15º Congresso de Inovação, Ciência e Tecnologia do IFSP - 2024

Aplicação Web para Importação de Glossários de Cursos Técnicos Profissionalizantes Representados como Ontologias em OWL

DANIEL F. D. PIAGGE e CRISTIANE A. YAGUINUMA

1 Discente do Curso Técnico em Informática Integrado ao Ensino Médio e Bolsista PIBIFSP, IFSP – Campus Araraquara, daniel.piagge@aluno.ifsp.edu.br
2 Professora Doutora e orientadora de projeto, IFSP – Campus Araraquara, cristiane.yaguinuma@ifsp.edu.br

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação

RESUMO: O projeto em questão visa aprimorar o acesso a glossários técnicos desenvolvidos para o Curso Técnico em Mecânica Integrado ao Ensino Médio. Trabalhos anteriores criaram glossários tríligues (português, inglês e espanhol) que, apesar de funcionais, apresentavam limitações para integração com sistemas computacionais devido ao formato textual utilizado. Para superar essa barreira, foi desenvolvido um modelo computacional usando ontologias, especificamente com a linguagem OWL, que formaliza o domínio dos termos técnicos e suas relações. No entanto, a necessidade de gerenciar e expandir esses glossários motivou o desenvolvimento de uma aplicação web capaz de importar essas ontologias OWL e armazená-las em um banco de dados relacional, pois o banco de dados relacional facilita os meios de consulta e manutenção dos termos existentes. O desenvolvimento da aplicação envolveu o uso da biblioteca SimpleXML, que facilitou a leitura dos arquivos OWL e a extração dos dados necessários. Essa abordagem permitiu a correta vinculação entre termos técnicos e seus respectivos atributos, como disciplinas, traduções e definições. Em suma, o projeto estabelece uma base para futuras expansões, onde novos glossários poderão ser facilmente integrados ao sistema sem a necessidade de processos manuais extensivos para inserção e manutenção dos termos já existentes.

PALAVRAS-CHAVE: OWL, SimpleXML, aplicação web, ontologias.

Web Application for Importing Glossaries of Professional Technical Courses Represented as Ontologies in OWL

ABSTRACT: The project in question aims to improve access to technical glossaries developed for the Technical Course in Mechanics Integrated into High School. Previous work created trilingual glossaries (Portuguese, English and Spanish) which, despite being functional, presented limitations for integration with computer systems due to the textual format used. To overcome this barrier, a computational model was developed using ontologies, specifically with the OWL language, which formalizes the domain of technical terms and their relationships. However, the need to manage and expand these glossaries motivated the development of a web application capable of importing these OWL ontologies and storing them in a relational database, as the relational database facilitates the means of querying and maintaining terms. existing. The development of the application involved the use of the SimpleXML library, which made it easier to read OWL files and extract the necessary data. This approach allowed the correct link between technical terms and their respective attributes, such as disciplines, translations and definitions. In short, the project establishes a basis for future expansions, where new glossaries can be easily integrated into the system without the need for extensive manual processes to insert and maintain existing terms.

KEYWORDS: OWL, SimpleXML, web application, ontologies.

INTRODUÇÃO

Projetos de iniciação científica anteriores na área da linguagens (Reis; Dantas, 2018) abordaram e representaram em formatos textuais termos presentes em disciplinas do Curso Técnico em Mecânica Integrado ao Ensino Médio, através de dicionários e glossários já elaborados. Os resultados apresentaram um glossário trilingue com tradução em espanhol e inglês. Porém, consultar a terminologia através de documentos textuais apresenta dificuldades por não conter muita estruturação para processamento em aplicações computacionais.

A fim de resolver essa limitação, um projeto de iniciação científica anterior (Piza; Yaguinuma, 2020) estruturou o glossário como uma ontologia computacional, um modelo de domínio de dados e conhecimento. Ademais, outro projeto (Silva; Yaguinuma, 2022) desenvolveu uma aplicação direcionada aos estudantes e docentes da área da mecânica para consulta ao glossário.

