

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**MÉTRICAS PARA ONTOLOGIAS:
REVISÃO SISTEMÁTICA E
APLICAÇÃO AO PORTAL ONTOLP**

LARISSA ASTROGILDO DE FREITAS

Dissertação apresentada como requisito parcial
à obtenção do grau de Mestre em Ciência da
Computação na Pontifícia Universidade Católica
do Rio Grande do Sul.

Orientadora: Renata Vieira

**Porto Alegre
2010**

Dados Internacionais de Catalogação na Publicação (CIP)

F866m Freitas, Larissa Astrogildo de
Métricas para ontologias : revisão sistemática e aplicação ao
portal OntoLP / Larissa Astrogildo de Freitas. – Porto Alegre,
2010.
96 f.

Diss. (Mestrado) – Fac. de Informática, PUCRS.
Orientador: Prof^a. Dr^a. Renata Vieira.

1. Informática. 2. Linguística Computacional. 3. Ontologia.
I. Vieira, Renata. II. Título.

CDD 006.35

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Métricas para Ontologias: Revisão Sistemática e Aplicação ao Portal Ontolp", apresentada por Larissa Astrogildo de Freitas, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Inteligência Computacional, aprovada em 15/03/10 pela Comissão Examinadora:

Profa. Dra. Renata Vieira -
Orientadora

PPGCC/PUCRS

Profa. Dra. Vera Lúcia Strube de Lima -

PPGCC/PUCRS

Profa. Dra. Maria Luiza Machado Campos -

UFRJ

Homologada em 29/06/10, conforme Ata No. 011 pela Comissão Coordenadora.

Prof. Dr. Fernando Gehm Moraes
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 - P32- sala 507 - CEP: 90619-900
Fone: (51) 3320-3611 - Fax (51) 3320-3621
E-mail: ppgcc@pucrs.br
www.pucrs.br/facin/pos

*“Mais importante do que
esperar ou procurar
oportunidades, é criá-las.”
Legrand*

AGRADECIMENTOS

À minha família Edú, Elaine, Eduardo e Bianca, pelo incentivo.

Ao meu namorado Ulisses, pela paciência.

À minha orientadora Renata, por tudo que aprendi.

Aos colegas do laboratório PLN Rodrigo, Sandra, Thiago, Josiane, Silvia, Anderson, Paulo, Douglas, Fernando, Igor, Roger, Lucelene, Kamila, Daniel e Guilherme, em especial, Tatiane, Patrícia, Clarissa e Mírian pelas dicas e pelo companheirismo.

Aos professores do PPGCC da PUCRS Vera e Marcelo, aos professores do ICMC da USP São Carlos Tiago, Sandra e Graça, aos professores do DL e do DC da UFSCar Ariani, Gladis, Helena e Lúcia pelas sugestões.

À CAPES e ao CNPq, pelo apoio financeiro durante o mestrado.

MÉTRICAS PARA ONTOLOGIAS: REVISÃO SISTEMÁTICA E APLICAÇÃO AO PORTAL ONTOLP

RESUMO

Ao tratarmos da construção e da disponibilização de ontologias nos deparamos com as seguintes limitações: a construção de ontologias exige a presença de especialistas do domínio e portanto apresenta um custo elevado, muitas ontologias podem conter dados protegidos por algum tipo de direito o que dificulta sua disponibilização, além disso para aquelas que podem ser compartilhadas existe o problema como encontrá-las.

O Projeto OntoLP visa atenuar os problemas acima, uma vez que se propõe a colaborar na disponibilização de ontologias em língua portuguesa, ferramentas e publicações relacionadas.

Dado um conjunto de ontologias é importante reunir informações que ajudem os pesquisadores a conhecer os recursos disponíveis de forma ágil. Uma área em desenvolvimento neste cenário é a área de métricas para ontologias, porém existem muitas propostas, não padronizadas, de métricas na literatura.

Sendo assim, procuramos neste trabalho fazer um levantamento exaustivo sobre métricas para ontologias através de uma revisão sistemática. Depois de analisadas as métricas, um sistema que realiza o cálculo de métricas em ontologias OWL foi desenvolvido e aplicado ao portal de ontologias OntoLP. O objetivo é ajudar na busca por ontologias. Por fim, o uso das métricas no portal foi avaliado por meio de um questionário.

Os resultados do questionário de maneira geral se mostraram satisfatórios, uma vez que foi possível detectar um alto nível de concordância perante respostas das questões sobre a utilidade das métricas em repositórios específicos para ontologias.

