

1. Identificação do Projeto

Título do Projeto PIBIC/PAIC

ONTOLOGIAS DE VOCABULÁRIOS PARA INDEXAÇÃO DE PROJETOS DE PIBIC: Estudo de caso da iniciação científica na Ufam.

Orientador

Kátia Viana Cavalcante

Aluno

Danielle Chaves Carmim

2. Informações de Acesso ao Documento

2.1 Este documento é confidencial?

SIM

NÃO

2.2 Este trabalho ocasionará registro de patente?

SIM

NÃO

2.3 Este trabalho pode ser liberado para reprodução?

SIM

NÃO

**2.4 Em caso de liberação parcial, quais dados podem ser liberados?
Especifique.**

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

**ONTOLOGIAS DE VOCABULÁRIOS PARA INDEXAÇÃO DE PROJETOS DE
PIBIC: Estudo de caso da Iniciação Científica na UFAM.**

Bolsista: Danielle Chaves Carmim, CPNq

MANAUS

2016

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL

PIB – SA – 0140/2015

**ONTOLOGIAS DE VOCABULÁRIOS PARA INDEXAÇÃO DE PROJETOS DE
PIBIC: Estudo de caso da Iniciação Científica na UFAM.**

Bolsista: Danielle Chaves Carmim, CPNq

Orientadora: Profa. Dra. Kátia Viana Cavalcante

MANAUS

2016

RESUMO

O presente relatório visa discutir sobre as ontologias de vocabulários para indexação de projetos de Pibic: Estudos de casos da Iniciação científica na UFAM. O objetivo da pesquisa é construir um conjunto de termos que possibilitem a catalogação e classificação dos projetos do Programa de Iniciação Científica. A metodologia executada foi exploratória e descritiva sendo priorizado o estudo de caso de cada projeto do triênio 2013, 2014 e 2015. Os resultados obtidos no levantamento serviram de base para interpretação dos aspectos das terminologias, ontologias e a indexação apresentadas nos projetos. A análise está disposta em forma de texto, e apresenta as informações de maior significância para com os objetivos propostos no projeto desta pesquisa.

Palavras-chave: Projetos. Ontologias. Iniciação científica.

SUMÁRIO

1	INTRODUÇÃO	6
2	ONTOLOGIAS	Erro! Indicador não definido.
2.1	USO DE ONTOLOGIAS.....	Erro! Indicador não definido.
3	ESTRATEGIA METODOLOGICA.....	Erro! Indicador não definido.
4.	RESULTADOS.....	Erro! Indicador não definido.
5	CONCLUSÃO	Erro! Indicador não definido.
	REFERÊNCIAS.....	Erro! Indicador não definido.



1 INTRODUÇÃO

Nos últimos anos, conforme o aumento exponencial dos dados disponíveis tem-se conferido importância fundamental e significativa às técnicas de estruturalização/organização da informação. Entre os diversos tipos de estruturação utilizadas nas organizações da informação, se faz necessário a utilização de termos estruturais que se arranjam na classificação e na criação das categorias entre os cabeçalhos de assuntos e esquemas de classificação (ou taxonomias). Há estruturas que se organizam a partir de conceitos e de seus relacionamentos entre: ontologias, indexação e tesouro, dentro dos vocabulários existentes na plataforma e mapeamento existente na área de conhecimento comum de usuários e pesquisadores.

A utilização de ferramentas adequadas para a representação das informações contidas nos permite ainda, a redução do tempo de resposta no processo de busca, além de eliminar ambiguidades criadas na terminologia em razão de sinônimos, homógrafos, diferenças de escrita, formas das palavras, que normalmente se transformam em barreiras nos sistemas de comunicação, armazenamento e recuperação de informação. As estruturas de indexação são ferramentas que precisam ser previamente definidas, de modo a serem direcionadas para atender às demandas dos usuários e ao mesmo tempo agregar valor à informação. Na medida em que a indexação trabalha com a representação do conteúdo temático dos documentos, utiliza termos de vocabulários controlados, listas de termos significativos e a própria linguagem natural, de forma a possibilitar o armazenamento, as estratégias de busca, e a pesquisa e recuperação das informações, tendo como ponto de acesso os termos que lhes foram atribuídos. Na complexidade da evolução tecnológica dos dados, se faz necessário uma abordagem na utilização das ontologias na organização do conteúdo das fontes de dados.