Com o sucesso proveniente do feedback de colaboradores, houve a necessidade de ampliar o projeto com um gerenciamento de glossários (Silva; Yaguinuma, 2023) que possibilita criação de novos glossários, inserção de termos, cursos, discplinas, traduções e fontes relacionadas. Entretanto, não fora desenvolvida a funcionalidade de importação de glossários existentes, e o cadastro manual do glossário organizado no projeto anterior (Piza; Yaguinuma, 2020) seria inviável por envolver mais de 600 termos, se tornando um processo exaustivo.

Buscando atender essa demanda, o objetivo do atual projeto atribui o desenvolvimento de uma aplicação web para importação de glossários representados como uma ontologia em OWL, que permita criar e armazenar glossários por intermédio de um banco de dados relacional, para que estejam disponíveis para outras aplicações voltadas para o gerenciamento e consulta dos termos.

MATERIAL E MÉTODOS

Para a compreensão do que foi inicialmente proposto para o projeto, contamos com materiais teóricos relacionados à ontologia no seu contexto filosófico e computacional (Machado, 2021), trabalhando com revisão bibliográfica, atribuindo seus conceitos com situações reais e relacionando suas ideias em aplicações para testes para obter uma dimensão da complexidade por trás do que seria manipulado durante a execução.

De acordo com (Machado, 2021), uma ontologia representa um conjunto de primitivas representacionais com um preceito lógico de afirmações de anotação, relacionamento e propriedade, atribuindo agrupamentos e classificações. Em seu sentido filosófico, é o ramo da metafísica que estuda os tipos de coisas que existem no mundo. Aristóteles apresenta categorias que servem de base para classificar entidades.

Já no sentido computacional a abordagem deve ter presente diferentes variantes interpretativas associadas, evitando entendimento paralelo. O entendimento do processo é desviante do contexto filosófico, focando no campo dos sistemas de informação visando estabelecer uma estrutura de dados complexa e sem ambiguidade (Horridge, 2009), não se enquadrando no que se denomina modelos conceituais, que é um conjunto de suposições baseadas no mundo real que indicarão as regras de negócio de um sistema.

Para aplicação de uma ontologia computacional, é possível contar com a OWL (*Ontology Web Language*) (Horridge, 2009), uma linguagem para definir e instanciar ontologias na World Wide Web. OWL é uma linguagem para definição e instanciação de ontologias Web. Uma ontologia OWL pode formalizar um domínio, definindo classes e propriedades destas classes, definir indivíduos e afirmações sobre eles, podendo atribuir consequências lógicas.

Para replicar o conteúdo presente na ontologia, foi utilizado um banco de dados relacional criado por um projeto anterior (Silva; Yaguinuma, 2023). O banco foi gerado por meio do Mapeador Objeto-Relacional (ORM) Prisma, com uma estrutura voltada para a plataforma SQLite, com domínios de tipo texto para ambas colunas das 11 tabelas estruturadas, sendo possível integrar com outros SGBDs como MySQL e PostgreSQL, alterando-se configurações do ORM Prisma.

Buscando compreensão da estrutura e relacionamentos entre os elementos, foi utilizada a aplicação MYSQL Workbench, que fornece a engenharia reversa de transformar um banco de dados SQL em um modelo lógico de fácil interpretação. Entretanto, os domínios demarcados no banco de dados não eram compatíveis com a aplicação MYSQL Workbench, considerando restrições específicas, como exigência de que valores inseridos em uma chave primária devessem ser curtos, sendo assim domínios de tipo texto não seriam uma alternativa viável.

Sendo assim, a estrutura do banco de dados foi alterada em seus domínios, a fim de resolver o conflito de compatibilidade e assegurar que toda a ontologia seria armazenada após a implementação da ferramenta de importação de ontologias.

Como definido, sendo necessário transcrever as informações da ontologia para um banco de dados relacional com intermédio da linguagem de programação PHP, foram pesquisadas bibliotecas comprovadamente eficazes e que pudessem ser o meio de apoio para chegar ao objetivo de obter o banco de dados relacional completo como previsto anteriormente.

Em primeira análise, foi revisada a biblioteca EasyRDF (De Jesus, 2021), se trata de uma biblioteca mais atual, com resultados promissores e que possui um meio de navegação por estruturas ontológicas de fácil manipulação. Por outro lado, essa manipulação se limita à linguagem de marcação RDF, então não seria possível trabalhar com toda a expressividade das construções da ontologia do glossário de termos da mecânica, tornando o processo árduo e trabalhoso.