Palavras-chave: Ontologia; Métricas; Língua Portuguesa.

METRICS FOR ONTOLOGIES: SYSTEMATIC REVIEW AND APPLICATION IN ONTOLP PORTAL

ABSTRACT

Ontology construction and availability face various constraints such as the necessity of domain experts (leading to high costs). Besides that ontologies may have protected data, which can prevent it from being publicly available. In addition to that, there is the problem of where to find the ontologies, once they are not restricted.

In face of that, the OntoLP Project aims to collaborate in the task of sharing ontologies in Portuguese language, as well as ontology research related tools and publications.

Given an ontology set it is important to gather information that can support researchers to inspect the available resources in a general way. In this context, metrics for ontology is a novel area. Although there are many proposal for metrics in the literature, they are not standardized.

So, this work aims to do a survey about metrics for ontologies by doing a systematic review of the literature. After this exhaustive search the gathered metrics were selected and implemented to be applied to a group of ontologies written in OWL. The goal was giving support to the reuse of Portuguese ontologies. The usage of metrics for ontologies in the OntoLP portal was evaluated through a survey.

The survey results were satisfactory, the utility of metrics for ontologies in specialized repositories for ontologies was considered relevant by the portal potential users.

Keywords: Ontology; Metrics; Portuguese Language.

LISTA DE FIGURAS

Figura 2.1	Visão da Web Semântica. Fonte: adaptado [ANT08]	30
Figura 2.2	Categorização de ontologias segundo Guarino. Fonte: adaptado [GUA98] . . .	31
Figura 2.3	Categorização de ontologias segundo Lassila. Fonte: adaptado [LAS01] . . .	32
Figura 3.1	Etapas da revisão sistemática. Fonte: adaptado [BIO05]	39
Figura 4.1	Portal OntoLP descrição das métricas.	56
Figura 4.2	Portal OntoLP gráfico das métricas.	56
Figura 5.1	Retorno para as questões relativas às métricas implementadas.	59
Figura B.1	Ontologia de Ecologia.	69
Figura B.2	Ontologia de Nanotecnologia.	70
Figura B.3	Ontologia de Arte Contemporânea.	71
Figura B.4	Ontologia de Currículo Lattes.	72
Figura B.5	Ontologia de Emoção.	73
Figura B.6	Ontologia de Privacidade de Dados.	74
Figura B.7	Ontologia de Rede de Conhecimento Científico (RCC).	75
Figura B.8	Ontologia de Smartphone.	76
Figura B.9	Ontologia de Música.	77
Figura B.10	Ontologia de Equivalência de Estímulos (EE).	78
Figura C.1	Número de classes raiz.	83
Figura C.2	Número de classes folha.	83
Figura C.3	Número total de classes.	84
Figura C.4	Número de propriedades (atributos).	84
Figura C.5	Número de propriedades (relações).	85
Figura C.6	Número de propriedades.	85
Figura C.7	Número total de instâncias.	86
Figura C.8	Profundidade máxima.	86
Figura C.9	Profundidade média.	87
Figura C.10	Média da profundidade máxima pela profundidade média.	87
Figura C.11	Largura máxima.	88
Figura C.12	Largura média.	88
Figura C.13	Média do número de propriedades (atributos) pelo número total de classes. .	89
Figura C.14	Média do número de propriedades (relações) pelo número total de classes. .	89
Figura C.15	Média do número de propriedades pelo número total de classes.	90
Figura C.16	Média do número de classes folha pelo número total de classes.	90
Figura C.17	Média do número de classes folha pelo número de classes raiz.	91
Figura C.18	Média do número de classes populadas pelo número total de classes.	91
Figura C.19	Média do número total de instâncias pelo número total de classes.	92

Figura C.20 Desvio padrão do número de total instâncias pelo número total de classes. . 92

LISTA DE TABELAS

Tabela 3.1	Fontes selecionadas.	42
Tabela 3.2	Retorno para cada <i>string</i> de busca em cada fonte selecionada.	43
Tabela 3.3	Retorno para cada fonte selecionada em cada etapa.	44
Tabela 3.4	Apresentação dos artigos selecionados.	45
Tabela 3.5	Extração de informações gerais sobre os trabalhos.	47
Tabela 3.6	Extração de informações gerais sobre as ontologias.	47
Tabela 3.7	Extração de informações gerais sobre as métricas.	47
Tabela B.1	Resultado da aplicação das métricas.	79
Tabela D.1	Resultado questionário (pré-teste).	95
Tabela D.2	Resultado questionário (final).	96

