 101: a guide to creating your first ontology* (Noy; McGuinness, 2001) e o *software* de engenharia de ontologias Protégé, desenvolvido pelo Instituto Nacional de Ciências Médicas Gerais, da Universidade de Stanford.

2.3 Resultados e discussão

A metodologia *Ontology Development 101* é uma abordagem simples e prática, ideal para a criação de ontologias, especialmente em projetos menores ou em equipes com pouca experiência. Ela oferece um roteiro claro e detalhado, desde a definição do escopo até a criação de instâncias, sendo uma solução eficaz para quem busca um processo iterativo e acessível. Aplicá-la ao RI da UNESP pode resultar em uma organização mais estruturada dos metadados, aprimorar a recuperação de informações e facilitar a interoperabilidade com outros sistemas. Seguindo o passo a passo da metodologia, adaptamos sua aplicação ao RI da UNESP:

1. Definir o Domínio e o Escopo da Ontologia

- Domínio: O Repositório Institucional da UNESP. Isso inclui recursos como artigos científicos, teses, dissertações, relatórios técnicos e outros tipos de publicações acadêmicas.
- Escopo: Uma ontologia direcionada para a estruturação de informações bibliográficas, autores, áreas de conhecimento, instituições envolvidas, além de metadados relacionados, como tipos de documentos, datas de publicação, acesso etc.

- Objetivo: Organizar os documentos acadêmicos, permitir consultas mais precisas e garantir que os recursos estejam estruturados de acordo com padrões acadêmicos nacionais e internacionais.

O RI da UNESP visa armazenar, preservar, divulgar e disponibilizar de forma aberta documentos científicos, acadêmicos, artísticos e técnicos, bem como, dados e planos de gestão produzidos por pesquisadores e estudantes da UNESP.

Possui acesso aberto, segundo o *site* do RI da UNESP, consiste na disponibilização gratuita da produção intelectual na *internet*, mas ressalva que pode haver algumas restrições quanto ao uso desses materiais. Destaca o aumento da visibilidade do trabalho dos autores e a acessibilidade dos leitores, ampliando o impacto da universidade e de sua produção intelectual na sociedade. Apresenta tutoriais de auto arquivamento de Tese, Dissertações e Trabalho de Conclusão de Residência. A figura 1, apresenta os tipos de matérias que compõem o RI da UNESP.

Figura 1 – Matérias do Repositório Institucional da UNESP

Tipos de Materiais

Artigo: 115219	Dissertação de mestrado: 34522	Tese de doutorado: 19017
Trabalho de conclusão de curso: 15986	Trabalho apresentado em evento: 10597	Resumo: 6776
Resenha: 3092	Capítulo de livro: 1457	Editorial: 1045
Carta: 854	Livro: 581	Patente: 514
Errata: 355	Nota: 314	Podcast: 277
Tese de livre-docência: 213	Dado de pesquisa: 131	Trabalho de conclusão de residência: 85
Boletim: 78	Revista: 74	Plano de gestão de dados: 53
Jornal: 34	Data paper: 31	Objeto educacional: 28
Relatório: 28	Anais: 16	Biografia: 8
Relatório de pesquisa: 7	Tutorial: 3	Partitura: 2
Regulamento: 1	Relatório de pós-doc: 1	Vídeo: 1

Fonte: Repositório Institucional da UNESP (2016)

Para navegar no Repositório Institucional da UNESP, o usuário seleciona uma comunidade de suas coleções (Produção Acadêmica e Científica; Produção cultural e Produção técnica).

2. Considerar o Reuso de Ontologias Existentes

A metodologia *Ontology Development* 101 recomenda que antes de construir uma ontologia do zero, devemos procurar por ontologias existentes que possam ser reutilizadas ou adaptadas. No entanto, optamos por uma abordagem híbrida, combinando ontologias

padronizadas, como *Dublin Core* ou BIBO para a descrição bibliográfica de publicações, com uma ontologia específica da UNESP para refletir estruturas internas, políticas institucionais e linhas de pesquisa. A UNESP é uma universidade com múltiplos campi, cursos e linhas de pesquisa. Criar uma ontologia padronizada pode permitir que todas essas complexidades sejam bem representadas, organizando de maneira única a produção acadêmica e a relação entre departamentos, unidades de ensino, pesquisadores e projetos.