Ademais, foi encontrada e utilizada a biblioteca simpleXML (Calegari, 2016), que permite a leitura em PHP de arquivos na linguagem de marcação XML, como era o caso da ontologia que precisávamos utilizar. A simpleXML garante manipulação dos dados e de seu conteúdo, mas se trata necessariamente de percorrer estruturas e seu interior de forma bem incisiva e específica sendo possível navegar não no produto final de uma estrutura XML, mas sim na estrutura propriamente dita, sendo necessário conhecer a base do que se foi feito e todos os seus elementos que derivam de outros para receber seus valores. Porém, mesmo que o trabalho parecesse exaustivo, foram utilizados laços para percorrer o arquivo por se tratar de estruturações repetidas, assim então, intermediando a passagem de dados. Com isso, o código de processamento estruturado através da plataforma Visual Studio Code foi mais específico para o domínio do arquivo de ontologia criada, buscando extrair todas as informações presentes e se obter o preenchimento do banco de dados relacional, presente na plataforma phpMyAdmin via Xampp.

A execução do processamento do arquivo envolveu um processo muito detalhado considerando que houve a necessidade de armazenamento primário através de tipos de memória em matriz utilizando sessões. Esse método mais simples aproxima-se ao que seria uma estrutura de tabelas, podendo ter uma dimensão de como informações de termos, disciplinas e definições estariam dispostos no produto final. Com isso também foi possível estabelecer um meio de comparação de identificadores IRI (*Internationalized Resource Identifier*), que é um identificador único atribuído a cada elemento na linguagem OWL, esse meio de identificação foi utilizado para definição de relacionamentos entre as informações no banco, onde para cada elemento que existia em função de outro no glossário, como uma tradução que depende da existência do termo na sua língua padrão, seria necessário afirmar a relação utilizando o IRI do elemento que é referenciado na afirmação lógica.

A prioridade de inserção ao banco de dados inicialmente foram as tabelas que não possuiam chaves estrangeiras de outras tabelas dentre as colunas, isso porque não seria possível ter como referenciá-las sem que estivessem cadastradas no banco. O intermédio de transcrição foi um esquema de funções PHP que conectou o banco de dados ao arquivo de processamento via mysqli. Sendo assim, tabelas sem referências das demais foram armazenadas primeiro, e para cada vez que uma linha era adicionada ao banco, o identificador gerado automaticamente pelo SQL era armazenado na sessão referente aquela tabela.

Ademais, tabelas que possuiam relacionamentos puderam ser adicionadas ao banco, já que os identificadores SQL necessários estavam presentes nas sessões assim como os identificadores IRI, podendo assim então atribuir comparações de IRI na sua estrutura escrita dependendo do seu formato de valor. Com isso, o mesmo IRI em diferentes sessões representava que existia uma relação definida e afirmada entre eles no sistema ontológico. Sendo assim, caso os identificadores estivessem sendo referenciados, seria necessário apenas utilizar os identificadores obtidos anteriormente na primeira fase de transcrição. As funções PHP atribuem as relações de chave estrangeira na segunda fase de inserção, sendo possível armazenar a linha completa apenas se o item na tentativa de inserção à coluna correspondesse com algum item que existisse na tabela referenciada, o que estabelece uma base que impossibilita a entrada de informações divergentes quando se trata de relacionamentos SQL.

RESULTADOS E DISCUSSÃO

Para a interface gráfica de importação de glossários no formato OWL, foi alocada uma área de retorno referente ao sucesso de importação de análise quantitativa de elementos que foram inseridos

com o processamento, essas mensagens estão dispostas logo abaixo do campo de upload de arquivo e o botão de submissão (figura 1). Além dela, outro avanço tão importante quanto a interface gráfica foi o banco de dados completo após o processo de transcrição dos termos, disciplinas, definições, fontes e traduções referentes aos termos do Curso Técnico em Mecânica.