LISTA DE FRAGMENTOS

Fragmento 2.1 Exemplo de classes raiz.	34
Fragmento 2.2 Exemplo de referências.	34
Fragmento 2.3 Exemplo de propriedades (relações).	35
Fragmento 2.4 Exemplo de propriedades (atributos).	35
Fragmento 2.5 Exemplo de instâncias.	35

LISTA DE SIGLAS

SHOE	<i>Simple HTML Ontology Extension</i>
OIL	<i>Ontology Interchange Language</i>
OWL	<i>Web Ontology Language</i>
W3C	<i>World Wide Web Consortium</i>
WWW	<i>Word Wide Web</i>
URI	<i>Uniform Resource Identifier</i>
XML	<i>Extensible Markup Language</i>
NS	<i>Namespaces</i>
RDF	<i>Resource Description Framework</i>
HTML	<i>HyperText Markup Language</i>
DARPA	<i>Defense Advanced Research Projects Agency</i>
DAML	<i>DARPA Agent Markup Language</i>
GPL	<i>General Public Licence</i>
HP	<i>Hewlett-Packard</i>
CSS	<i>Cascading Style Sheets</i>
GUI	<i>Graphical User Interface</i>

SUMÁRIO

1. Introdução	25
1.1 Motivação e Contexto da Dissertação	25
1.2 Objetivo da Dissertação	27
1.2.1 Objetivo Geral	27
1.2.2 Atividades Realizadas para Atingir o Objetivo Geral	27
1.3 Organização da Dissertação	27
2. Fundamentação Teórica	29
2.1 Web Semântica	29
2.2 Ontologias	31
2.2.1 Linguagens para Ontologias	32
2.2.2 Levantamento de Ontologias em Português	35
2.3 Métricas	36
3. Revisão Sistemática	39
3.1 Planejamento da Revisão	39
3.1.1 Formulação da Questão	40
3.1.2 Seleção das Fontes	42
3.2 Condução da Revisão	43
3.2.1 Seleção dos Estudos	43
3.2.2 Extração de Informação	45
3.3 Publicação dos Resultados	46
4. Protótipo	49
4.1 Bibliotecas Utilizadas	49
4.2 Ontologias Utilizadas	49
4.3 Métricas Implementadas	52
4.4 Protótipo no Portal OntoLP	55
5. Avaliação	57
5.1 Elaboração do Questionário	57
5.2 Aplicação do Questionário	58

6. Considerações Finais	61
6.1 Contribuições	61
6.2 Trabalhos Futuros	61
Bibliografia	63
Apêndice A. Protocolo Elaborado	67
Apêndice B. Resultado da Aplicação das Métricas	69
Apêndice C. Questionário Elaborado	81
Apêndice D. Resultado da Aplicação do Questionário	95

1. Introdução

Neste capítulo apresentamos a motivação, o contexto, o objetivo e a organização desta dissertação.

1.1 Motivação e Contexto da Dissertação

O conceito de ontologia remonta à época de Aristóteles e seus discípulos, os metafísicos, que estudavam a natureza do ser, a existência ou a realidade em geral, bem como categorias básicas do ser e suas relações. Apesar desta longevidade do conceito, a palavra em si só surgiu alguns séculos depois, em 1613, derivada dos termos gregos *ontos*, ser, e *logos*, discurso, sob o cunho de Rudolf Gockel [WEL01].

Na década de 80 pesquisadores de Inteligência Artificial, mais especificamente da área de representação do conhecimento, notaram que trabalhar em ontologias era relevante aos sistemas inteligentes para o processo de descrição do mundo, para que estes pudessem raciocinar e atuar no mundo [WEL01].

Uma definição recorrente em publicações é de que ontologias sejam especificações explícitas de conceitualizações. Onde conceitualizações são pontos de vista, resumos simplificados do mundo que desejamos representar para algum propósito [GRU95].

Durante a última década as ontologias têm sido amplamente utilizadas em aplicações relacionadas a diversas áreas, como gestão do conhecimento, processamento de linguagem natural, comércio eletrônico, integração inteligente de informação, recuperação da informação, integração de bases de dados, bioinformática, ensino e *Web Semântica* [GOM04].