A ontologia criada por especialistas e definida por regras, regula a combinação entre temas e relação de um domínio do conhecimento. Sendo que, o usuário que formula consultas usa conceitos definidos pelas ontologias. Desta forma, melhorias nos processos de recuperação de informação. E com a meta de aperfeiçoar o processo de recuperação a proposta que norteia é este projeto é o de construir um conjunto de termos que possibilitem a catalogação e classificação nos projetos de PIBIC do Programa de Iniciação



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



a Pesquisa da Universidade Federal do Amazonas, visando à construção de um sistema de acompanhamento e modernização do referido programa. Desta forma, buscando maior embasamento e internalização da temática abordada, a execução do projeto foi direcionada para dois momentos, ou seja, o primeiro abrangeu o universo conceitual da temática, enquanto que a segundo percorreu o sentido prático, a partir do mapeamento e análise do banco de dados do Portal Lira no triênio 2013, 2014 e 2015, identificar quantos projetos eram trabalhados na área do Campus Universitário, uma vez que a pesquisa possuía abordagem quali-quantitativa. Os dados foram tratados quantitativamente, sendo interpretados de modo a construir um panorama da distribuição desses projetos no campus Manaus e qualitativamente a geração de termos para catalogação dos projetos. Desta feita, espera-se contribuir de forma norteadora para as discussões ontológicas das propriedades do assunto, e alavancar o interesse de profissionais da informação pelo tema.

2. ONTOLOGIAS

Os conceitos existentes de ontologias, sendo estas inúmeras, que compõem as classes gramaticais e a terminologia dos vocabulários. O termo ontologia possui um sentido especial na organização da informação existente, com diversas definições apresentadas. Dentre elas, utilizaremos a apresentada por Sowa, uma vez que ontologias definem-se como categorias para as coisas que existem em um mesmo domínio. Na conspeção do autor, ontologia é um “catálogo de tipos de coisas” em que se supõe existir um domínio, na perspectiva de uma pessoa que usa uma determinada linguagem (Sowa, 1999).

Ainda do ponto de vista conceitual, Borst (1997) explica a questão da ontologia de forma relevante e extremamente atual, uma vez que a sociedade da informação requer cada vez mais compartilhamento de termos e ações, vejamos:

“Uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Nessa definição, “formal” significa legível para computadores; “especificação explícita” diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas, explicitamente definidos; “compartilhado” quer dizer conhecimento consensual; e “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real. (BORST, 1997, p. 12).



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



Na sociedade contemporânea, os sistemas convencionais de consultas utilizam técnicas de bases sintáticas sobre uma forma de adequação léxica, mais do que uma aplicação da base de conhecimento do campo de interesse. Em muitos domínios, o usuário está interessado em encontrar informação onde a relevância dos documentos não pode ser medida de mediante o uso de sistemas de busca por palavras-chaves (Keywords) (BECKS, 1998). A relevância deve ser preferencialmente estimada num nível de conhecimento profundo do domínio de problema em questão. Neste sentido, algumas propostas envolvem a criação de metadados que seguem modelos de ontologias. O termo “ontologia” é emprestado da filosofia, onde uma Ontologia é uma descrição da Existência. Na Ciência da Computação, vem sendo aplicado desde o início da década de 90 na área de inteligência artificial para representação computacional de conhecimento em áreas como engenharia de conhecimento e processamento de linguagem natural. Para sistemas de Inteligência Artificial, o que “existe” é o que pode ser representado. Neste contexto, podem ser entendidas como uma especificação formal e explícita de uma conceitualização consensual (GRUBER, 1993).

Desta forma, podemos admitir que uma ontologia se assemelhe a uma linguagem documental codificada e controlada, concebendo-se na sua ordenação uma maneira diferente levando em conta certas peculiaridade e propriedade dos termos.

2.1 O USO DE ONTOLOGIAS

As ontologias podem ser utilizadas em diversas áreas (comunicação, biblioteconomia etc.), sendo na investigação do objeto da linguagem (campo) e no estudo dos vocábulos (processo), desta forma, podemos observar as principais necessidades de utilização da ontologia no texto.

Com a evolução e construção da Web, a informação está ao passo de um click, mas este enorme volume de informação constituem um fator complicado para quem busca a informação específica e objetivamente útil, desta forma surge a problematização e necessidade eficaz na pesquisa da informação, capaz de tratar semanticamente a informação contida nos documentos e estruturando através das ontologias no banco de dados.

Jackson e Moulinier (2002) observaram a tendência e a necessidade do desenvolvimento de programas capazes de executar automaticamente tarefas múltiplas relacionadas na recuperação de informação, tais como selecionar documentos de uma



base e agrupá-los em categorias ou classes e deles extrair determinados conjuntos de informação.