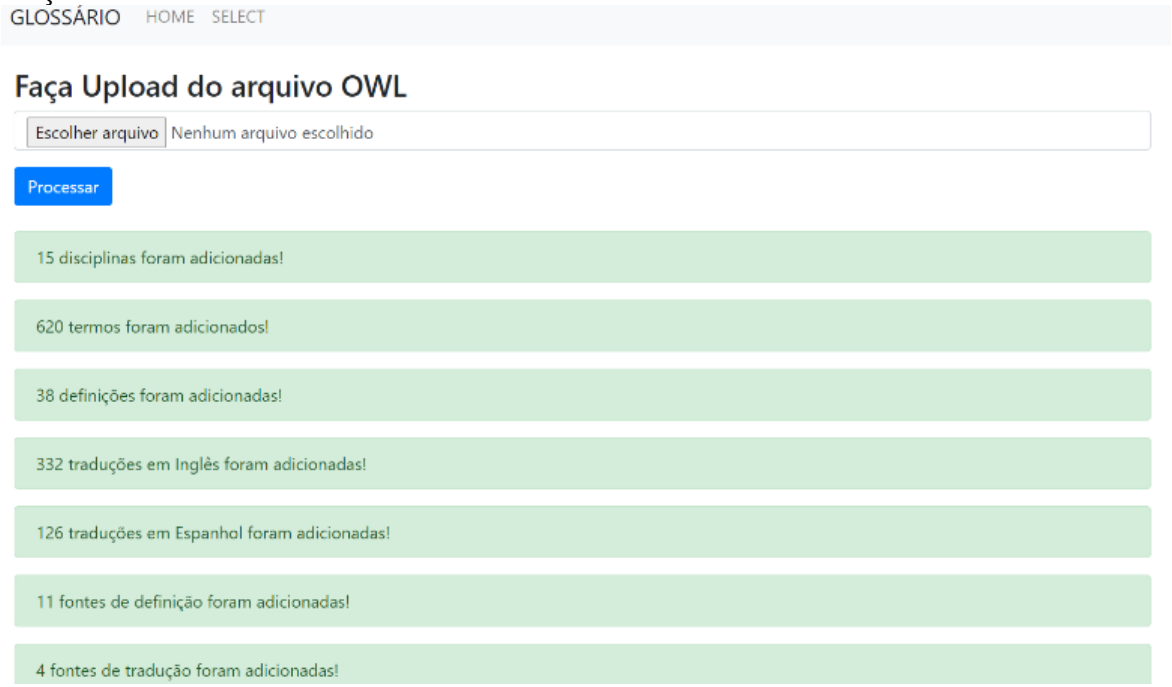

FIGURA 1. Retorno obtido com a importação do glossário de termos.

O principal resultado desse projeto é a aplicação web de importação que fornece um feedback imediato e detalhado sobre o processo de importação dos dados do glossário para o banco de dados. Essa aplicação, conectada ao banco de dados, é responsável por confirmar o sucesso da importação, além de apresentar um resumo quantitativo dos elementos que foram adicionados a cada tabela.

Durante o processo de upload do glossário de termos do curso de Mecânica, a aplicação web não apenas assegura que a importação foi realizada com êxito, mas também contabiliza e exibe o número de itens que foram inseridos em cada uma das tabelas do banco de dados. Especificamente, com a importação do glossário, foram contabilizadas 15 disciplinas, 620 termos técnicos, 38 definições e cerca de 450 traduções, que variam entre os idiomas inglês e espanhol, como ilustrado na figura 1.

Essa contagem detalhada é apresentada ao usuário de forma clara e organizada, permitindo um acompanhamento preciso do sucesso da operação e garantindo a confiabilidade do processo. A interface também oferece um ponto de verificação para assegurar que todos os dados essenciais foram importados corretamente, dando ao usuário a confiança de que o glossário está completo e devidamente integrado ao sistema. Essa funcionalidade não só facilita a manutenção do banco de dados, como também assegura que a informação disponibilizada esteja sempre atualizada e completa para consulta e uso futuro.

O código criado para manipular a estrutura do arquivo OWL de termos da mecânica utiliza arrays associativos para organizar e acessar os dados extraídos da ontologia, especificamente para o tratamento das traduções dos termos técnicos. No contexto do PHP, arrays associativos permitem associar chaves a valores, facilitando o manejo de dados complexos como os encontrados em uma ontologia OWL.