Inicialmente estas estruturas foram construídas usando técnicas de modelagem de Inteligência Artificial baseada em frames e lógica de primeira ordem. Nos últimos anos outras técnicas de representação do conhecimento fundamentadas em lógica de descrição têm sido utilizadas. É importante salientar que no contexto da *Web Semântica* novas linguagens de lógica de descrição surgiram tais como SHOE, OIL, DAML + OIL e OWL. Estas linguagens são usualmente chamadas de linguagens de ontologia baseada na *Web* ou linguagens de marcação de ontologia [GOM04].

Em fevereiro de 2004 OWL tornou-se uma recomendação do W3C, ou seja, uma especificação estável (desenvolvida por um grupo de trabalho e revisada por membros do consórcio) considerada como um padrão *Web* tanto pela indústria quanto pela academia. Podemos elencar como componentes básicos desta linguagem de marcação de ontologia classes, propriedades e indivíduos.

Hoje devido principalmente ao grande volume de informação disponível na *Web*, a compreensão compartilhada de um domínio se faz necessária para superar as diferenças de terminologia, evitando problemas, como por exemplo, o uso de um mesmo termo com significados diferentes [ANT08].

O crescente interesse pela *Web Semântica* fez aumentar o número de ontologias de domínios disponíveis [SPE07]. No entanto, sua construção requer a presença de especialistas do domínio a ser modelado, ou seja, implica esforço humano altamente qualificado. Depende de um engenheiro de conhecimento que investigue um domínio em específico, aprenda quais conceitos são importantes nesse domínio e crie uma representação formal dos objetos e relações [RUS04]. Por este motivo, possibilitar o reuso de ontologias, evitando gastos desnecessários, se torna um ramo recente e interessante a ser estudado.

A importância do reuso de ontologia é comparável às questões de reuso de software. Segundo [SPE07] este pode ocorrer em diversos níveis (desde código até conhecimento), maiores benefícios podem ser atingidos ao se reutilizar artefatos de mais alto nível de abstração, isto é, ao se reusar conhecimento. No caso das ontologias este conhecimento é representado através de uma linguagem de marcação de ontologia. [NOY01] diz que uma das razões para se construir ontologias é possibilitar posteriormente o reuso de conhecimento do domínio. A reutilização para [PIN00] pode ser observada sob dois pontos de vista: (1) criar, estender, especializar e adaptar uma ontologia a partir dos conceitos de outras, preservando as ontologias originais e (2) unir ontologias diferentes sobre domínios iguais ou similares em uma única.

Visto que o reuso de qualquer entidade está vinculado a sua qualidade e que métricas quantificam atributos (propriedades ou características) de uma determinada entidade (produto, processo ou recurso) [IEE90], elaboramos uma revisão sistemática sobre métricas para ontologias com a finalidade de conhecer o estado da arte a respeito deste assunto. Um próximo passo foi implementar um subconjunto de métricas processáveis computacionalmente e cujos resultados fossem numéricos.

Para avaliar os resultados da aplicação das métricas em ontologias, elaboramos e aplicamos um questionário a potenciais usuários, pesquisadores e engenheiros de ontologias. O objetivo foi verificar a utilidade de um subconjunto de métricas na inspeção de ontologias presentes em bibliotecas especializadas, no contexto deste trabalho o portal OntoLP ¹, já que a busca por ontologias é uma das atividades vinculadas ao processo de reuso.

A ênfase do portal OntoLP e do projeto que o contempla é em disponibilizar recursos (ontologias, ferramentas e outros) para a língua portuguesa. Portanto, pretendemos também através deste trabalho contribuir de forma que a comunidade obtenha maiores informações a respeito do que está sendo desenvolvido em nosso idioma.

¹<http://www.inf.pucrs.br/~ontolp>

1.2 Objetivo da Dissertação

1.2.1 Objetivo Geral

O objetivo geral deste trabalho foi elaborar uma revisão sistemática sobre métricas para ontologias para compreender as métricas propostas na literatura e então extrair um subconjunto de métricas das ontologias disponíveis no portal OntoLP.

1.2.2 Atividades Realizadas para Atingir o Objetivo Geral

Como atividades realizadas para atingir o objetivo geral, podemos citar:

- Estudar as áreas de *Web Semântica*, ontologias e métricas.
- Realizar uma revisão sistemática sobre métricas para ontologias, a fim de conhecer o que está sendo feito atualmente a respeito deste assunto.
- Buscar por ontologias de diferentes domínios em língua portuguesa para alimentar o repositório do portal OntoLP.
- A partir do resultado obtido na revisão sistemática extrair as métricas e, em seguida, analisar, comparar e identificar suas diferentes abordagens.
- Definir o subconjunto de métricas a ser implementado e sua classificação em grupos. Os critérios de escolha foram a frequência no conjunto de artigos, aplicabilidade nas ontologias disponíveis, retornos numéricos e abrangentes.
- Elaborar e aplicar um questionário para avaliar os resultados obtidos, a ser respondido por um grupo de usuários, pesquisadores e engenheiros de ontologias.