Acredita-se que a utilização das ontologias aliada aos sistemas capazes de inferir sobre a relevância da informação pode fornecer uma compreensão mais apurada do domínio, resultando numa classificação textual eficiente. Desta união Baseada em conhecimento favorece a manipulação de informação, entendida como a realização de mais de uma tarefa, como por exemplo, classificação, recuperação e extração, proporcionando flexibilidade nas ferramentas de diferentes domínios.

Observa-se que a vantagem para troca de domínio, basta trocar as ontologias. Contudo, Guarino (1998), afirma que ontologia é como um ramo da filosofia se refere a um sistema de categorização dos objetos do mundo para a organização da realidade. Assim, ontologias especificam conceitos sobre um dado campo de conhecimento e as relações, restrições e axiomas válidos entre esses conceitos e suas instâncias, provendo uma boa taxonomia tanto de entidades como de instâncias, desde universo.

Atualmente, o uso de ontologias vem se difundindo em diversas áreas com domínios como gestão de conhecimento, comércio eletrônico, processamento e linguagens naturais, recuperação da informação na Web de cunho educacional e outras áreas que buscam desenvolver um vocabulário contendo os conceitos relativos ao domínio de aplicação.

3. ESTRATÉGIA METODOLÓGICA

A pesquisa executada classificou-se como exploratória e descritiva sendo utilizado o estudo de caso, por tratar-se de uma investigação empírica que pretende descrever um contexto real para que haja uma intervenção institucional.

Mediante solicitação formal de acesso ao Portal Lira a PROPESP, a petição foi analisada e aprovada pelo setor jurídico, a partir da indicação dos itens específicos aos quais teríamos acesso, ou seja, introdução, objetivos e desenvolvimento, e somente os projetos executados na capital, dentro dos padrões de análises de interesses para uma catalogação.

Após o cadastramento e manuseio do Portal, identificamos que o sistema de análise do portal Lira não fornecia relação total de projetos aprovados por unidade e, sim um total de projetos que foram submetidos e analisados – partindo do código. Desta



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



UFAM

forma, geramos planilhas no Excel, de forma separar os projetos aprovados para análise por ano e área do conhecimento. Posteriormente de posse do Código do projeto acessá-lo no Portal.

Ao acessá-lo buscamos identificar a partir da metodologia que envolve o levantamento dos termos utilizados na organização e estruturação de thesaurus, com base em padrões internacionais e num conjunto de critérios científicos previamente definidos, como por exemplo, local de execução da proposta de forma a correlacioná-lo com a área do campus Manaus.

A seguir os resultados serão apresentados por ano, com descrição das dificuldades encontradas e o total de projetos executados na área do campus.

4 RESULTADOS

Ao categorizar cada área, foi possível identificar e traçar passos para localizar as respostas desejadas. A busca nos itens indicados – introdução, objetivo e metodologia, nem sempre continham respostas referentes áreas e local de pesquisa, tendo como maior dificuldade a leitura de todo o projeto e muitas vezes não encontrado dados relacionados à pesquisa e local.

Alguns projetos continham títulos repetidos ou como em muitos casos, com mudança total do título, no entanto o conteúdo era o mesmo apresentado em ambos, vide exemplo aleatório coletado na base do Portal Lira: Fungos em silagem de resíduo úmido de cervejaria com inclusão gradual de grãos de milho (Título analisado pelo comitê)/ Fungos anemófilos em uma granja experimental de Manaus (Título postado no portal Lira). Essa checagem tomou tempo considerável, pois acabávamos lendo duas vezes o mesmo conteúdo.

Outros projetos explanavam a área técnica sem indicar o local de coleta, simplesmente os locais de análise, tipo no laboratório x ou y. Em muitos dos projetos analisados, somente foi possível encontrar os dados na justificativa e ou conclusão.

O que nos levou a justificar junto a departamento de pesquisa, a necessidade da leitura integral do projeto, de forma a detectar os dados necessários para relacionar e diferenciar locais aja vista o documento de autorização emitido pela procuradoria jurídica preocupada com os direitos de segurança da informação.



2013

No ano de 2013, foram consultados e analisados 628 (seiscentos e vinte e oito) projetos aprovados. Esses projetos foram distribuídos em seis áreas do conhecimento conforme tabela a seguir.

Tabela 1: Quantidade de projetos consultados para o ano de 2013.

ÁREA DO CONHECIMENTO	TOTAL DE PROJETOS
Agrárias	83
Biológicas	74
Saúde	120
Ciências humanas	105
Ciências sociais aplicadas	128
Ciências exatas da terra e engenharias	118

Fonte: Coleta de campo. 2015

Nos projetos da área de Agrárias identificamos uma sobreposição ou falta de clareza em relação à área de estudo (AE) e os locais de coleta (LC). Assim como, entre os locais de coleta (LC) e locais de análise (LA).