Cada elemento presente na ontologia possui uma ou mais traduções que precisam ser tratadas de maneira única. Para isso, o código utiliza o IRI, que assegura que cada elemento possa ser distinguido, mesmo que o texto em si seja semelhante a outro. O processamento lê o arquivo OWL utilizando a biblioteca SimpleXML e armazena essas índices em um array associativo, onde a chave do array é o IRI e o valor é a índice correspondente.

Finalmente, o código insere os itens ao banco de dados relacional. O IRI é utilizado como uma referência para assegurar que o elemento correto seja associada à cada relação correspondente via chave estrangeira. O identificador único gerado automaticamente pelo banco de dados (ID) é então utilizado para manter a integridade relacional.

Assim, o processo de codificação foi fundamental para garantir que, ao importar o glossário de termos técnicos, todas as informações sejam corretamente armazenadas e fiquem prontamente disponíveis para consulta e gerenciamento. A combinação de arrays associativos e laços em PHP permite uma manipulação eficiente e precisa dos dados, assegurando que o sistema seja robusto e flexível o suficiente para lidar com a complexidade da ontologia OWL.

Após a transcrição das informações da ontologia para o banco de dados, foi realizada uma consulta detalhada ao banco de dados com o objetivo de extrair as informações necessárias para cada termo. Essa consulta foi projetada para não apenas listar os termos isoladamente, mas também para revelar as relações com eles, conforme foram definidas na ontologia. Uma parte dos resultados da consulta é ilustrada na Figura 2.

termo	disciplina	definicao	traducaoEspanhol	traducaoIngles
Tempo de Espera para Quebra de Cavaco no Desbaste	Manufatura Assistida Por Computador	*NULL*	*NULL*	Dwell Time Fore Chip Breaking When Roughing
Temporização no Final da Rosca	Manufatura Assistida Por Computador	*NULL*	*NULL*	Dwell Time at Thread Depth (Chip Breakage)
Tenacidade	Tecnologia Dos Materiais	*NULL*	*NULL*	Tenacity
Tensão	Elétrica	força por unidade de área normalmente calculada co...	Tensión	Stress
Tensão	Resistência Dos Materiais E Elementos De Máquinas	força por unidade de área normalmente calculada co...	Tensión	Stress
Tensão Admissível	Resistência Dos Materiais E Elementos De Máquinas	*NULL*	*NULL*	Stress Permissible
Tensão de Cisalhamento	Tecnologia Dos Materiais	*NULL*	*NULL*	Cisalhante Stress
Tensão de Escoamento	Resistência Dos Materiais E Elementos De Máquinas	*NULL*	*NULL*	Stress Flow
Tensão de escoamento	Tecnologia Dos Materiais	*NULL*	*NULL*	Yield Stress
Console Superposição	Elétrica	*NULL*	*NULL*	*NULL*

FIGURA 2. Consulta ao banco de dados preenchido pós processamento do arquivo.

Os resultados da consulta demonstram que o objetivo pré-estabelecido foi cumprido, e que o produto final está pronto e disposto para uso quando integrado com aplicações de manutenção dos termos listados ao banco de dados agora preenchido.

O banco de dados, estruturado para refletir a ontologia original, permite que cada termo seja consultado junto com suas respectivas disciplinas, definições e traduções. A consulta elaborada foi capaz de reunir informações que ligam os termos a seus contextos dentro do curso de Mecânica, demonstrando as relações entre os termos, suas traduções em inglês e espanhol, e as disciplinas às quais estão relacionados, dentre as informações necessárias.

Esse tipo de consulta é essencial para a utilização do banco de dados, pois aproveita a riqueza de detalhes armazenados, garantindo que os usuários possam acessar informações completas e contextualmente relevantes. A capacidade de consultar o banco de dados dessa forma reflete o sucesso do projeto em transformar o glossário textual em uma estrutura de dados, que pode ser facilmente navegada e utilizada para fins educacionais e de pesquisa.