1.3 Organização da Dissertação

O texto encontra-se organizado da seguinte forma. No Capítulo 2 apresentamos o embasamento teórico a respeito de *Web Semântica*, ontologias e métricas. No Capítulo 3 abordamos a metodologia de estudo denominada revisão sistemática (desde o desenvolvimento do protocolo até a análise dos resultados) sobre métricas para ontologias. No Capítulo 4 descrevemos o protótipo, o conjunto de métricas e de ontologias. No Capítulo 5 detalhamos o questionário elaborado e aplicado, utilizado como método de avaliação dos resultados obtidos neste trabalho. E finalmente, no Capítulo 6, expomos as considerações finais e os trabalhos futuros. Seguindo das referências bibliográficas.

2. Fundamentação Teórica

Neste capítulo apresentamos as áreas de Web Semântica, Ontologias e Métricas. E suas respectivas peculiaridades tendo em vista o contexto deste trabalho.

2.1 Web Semântica

A *Web* mudou a maneira das pessoas se comunicarem e negociarem. Seu uso típico envolve procurar e fazer uso de informação, pesquisar e entrar em contato com outras pessoas, ver catálogos de lojas *on-line* e encomendar produtos, preencher formulários e visualizar materiais [ANT08].

Tendo em vista este cenário, consideramos as palavras-chave utilizadas em motores de busca, como Yahoo ¹ e Google ², as principais ferramentas na *Web* atual. No entanto, existem algumas limitações destes motores de busca, tais como: um grande número de página como resultado da busca, porém apresentando um baixo grau de precisão; dependendo da especificidade da consulta o retorno pode ser baixo ou inexistente; outro problema é que as ferramentas atuais são altamente sensíveis ao vocabulário, ou seja, palavras-chave com mesmo significado semântico, que deveriam obter resultados iguais, retornam resultados diferentes; atualmente se precisamos de informações que estão distribuídas por vários documentos, devemos iniciar várias buscas, extrair manualmente as informações parciais e agrupar estas informações [ANT08].

Sabemos que o volume de informações na *Web* é crescente, em vista disso pesquisadores da indústria e da academia vêm explorando a possibilidade de criar uma *Web Semântica*. A proposta da *Web Semântica* surgiu com Tim Berners-Lee, o inventor da WWW no final dos anos 80. O termo foi utilizado pela primeira vez no livro "*Weaving the Web*" (Tecendo a Teia). Sua visão dessa nova *Web* dá importância ao significado da informação. Central a esta ideia está a utilização de ontologias, que fornecem uma língua franca permitindo que máquinas processem e integrem recursos *Web* de maneira inteligente, possibilitando buscas mais rápidas e aperfeiçoadas, facilitando a comunicação entre dispositivos heterogêneos acessíveis via *Web* [BER01]. Com esta nova *Web* se pretende solucionar as limitações expostas.

Abaixo descrevemos o modelo de camadas idealizado por Tim Berners-Lee (Figura 2.1), o qual apresenta uma visão geral da arquitetura da *Web Semântica* [ANT08].

A camada Unicode + URI fornece a interoperabilidade em relação à codificação de caracteres, ao endereçamento e à nomeação de recursos da *Web Semântica*. O Unicode é um padrão de codificação que possibilita uma representação numérica universal (independente da plataforma de software e do idioma) para cada caractere. O URI é um padrão de identificação de um recurso físico ou abstrato

¹<http://m.br.yahoo.com//>

²<http://www.google.com.br/>

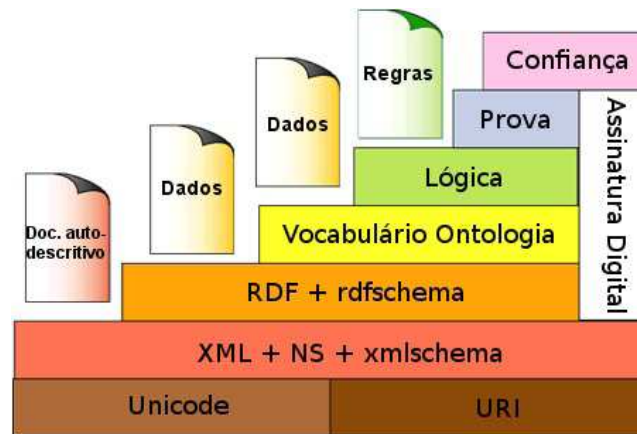


Figura 2.1 – Visão da Web Semântica. Fonte: adaptado [ANT08]

de maneira única e global.