Em muitos dos projetos a área de estudo e os locais de coleta eram repetidos, sendo que deve ocorrer distinção. Após o filtro identificamos a seguinte distribuição por áreas de estudo (AE):

- 38 (trinta e oito) projetos tiveram como AE diversos municípios do Estado;
- somente 39 (trinta e nove) projetos tiveram como AE o município de Manaus;
- 06 (seis) projetos foram desenvolvidos na AE da Fazenda da Ufam.

Os locais de coleta foram assim identificados:

- 23 (vinte e três) projetos com LC em municípios;
- 06 (seis) com LC na cidade de Manaus;
- 03 (três) projetos com LC no INPA;
- 01 (um) na Embrapa;
- 37 (trinta e sete) projetos com LC em laboratórios da UFAM;
- 12 (doze) na fazenda experimental da Ufam;
- 03 (três) projetos com LC na granja;



h) 01 (um) sem identificação.

Na área Biológica os problemas se repetiram, ou seja, sobreposição de local de coleta com local de análise. Registra-se que alguns projetos, não possuíam identificação para AE bem como LC ou LA. Outro fator identificado nesta área diz respeito a grande abrangência da AE, por exemplo: Manaus, mas o LC é laboratório. Observa-se que o conceito de área de estudo não específica.

Como resultado para área do conhecimento Biológica, temos como AE a seguinte distribuição:

- a) 21 projetos com AE em diversos municípios do Estado;
- b) 43 (quarenta e três) projetos com AE no município de Manaus;
- c) 04 (quatro) projetos com AE na Fazenda da Ufam
- d) 01 (um) projeto com AE na Granja da Ufam
- e) 01 (um) projeto com AE na floresta do Campus
- f) 04 projetos sem identificação de AE

Sendo que os locais de coleta desses projetos estão assim distribuídos:

- a) 03 projetos com LC em município do Estado;
- b) 09 projetos com LC no INPA (claramente análise);
- c) 03 projetos com LC na Fazenda da Ufam;
- d) 01 projeto com LC na Granja da Ufam;
- e) 39 projetos com LC em laboratórios do UFAM;
- f) 01 projeto com LC no Hospital Universitário
- g) 04 projetos com LC na Floresta do Campus
- h) 18 projetos sem identificação do LC

As demais áreas apresentam os mesmos conflitos já identificados. Destaca-se que na área das Ciências Humanas, grandes partes dos projetos indicam como local de coleta as referências bibliográficas.

Nos projetos da área de saúde não ocorre variação quanto a AE, todos apontam cidade de Manaus e a grande maioria dos cento vinte projetos trazem como LC os hospitais da cidade.



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



No ano 2014 identificou-se uma diminuição na quantidade de projetos aprovados somente 589 (quinhentos e oitenta e nove) projetos, no entanto ocorre um aumento na quantidade de projetos na área das ciências exatas e engenharias.

Tabela 2: Quantidade de projetos consultados para o ano de 2014.

Área do conhecimento	Total de projetos
Agrárias	82
Biológicas	53
Saúde	94
Ciências Humanas	121
Ciências Sociais Aplicadas	105
Ciências Exatas da Terra e Engenharias	134

Fonte: Coleta de campo. 2015

A maioria das dificuldades apresentadas nos anos anteriores, também foi detectada em 2015 e, acrescidas ao fato do calendário acadêmico a típico pelo que passou a universidade, o fato do processo está em andamento ante as inclusão e exclusão de projetos, o que dificultou a esquematizar todos os projetos em aberto.

2015

No ano de 2015 foram avaliados 584 (quinhentos e oitenta e quatro) projetos distribuídos nas seguintes áreas conforme tabela abaixo.

Tabela 3. Quantidade de projetos consultados para o ano de 2015

Área do conhecimento	Total de projetos
Agrárias	103
Biológicas	58
Saúde	106
Ciências Humanas	94
Ciências Sociais Aplicadas	127
Ciências Exatas da Terra e Engenharias	96

Fonte: Coleta de campo. 2015



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



Ao buscar e arrolar os projetos Pibic dos anos 2013, 2014 e 2015, foi possível comprovar, que além das limitações/restrições relacionada ao sistema (Portal Lira), observamos que os projetos de pesquisa não apresentavam uniformização de dados de modo a permitir determinadas análises gerencias do ponto de vista administrativo e de gestão da informação.