CONCLUSÕES

Os objetivos propostos no início deste projeto foram plenamente atingidos, marcando um avanço significativo na gestão e acessibilidade dos glossários técnicos do Curso Técnico em Mecânica. A importação bem-sucedida do glossário para a plataforma de banco de dados phpMyAdmin representa uma grande conquista, pois estabelece a base necessária para consultas ao conteúdo agora disposto em uma estrutura relacional e que favorece a finalização do sistema dedicado ao glossário estabelecido anteriormente (Silva; Yaguinuma, 2023).

Esse progresso não apenas otimiza o processo de consulta e manutenção dos termos, mas também viabiliza a futura integração do banco de dados com o sistema desenvolvido em projetos anteriores (Silva; Yaguinuma, 2022). Com essa integração, o glossário estará completamente preparado para ser acessado por estudantes e docentes, oferecendo uma ferramenta eficaz para a consulta de termos

técnicos, bem como favorece que futuros projetos possam garantir a atualização e manutenção contínua dos dados visando o bom funcionamento ds plataforma.

Além disso, a aplicação desenvolvida neste projeto garante o bom funcionamento do sistema, permitindo que novos glossários sejam incorporados com facilidade, sem a necessidade de processos manuais exaustivos. Assim, o projeto não só atende às demandas imediatas, mas também prepara o espaço para futuras expansões em novos projetos.

CONTRIBUIÇÕES DOS AUTORES

Daniel Fellipe Dalle Piagge trabalhou na metodologia e procedeu com os experimentos. Cristiane Akemi Yaguinuma atuou na orientação do projeto. Todos os autores contribuíram com a revisão do trabalho e aprovaram a versão submetida.

AGRADECIMENTOS

Agradecemos o Programa Institucional de Bolsas de Iniciação Científica e Tecnológica (PIBIFSP) do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) pela concessão de bolsa de estudos para realização dessa pesquisa.

REFERÊNCIAS

CALEGARI, Newton Juniano. Proposta de uma ferramenta de anotação semântica para publicação de dados estruturados na Web. 2016. 87 f. Dissertação (Mestrado) - Curso de Programa de Estudos Pós-Graduados em Tecnologias da Inteligência e Design Digital, Pontifícia Universidade Católica de São Paulo, São Paulo, 2016. Disponível em: https://tede2.pucsp.br/handle/handle/18992. Acesso em: 06 set. 2024.

DA SILVA, Rafael Matias; YAGUINUMA, Cristiane. Aplicaç ão Web para Gerenciamento de Glossários Utilizados em Cursos Técnicos Profissionalizantes. In: VIII ENCONTRO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA-EnICT (ISSN: 2526-6772). 2023.

SILVA, R. M. D.; YAGUINUMA, C. A. OntoMec: Aplicação Web Baseada em Serviços Web para Consultas a Ontologia sobre Termos da Área de Mecânica. In: ENCONTRO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA, 7., 2022, Araraquara. Anais do VII Encontro de Iniciação Científica e Tecnológica (EnICT). Araraquara: IFSP, 2021. Disponível em: https://arq.ifsp.edu.br/eventos/index.php/enict/7EnICT/paper/viewFile/582/352. Acesso em: 29 jun. 2022.

JESUS, Vanessa Marta de. Dados abertos conectados a partir de catálogos online de bibliotecas. 2021. 148 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2021. Disponível em: https://repositorio.ufmg.br/handle/1843/39188. Acesso em: 06 set. 2024.

HORRIDGE, Matthew. A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools. Manchester: University Of Manchester, 2011. Disponível em: http://mowlpower.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf. Acesso em: 06 set. 2024.

MACHADO, Luís Miguel Oliveira. Ontologias, dos sistemas aos modelos: Uma abordagem introdutória no contexto dos sistemas de organização do conhecimento. Fronteiras da Representação do Conhecimento, v. 1, n. 1, p. 1-18, 2021.

PIZA, G. V.; YAGUINUMA, C. A. Construç ão de Ontologia sobre Termos da Área de Mecânica Utilizados no Curso Técnico em Mecânica Integrado ao Ensino Médio. Araraquara: IFSP, 2020.

REIS, C. F.; DANTAS, G. L. Relação entre Língua em um Glossário da Área de Mecânica. III Encontro de Iniciação Científica e Tecnológica - EnICT, 2018, Araraquara. Anais do III Encontro de Iniciação Científica e Tecnológica - EnICT, 2018.