A camada XML + NS + xmlschema fornece a interoperabilidade em relação à sintaxe de descrição de recursos da *Web Semântica*. O XML é uma linguagem que fornece uma sintaxe básica para estruturação de documentos. Os NS são úteis para distinguir elementos e atributos de documentos XML que possuem o mesmo nome. O xmlschema é uma linguagem que faz o controle sintático a linguagem XML, definição de tipo e dados, restrições à estrutura.

A camada RDF + rdfschema fornece a interoperabilidade em relação à semântica de descrição de recursos da *Web Semântica*. O RDF é uma linguagem que fornece um modelo de dados expressos através de declarações simples, ou seja, de triplas (recurso-propriedade-valor). O rdfschema é uma linguagem que permite a definição de vocabulários, representação de ontologias simples.

A camada Vocabulário Ontologia dá suporte para a evolução dos vocabulários e para o processamento e integração das informações existentes evitando problemas como falta de definição ou conflito de terminologia. A OWL, detalhada na próxima seção, é uma linguagem que estende o vocabulário da rdfschema apresentando maior poder de expressividade e de inferência.

A camada Lógica é utilizada para melhorar ainda mais a linguagem de marcação de ontologia e permitir a escrita de aplicações específicas. Fornece suporte para a descrição de regras para expressar relações entre as classes de uma ontologia, as quais não podem ser expressas com a linguagem de marcação de ontologia.

A camada de Prova envolve a execução das regras da camada Lógica, funcionando como um mecanismo de inferência, verifica a consistência das informações e das regras.

A camada Confiança surge com o uso de assinaturas digitais e outros tipos de conhecimento, garantindo que as informações estejam representadas de modo correto, possibilitando certo grau de confiabilidade.

A arquitetura da *Web Semântica* está atualmente sendo debatida pela comunidade e pode estar sujeita a refinamentos e a modificações no futuro.

2.2 Ontologias

Na literatura [GUA98] diz que diferentes categorias de ontologia podem ser desenvolvidas de acordo com seu nível de generalidade. O autor identifica quatro categorias (Figura 2.2):

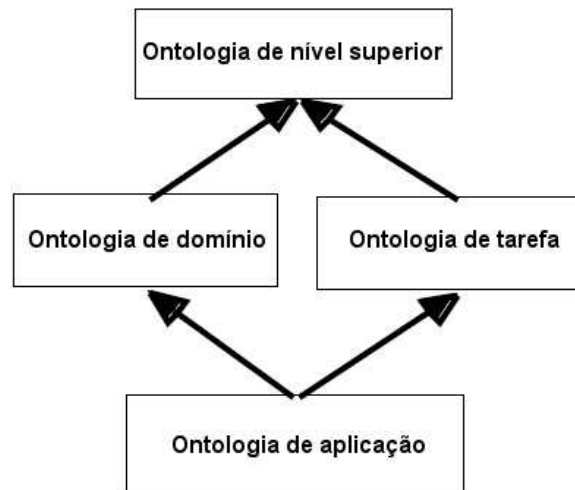


Figura 2.2 – Categorização de ontologias segundo Guarino. Fonte: adaptado [GUA98]

- Ontologias de nível superior - descrevem conceitos muito genéricos, tais como espaço, tempo, matéria, objeto, evento, ação e outros. Estes seriam, a princípio, independentes de domínio e poderiam ser reutilizados na construção de novas ontologias.
- Ontologias de domínio e de tarefa - descrevem vocabulários relativos a um domínio, uma tarefa ou uma atividade genéricos através da especialização de conceitos presentes na ontologia de nível superior.
- Ontologias de aplicação - descrevem conceitos em função tanto de um domínio quanto de uma tarefa em particular, que muitas vezes são especializações de ambas as ontologias relacionadas. Conceitos em ontologias de aplicação correspondem, de maneira geral, aos papéis desempenhados por entidades do domínio no desenrolar de alguma tarefa.