Destarte as ontologias como especificações formais e explícitas de conceitualizações compartilhadas capturam e explicitam o vocabulário utilizado. Desta forma sugerimos que sejam incluídos nos formulários para apresentação de projetos junto ao Portal Lira, os seguintes campos em que o pesquisador deverá indicar os temas listados a seguir, de forma a facilitar a recuperação da informação.

Os projetos deverão apresentar:

a) Especificação geral conhecimento

- Grande área do conhecimento;
 - Área do conhecimento
 - Subárea do conhecimento
 - Disciplina

b) Especificação de Localização

- **ÁREA DE ESTUDO**
 - Município (indicar as subclasses: Cidade; Bairro; comunidade)
 - Instituição (secretarias; institutos; ong's)
 - Áreas de proteção (Apas, reservas)
 - Dependência das Ufam (fazenda; prédios, floresta; laboratórios)
- **LOCAIS DE COLETA**
 - Município (indicar as subclasses: Cidade; Bairro; comunidade)
 - Instituição (secretarias; institutos; ong's)
 - Áreas de proteção (Apas, reservas)
 - Dependência das Ufam (fazenda; prédios, floresta; laboratórios)

Outra sugestão é adotar a padronização de cadastro dos orientadores e orientandos utilizando apenas as iniciais. Este recurso mediante link pode realizar o cruzamento com a Plataforma Lattes facilitando a recuperação da informação e o compartilhamento de publicações.



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



O uso de um dicionário de palavras-chave pré-indicadas possibilita a coerência na escolha dos termos. As áreas do conhecimento já possuem Thesaurus, basta disponibilizar os termos no Portal Lira, de forma a uma homogeneização evitando duplicidade de termos.

CONCLUSÃO

Com o avanço tecnológico e da automatização, ficou possível agilizar os processos operacionais, ressaltando as potencialidades pouco exploradas de forma viabilizar melhores formas de acesso às informações de interesses dos usuários.

A uniformização terminológica técnica é importante para facilitar a comunicação, promover a pesquisa e a valorização do patrimônio científico no momento da recuperação. Acresce que essa uniformização é cada vez mais relevante devido às bases de dados em suporte informático, à gestão moderna e eficiente de coleções e à acessibilidade das coleções online.

As ontologias e tesouros são compostos de uma estrutura de classes e subclasses. As classes ontológicas representam o que chamamos de “mundo real”, que podem ser agrupados e categorizados de acordo com suas similaridades, mediante suas considerações de domínio concreto, tais entidades representam físicas ou conceituais, sendo objetos inanimados até teorias científicas/teóricos, sendo que uma classe pode possuir diversas subclasses, que podem organizar as entidades a partir de características mais específicas.

Assim sendo, esperamos ter contribuído com a especificação formal de vocabulários a ser utilizado nos projetos de pesquisa, permitindo maior rapidez na recuperação da informação.

Referências

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

BECKS, Andreas; SKLORZ, Stefan; TRESP, Christopher. Sematic structuring and visual querying of document abstracts in Digital Libraries. In: EUROPEAN CONFERENCE ON DIGITAL LIBRARIES, ECDU, 2, 1998. Heraklion. **Research and advanced technology for digital libraries**: proceeding. Berlin: Springer-Verlag, 1998 p. 443-458

FRAGOSO, Suely; RECUERO, Raquel; AMARAL, Adriana. **Métodos de pesquisa para internet**. Porto Alegre: Sulina, 2013. 239 p. (Coleção Cibercultura).



UNIVERSIDADE FEDERAL DO AMAZONAS

RELATÓRIO FINAL PIBIC/PAIC 2015-2016



GENESERETH, M.R.; NILSSON.N.S. **Logical foundation of artificial intelligence**. Los Altos. CA: Morgan Kaufmann, 1987

GRUBER, T. R. **Towards principles for the design of ontologies used** for knowledge sharing. Stanford University, Knowledge Systems Laboratory Technical Report KSL93-04. Disponível em: <<http://gicl.mcs.drexel.edu/people/regli/Classes/KBA/Readings/KSL-93-04.pdf>>. Acesso em: 10 jan. 2015

MUELLER, Suzana P. M. (Org.). **Métodos de pesquisa em ciência da informação**. Brasília: Thesaurus, 2007. 190 p. (Série Ciência da Informação e da Comunicação).

SANTAELLA, Lucia. **Linguagens líquidas na era da mobilidade**. São Paulo: Paulus, 2007. 468 p. (Coleção Comunicação).

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, v. 1, n.1, p. 41-62, jan./jun. 1996.