3. Enumerar os Termos Importantes

Os principais conceitos e termos usados no RI da UNESP. Conceitos: Materiais (Artigos científicos, teses, dissertações, livros, capítulos de livros etc.); Autor (Pesquisadores, estudantes, professores vinculados à UNESP ou outras instituições); Instituição (Departamentos, faculdades, centros de pesquisa); Área de Conhecimento (Ciências Exatas, Humanas, Biológicas, etc.); Data de Publicação (Ano e data de publicação dos documento); Tipo de Acesso (Aberto, restrito, embargado); Revisão por Pares (Informação sobre se o documento foi revisado ou não por pares). Lista de Termos (Artigo Científico, Autor, Área de Conhecimento, Data de Publicação, Palavras-chave, Idioma, Revisão por Pares).

4. Definir as Classes e Organizar em Hierarquias

A ontologia tem como função organizar os materiais em grandes categorias ou classes, facilitando a estruturação e a recuperação de informações. As classes são o elemento central e descrevem os conceitos dentro de um domínio específico. Por exemplo, em uma ontologia para RIs, uma classe representaria todos os materiais, enquanto, um material específico seria uma instância dessa classe.

Além disso, uma classe pode possuir subclasses que representam conceitos mais específicos relacionados à sua superclasse. Portanto, dividimos os materiais (figura 1) em seis categorias, conforme, apresentado no quadro seguinte:

Quadro 3 – Categorização da ontologia dos tipos de materiais acadêmicos da UNESP

Categorias	Tipos de materiais	Caracterização
1. Produções Científicas	Artigos, resenhas, resumos, capítulos de livro, editoriais, cartas, livros, notas, revistas, data <i>papers</i> , anais, boletins, jornais, biografias, relatórios de pesquisa, tutoriais, regulamentos e vídeos	Estes tipos de materiais são frequentemente utilizados para divulgar resultados de pesquisa e contribuições acadêmicas.
2. Trabalhos Acadêmicos	Dissertações de mestrado, teses de doutorado, trabalhos de conclusão de curso (TCC), teses de livre-docência e relatórios de pós-doutorado.	Estes documentos são produzidos como parte dos programas acadêmicos e de pesquisa.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

3.Materiais de Eventos	Trabalhos apresentados em eventos como conferências, seminários, <i>workshops</i> , entre outros.	Estes materiais documentam apresentações e discussões que ocorrem em encontros acadêmicos e profissionais.
4.Materiais Educacionais	Compreende objetos educacionais que são utilizados para fins de ensino e aprendizagem.	Este tipo de material pode incluir recursos como módulos de ensino, guias, manuais e outros recursos didáticos.
5.Materiais Técnicos e Jurídicos	Abrange patentes e planos de gestão de dados.	Estes materiais são essenciais para proteger invenções e organizar a gestão de dados de pesquisa.
6.Mídias e Comunicação:	Inclui <i>podcasts</i> e outros formatos de mídia	Materiais utilizados para a comunicação de informações acadêmicas e científicas de forma acessível e ampla.

Fonte: elaborado pelos autores (2024)

Os principais critérios utilizados para definir as categorias dos tipos de materiais acadêmicos da UNESP apresentados no quadro anterior são:

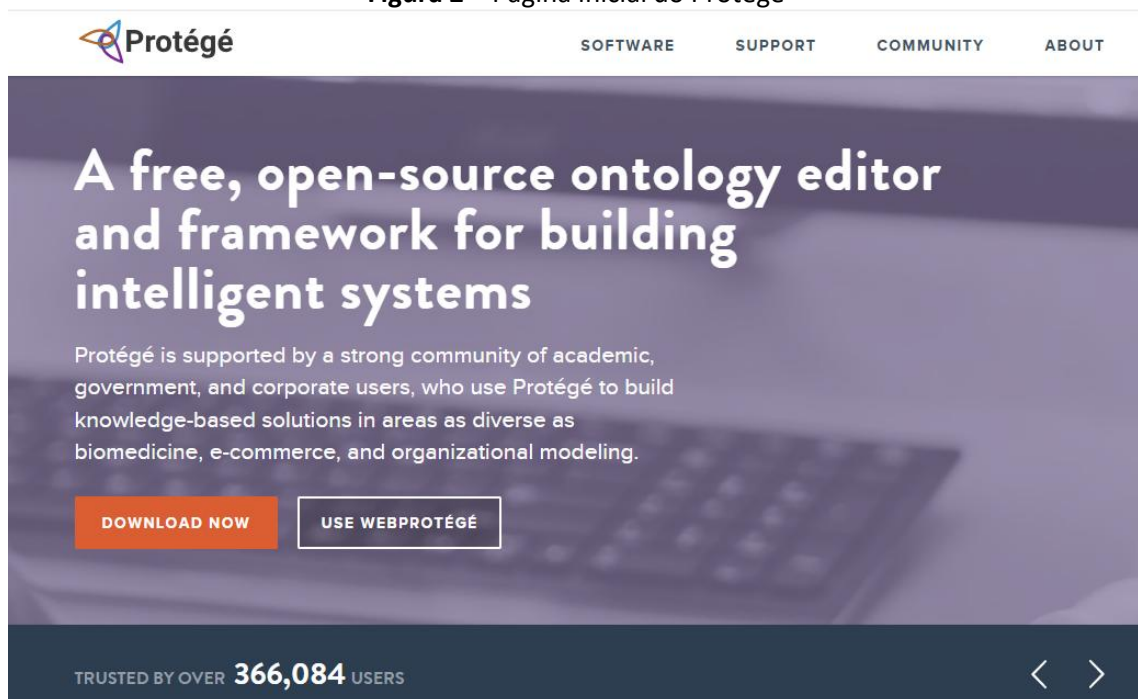
1. **Produções Científicas:** Inclui os materiais publicados e revisados por pares, que são fundamentais para a disseminação de conhecimento acadêmico e científico. Categorizar esses materiais permite uma organização detalhada e uma busca eficiente dentro do repositório da UNESP.
2. **Trabalhos Acadêmicos:** Representam a produção acadêmica formal dos alunos e pesquisadores, essencial para a documentação de progressos acadêmicos e obtenção de graus acadêmicos.
3. **Materiais de Eventos:** Captura a produção gerada a partir de eventos acadêmicos e profissionais, importante para registrar e disponibilizar as contribuições apresentadas em conferências, eventos e *workshops*.
4. **Materiais Educacionais:** Inclui recursos utilizados diretamente no processo de ensino e aprendizagem, assegurando que materiais pedagógicos estejam organizados e acessíveis para fins educacionais.
5. **Materiais Técnicos e Jurídicos:** Abrange documentos especializados que são fundamentais para a proteção da propriedade intelectual e gestão de dados, refletindo a importância da documentação técnica e regulamentar da UNESP.

6. Mídias e Comunicação: Adiciona formatos modernos de mídia que são cada vez mais usados para disseminação de conhecimento e comunicação científica, como *podcasts* e vídeos.

Os critérios garantem que as categorias escolhidas sejam abrangentes, funcionais e adaptáveis, além de favorecerem a organização semântica dos dados no contexto do repositório institucional da UNESP.

Utilizamos o *software* Protégé para ilustrar a categorização da ontologia dos tipos de materiais acadêmicos. O Protégé é uma plataforma gratuita e de código aberto amplamente empregada para modelar ontologias e desenvolver sistemas baseados em conhecimento. Criado pela Universidade de Stanford, seu foco principal é simplificar o processo de criação, visualização e gerenciamento de ontologias.