Já [LAS01] classifica as ontologias de acordo com as informações que elas precisam expressar e a riqueza de suas estruturas. As seguintes categorias podem ser vistas no espectro (Figura 2.3): vocabulário controlado, glossário, tesouro, hierarquia informal *is-a*, hierarquia formal *is-a*, hierarquia formal com instâncias, frames, restrição de valor, disjunta, inversa, *part-of* e restrição lógica.

Neste trabalho não entraremos no mérito de classificar as ontologias encontradas e inseridas no portal OntoLP. Mas adotaremos a estrutura proposta por Maedche [MAE02], onde uma ontologia pode ser descrita através de uma sextupla composta por elementos primitivos, $O := \{C, P, A, H^C, prop, att\}$ que consiste de: dois conjuntos disjuntos, C (conceitos) e P (relações). Um tipo específico de relações são atributos A . Uma hierarquia de conceitos, $H^C : H^C$ que é uma



Figura 2.3 – Categorização de ontologias segundo Lassila. Fonte: adaptado [LAS01]

relação direta transitiva $H^C \subseteq C \times C$ também chamada de taxonomia de conceitos. $H^C(C_i, C_j)$ significa C_i é um sub-conceito de C_j . A função $prop : P \rightarrow C \times C$ relaciona os conceitos de modo não taxonômico. A função $att : A \rightarrow C$ relaciona os conceitos com valores literais (strings).

Utilizamos esta estrutura porque ela pode mapear a maioria das linguagens de marcação de ontologias conhecidas e padronizar as métricas obtidas na revisão sistemática proposta. Na próxima subseção, descrevemos sucintamente algumas destas linguagens e enfatizamos uma em especial (OWL).

2.2.1 Linguagens para Ontologias

As linguagens de marcação para ontologias apresentadas a seguir (SHOE, OIL, DAML + OIL e OWL) correspondem as bases da *Web Semântica*.

SHOE

A linguagem SHOE foi desenvolvida na Universidade de Maryland em 1996, criada como uma extensão de HTML para anotar o conteúdo de páginas da *Web* com o objetivo de incorporar conhecimento semântico compreensível por máquina em documentos da *Web*. SHOE possibilita a definição de conceitos, relacionamentos e atributos. Contudo é menos expressiva que RDF (não permite negação) e apresenta grandes dificuldades na manutenção das páginas anotadas. O projeto SHOE³ foi descontinuado e os pesquisadores envolvidos migraram para as linguagens DAML + OIL e OWL [GOM04].

OIL

No âmbito do Projeto Europeu On-to-Knowledge foi construída a linguagem OIL. Assim como a outra linguagem anteriormente apresentada, OIL tinha como propósito expressar a semântica de recursos *Web*. OIL pode ser definida como uma linguagem baseada em frames que usa lógica

³<http://www.cs.umd.edu/projects/plus/SHOE/>

de descrição, dando suporte a mecanismos de inferência. Ontologias escritas em OIL podem ser traduzidas para RDF. A comunidade de pesquisadores que vem utilizando esta linguagem disponibiliza algumas ferramentas para a edição de ontologias, são elas: OilEd, Protégé-2000, OntoEdit e WebODE [GOM04].

DAML + OIL

Na mesma época em que estava sendo desenvolvida a OIL na Europa, a agência americana DARPA em conjunto com alguns membros do consórcio W3C desenvolvia a linguagem DAML ⁴.

A DAML + OIL foi criada da combinação das duas linguagens, que apesar de diferentes, apresentavam características similares. Esta linguagem é dividida em duas partes, domínio das classes dos objetos e domínio dos tipos de dados. Esta separação facilita a realização de inferências. DAML é composta por elementos de classe, expressões de classe e propriedades.

OWL

Em 2004 a W3C lançou a OWL, como uma revisão da DAML + OIL, projetada para atender as necessidades das aplicações da *Web Semântica* fornecendo informações sobre classes, relacionamentos e instâncias que fazem parte do domínio em questão. OWL possui três sublinguagens em função da expressividade: OWL Lite, OWL DL e OWL Full [SMI04]. e vários tipos de propriedades

- OWL Lite - permite menos expressividade. Suporta a criação de hierarquias e restrições mais simples (restrição existencial, de cardinalidade, de igualdade e vários tipos de propriedades);
- OWL DL - em comparação com a OWL Lite a OWL DL adiciona suporte total para negação, disjunção, restrição de cardinalidade, enumerações e restrição de valor, impõe restrições quanto a utilização de vocabulários e declarações;
- OWL Full - ao contrário da OWL Lite e da OWL DL esta sublinguagem não impõe restrições quanto a utilização de vocabulários e declarações.