Figura 2 – Página inicial do Protégé



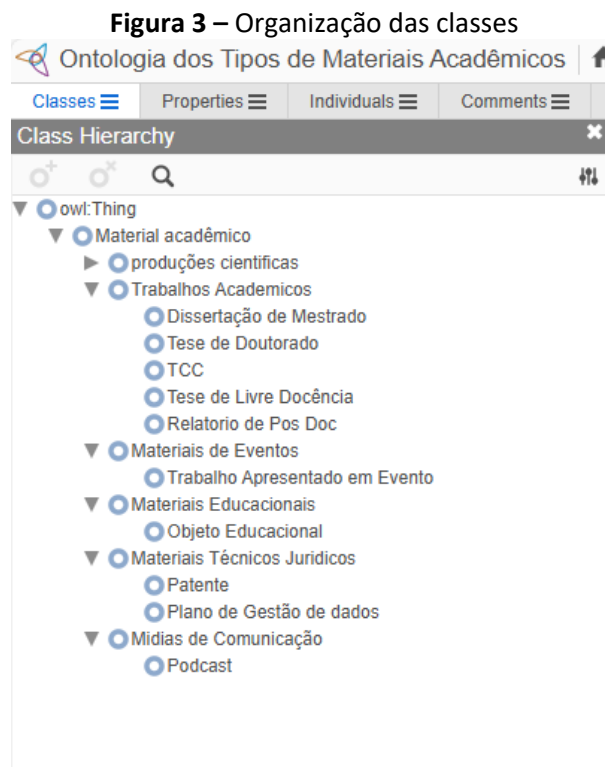
Fonte: Protégé (2024)

O funcionamento do Protégé envolve um conjunto de ferramentas e funcionalidades voltadas para a criação, edição e gerenciamento de ontologias, permitindo modelar o conhecimento de um determinado domínio de forma estruturada. As principais características do Protégé:

1. Modelagem de Ontologias: O Protégé permite aos usuários criar classes, propriedades, instâncias e axiomas para descrever formalmente um domínio de conhecimento. Através da criação de uma ontologia, é possível organizar e estruturar informações complexas.

2. Editor OWL: Ele oferece suporte completo à *Web Ontology Language* (OWL), um dos principais padrões para a definição de ontologias na *web* semântica. Isso facilita a integração de ontologias com outras ferramentas e plataformas que utilizam OWL.
3. Extensibilidade: O Protégé pode ser ampliado através de *plugins* que adicionam funcionalidades específicas, como integração com motores de inferência ou outras ferramentas para validação e visualização de dados.
4. Interface Gráfica e Usabilidade: A interface do Protégé é projetada para ser intuitiva e oferece diversos modos de visualização da ontologia, como visualizações hierárquicas, de rede, ou axiomáticas, permitindo uma navegação mais fácil pelo modelo.
5. Suporte à *Web Semântica*: Protégé é usado extensivamente em projetos de *web* semântica e é uma ferramenta central para desenvolver ontologias que suportam a interoperabilidade de dados na *web*.

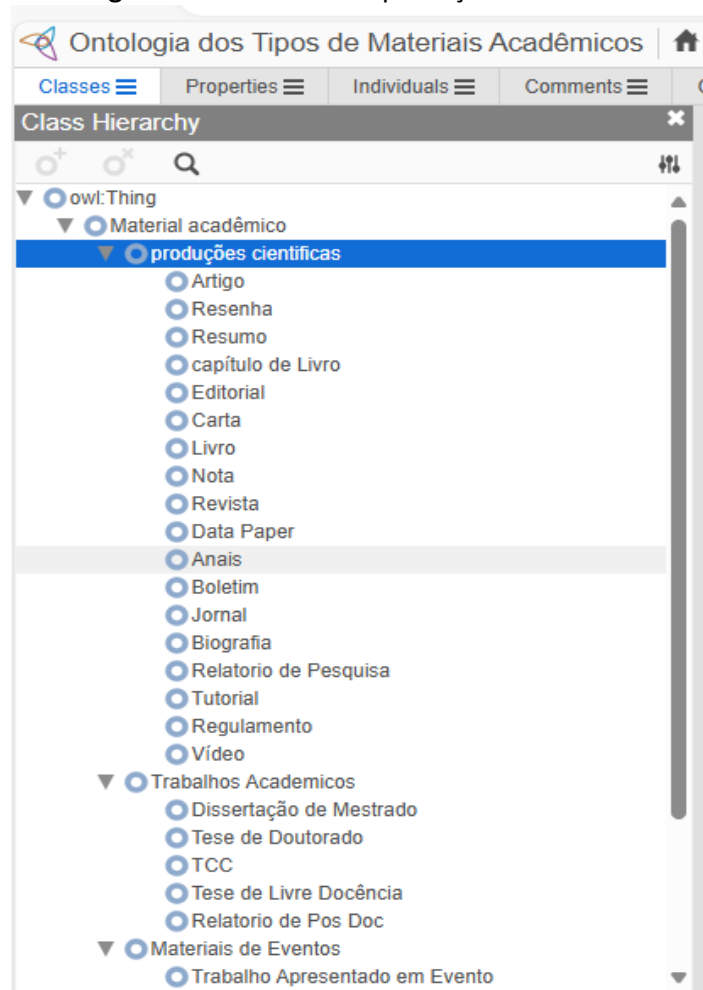
No que se refere as aplicações, o Protégé é uma ferramenta essencial para quem trabalha com organização de conhecimento e *web* semântica. O *software* é utilizado no campo das Ciências da Vida e Medicina (para desenvolver ontologias que estruturam informações nas áreas biomédica e genômica); Engenharia de *Software* (para a formalização de domínios complexos, facilitando a automação de processos de raciocínio); Gestão do Conhecimento (auxilia na organização de vastos conjuntos de dados em setores como educação, pesquisa e indústrias tecnológicas).



Fonte: elaborado pelos autores (2024)

Em seguida, definimos as relações hierárquicas entre os tipos de materiais para permitir uma categorização que facilite na busca e recuperação da informação. Estas relações hierárquicas estruturam a ontologia de maneira lógica, onde cada categoria principal contém subclasses que representam tipos específicos de materiais. Por exemplo, a categoria "Produções Científicas" contém subclasses como "Artigo", "Resenha", "Resumo", entre outras. Esta estrutura hierárquica não apenas organiza os materiais de forma intuitiva, mas também melhora a navegação e a acessibilidade dentro do RI da UNESP.

Figura 4 – Subclasse das produções científicas



Fonte: elaborado pelos autores (2024)

Para cada tipo de material terá um ou mais atributos adicionais, para enriquecer a descrição dos materiais e fornecer informações detalhadas para os usuários. Alguns dos atributos comuns incluem:

- Autor(es): Nome(s) do(s) autor(es) do material.
- Data de Publicação: Data em que o material foi publicado.

- Instituição: Instituição associada ao material, como universidades ou centros de pesquisa.
- Departamento: Departamento ou unidade acadêmica responsável pelo material.
- Resumo: Resumo ou descrição breve do conteúdo do material.
- Palavras-chave: Termos chave associados ao conteúdo do material.
- Idioma: Idioma em que o material está escrito.
- Identificador: Identificadores únicos como DOI, ISSN, ISBN, etc.
- Licença: Informação sobre os direitos de uso e licença do material.

Além desses atributos comuns, existem atributos específicos que se aplicam a determinados tipos de materiais. Por exemplo, um "Artigo" pode ter atributos como "Periódico", "Volume", "Número", e "Páginas", enquanto uma "Dissertação de Mestrado" pode incluir "Orientador(es)", "Programa de Pós-graduação", e "Área de Concentração". Estes atributos específicos fornecem um nível adicional de detalhe, permitindo uma descrição mais completa e precisa de cada material.

Ao incorporar estas classes, relações e atributos, a ontologia oferece uma estrutura robusta e flexível para organizar e gerenciar diferentes tipos de materiais acadêmicos e científicos no RI da UNESP, melhorando a eficiência na gestão de informações e facilitando o acesso e a recuperação dos dados pelos usuários.

5. Definir as Propriedades das Classes

Depois, dividimos as propriedades da ontologia para os materiais acadêmicos do RI da UNESP em dois tipos: em propriedades comuns, que se aplicam a todos os tipos de materiais, e propriedades específicas, que são exclusivas para certos tipos de materiais.

Quadro 4 – Propriedades da ontologia do RI da UNESP

Propriedades Comuns (para todos os materiais)	Propriedades Específicas (para tipos específicos de materiais)
<i>Título</i> <i>Escrito por</i> <i>Data de publicação</i> <i>Número de páginas</i> <i>publicado em</i> <i>resumo</i> <i>palavras-chave</i> <i>idioma</i> <i>instituição</i> <i>departamento</i> <i>tipo de material</i> <i>identificador</i> <i>licença</i>	Artigo: <i>periodico, volume, numero, paginas</i> Dissertação de Mestrado/Tese de Doutorado: <i>orientador, programade posgraduacao, areade concentracao</i> TCC: <i>curso, orientador</i> Trabalho Apresentado em Evento: <i>nomedoevento, localdoevento, datadoevento</i> Livro/Capítulo de Livro: <i>editora, isbn, edicao</i> Patente: <i>numerodepatente, datadeconcessao, titular</i> Dados de Pesquisa: <i>descricao do dataset, formato do dataset, tamanho do dataset</i> Podcast: <i>duracao, serie</i>

	Plano de Gestão de Dados: <i>responsavel pelo plano, versao</i>
--	--

Fonte: elaborado pelos autores (2024)

A finalidade dessas propriedades é enriquecer a descrição dos materiais, permitir uma organização mais eficiente e uma recuperação de informações mais precisa e detalhada dentro do RI da UNESP.

6. Definir Restrições das Propriedades

Para garantir consistência nos dados estabelecemos restrições nas propriedades. Por exemplo: Cardinalidade (Um documento pode ter múltiplos autores, mas apenas um título); Domínio e Alcance (A propriedade "escrito por" deve ter como domínio a classe "Produções científicas" e como alcance a classe "Trabalhos acadêmicos"); A propriedade "publicado em" deve ter como domínio a classe "Mídias e comunicação" e como valor uma data. Tipo de Dado: A propriedade "Data de Publicação" deve aceitar apenas valores no formato de data. A propriedade "Número de Páginas" deve aceitar apenas valores numéricos inteiros.

7. Criar Instâncias

Para popularizar a ontologia buscaremos criar instâncias de documentos e autores. Listando os seguintes elementos: Instância de Documento (Título; Data de Publicação; Área de Conhecimento e Autor). Instância de Autor (Nome; Afiliação e ORCID).

8. Revisão e Refinamento

Após a finalização da etapa de construção da ontologia, seguiremos com as etapas de revisão e refinamento com base no *feedback* dos usuários e nos testes se a ontologia pode responder a perguntas como: “Quais artigos sobre ontologias foram publicados pela UNESP em 2024?” ou “Quais autores da UNESP publicaram artigos na área de Ciência da Informação entre 2014 e 2024?”

3 CONSIDERAÇÕES FINAIS

Os Repositórios Institucionais (RIs) favorecem na coleta, organização, preservação e disseminação da produção intelectual de universidades e comunidades acadêmicas. Além de promover o acesso aberto, facilitar na preservação da informação. No entanto, a eficácia dessas plataformas depende da implementação de sistemas bem estruturados e semanticamente ricos. Neste contexto, as ontologias surgem como uma solução promissora, oferecendo uma abordagem sistemática para a organização, descrição e recuperação de conteúdos nos RIs.

Descrevemos as ontologias, como especificações explícitas e formais de uma conceitualização compartilhada, permitem a modelagem do conhecimento de forma compreensível tanto para humanos quanto para máquinas. Além do mais, estabelecem uma base para a interoperabilidade semântica, garantindo que diferentes sistemas possam compreender e utilizar a mesma informação de maneira consistente. Isso é particularmente vantajoso nos RIs, onde a organização lógica e coerente dos dados é essencial para a navegação intuitiva e a recuperação eficiente de informações.

A aplicação de ontologias para o Repositório Institucional da Universidade Estadual Paulista, demonstra como essas ferramentas podem estruturar e classificar os materiais acadêmicos de maneira eficaz. Utilizando o *software Protégé* e a metodologia de Noy e McGuinness, foi possível definir classes, propriedades de dados e objetos, criando uma taxonomia hierárquica que facilita a categorização e a busca de informações.

Em suma, as ontologias oferecem uma estrutura flexível para a gestão de informações nos RIs, melhora a visibilidade, acessibilidade e impacto da produção intelectual, portanto, representa um avanço significativo na organização e recuperação de informações acadêmicas, contribuindo para o desenvolvimento e a disseminação do conhecimento.

REFERÊNCIAS

ALMEIDA, M. B. Revisiting Ontologies: A Necessary Clarification. **Journal of the American Society for Information Science and Technology**, New Jersey, v. 64, n. 8, p. 1682-1693, 2013.

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, 2003. Disponível em: http://www.scielo.br/scielo.php?script=sci_a196520030003000002&lng=en&nrm=iso. Acesso em: 11 junho 2024.

CASTRO, S. **Ontologia**. Rio de Janeiro: Jorge Zahar, 2008.

CROW, R. **The case for institutional repositories**: A SPARC Position Paper. Washington, DC: Scholarly Publishing & Academic Resources Coalition, 2002.

DACONTA, M. C.; SMITH, K. T.; OBRST, L. J. **The semantic web**: a guide to the future of XML, Web services, and knowledge management. Indianapolis, Indiana: Wiley Pub, 2003.

FERNEDA, E. **Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação**. 2013. Relatório de Pesquisa. (Pós-Doutorado em Ciência da Informação) - Universidade Federal da Paraíba, João Pessoa, 2013. Disponível em:

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

<https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/pos-doutorado.pdf> . Acesso em: 11 jun. 2024.

GUARINO, N. Semantic matching: formal ontological distinction for information, extraction, and integration. *In*: PAZIENZA, M. T. (Ed.). **Information Extraction: a multidisciplinary approach to the emerging Information Technology**. Berlin: Springer Verlag, 1998.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, [S.l.], v. 5, n. 2, p. 199-220, 1993.

HOVY, E. Comparing Sets of Semantic Relations in Ontologies. *In*: GREEN, R.; BEAN, C. A.; MYAENG, S. H. (Ed.). **The semantics of relationships: an interdisciplinary perspective**. Norwell, USA: Springer, 2002. v. 3, p. 91–110.

LIMA, G. Â.; MACULAN, B. C. M. S. Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. **Ciência da Informação**, Brasília, v. 46, n. 1, p. 1–13, dez. 2017. Disponível em: <https://revista.ibict.br/ciinf/article/view/4014> . Acesso em: 11 jun. 2024.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the Digital Age. **Libraries and the Academy**, [S.l.], v. 3, n. 2, 2003.

MAEDCHE, A.; STAAB, S. Ontology learning for the Semantic Web. **IEEE Intelligent Systems**, IEEE Educational Activities Department, [S.l.], v. 16, n. 2, p. 72–79, 2001.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford Knowledge Systems Laboratory, [S.l.], v. 32, 2001.

PROTÉGÉ. **Protégé**. 2024. Disponível em: <http://protege.stanford.edu>. Acesso em 28: 11 jun. 2024.

RAMALHO, R. A. S. **Web semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação**. 2006. 120 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Marília, 2006. Disponível em: <http://hdl.handle.net/11449/93709>. Acesso em: 11 jun. 2024.

UNIVERSIDADE ESTADUAL PAULISTA. **Repositório Institucional UNESP**. São Paulo: UNESP, 2016. Disponível em: <http://repositorio.unesp.br>. Acesso em: 11 jun. 2024.