Os elementos que constituem uma ontologia no formato OWL são basicamente: classes (ou conceitos), relações (ou propriedades) e instâncias (ou indivíduos) [SMI04]. A seguir descrevemos por meio de exemplos os componentes essenciais destes elementos.

1. Classes

Os conceitos básicos de um domínio devem corresponder a classes que são as raízes de várias árvores taxonômicas. Assim, cada classe definida pelo usuário é implicitamente uma sub-classe de *owl : Thing*. Classes raiz de um domínio específico são definidas simplesmente ao nomearmos a classe. OWL também define a classe vazia, através da sintaxe *owl : Nothing*.

⁴<http://www.daml.org/>

Para a ontologia do domínio de arte contemporânea [TRI05] (Fragmento 2.1) nota-se somente a existência das classes raiz *Cultura* e *Obra_de_Arte*, cujos nomes são dados através da sintaxe *rdf : ID*. Formalmente, não sabemos quase nada sobre essas classes além de sua existência, apesar do uso de termos familiares como etiquetas. Poderiam até o momento muito bem terem sido chamadas de *Thing1* e *Thing2*.

Fragmento 2.1: Exemplo de classes raiz.

```
< owl : Class rdf : ID = "Cultura" / > < owl : Class rdf : ID = "Obra_de_Arte" / >
```

A sintaxe *rdf : ID = "Obra_de_Arte"* é usada para introduzir um nome, como parte de sua definição. Para referenciarmos uma classe utilizamos o caractere #. Por exemplo, *rdf : about = #Obra_Visual* e *rdf : resource = #Obra_de_Arte*, apresentado no Fragmento 2.2.

O construtor taxonômico fundamental para classes é *rdfs : subclassOf* que relaciona uma classe mais específica com uma classe mais geral. Neste caso, *Obra_Visual* é uma sub-classe de *Obra_de_Arte*, onde cada instância de *Obra_Visual* é também uma instância de *Obra_de_Arte*. Outra característica é que a relação *rdfs : subclassOf* é transitiva.

Através da sintaxe *rdfs : comment*, podemos visualizar a definição de *Obra_Visual* como sendo do tipo *string* e com valor *plástica*. As marcações *rdfs : comment* e *rdfs : label* (suportam múltiplos idiomas) são comentários opcionais, que têm como objetivo fornecer um nome legível às classes e nada contribuem para a interpretação lógica de uma ontologia.

Fragmento 2.2: Exemplo de referências.

```
< owl : Class rdf : about = "#Obra_Visual" >
< rdfs : comment rdf : datatype = "http : //www.w3.org/2001/XMLSchema#string" >
plástica < /rdfs : comment >
< rdfs : subclassOf rdf : resource = "#Obra_de_Arte" / > < /owl : Class >
```

2. Relações

Em OWL existem basicamente dois tipos de relações: propriedades do objeto que representa as relações entre instâncias de duas classes (*ObjectProperty*) e propriedades de tipo de dados que representa as relações entre as instâncias das classes e literais (*DatatypeProperty*).

Quando definimos uma propriedade há uma série de maneiras de restringir a relação. O domínio e o valor podem ser especificados. A propriedade pode ser definida como uma especialização de uma propriedade existente (*subPropertyOf*), assim como classes, podem ser organizadas em hierarquias.

No Fragmento 2.3 a propriedade *organiza* tem um domínio *Instituicao* e um valor *Evento*. Ou seja, ele relaciona as instâncias de *Evento* com as instâncias de *Instituicao*. O domínio da propriedade, quando este apresenta vários, é a intersecção das classes identificadas (e da mesma forma para valor).

Fragmento 2.3: Exemplo de propriedades (relações).

```
< owl : ObjectProperty rdf : ID = "organiza" >
< rdfs : range rdf : resource = "#Evento" / >
< rdfs : domain rdf : resource = "#Instituicao" / > < /owl : ObjectProperty >
```

A propriedade *é_pública* é definida como *DatatypeProperty*, Fragmento 2.4, quando refere-se ao relacionamento entre instâncias de classes e tipos de dados (*string*, *boolean* e *integer*). Neste exemplo a relação entre *Instituicao* e *boolean*.

Algumas das características das propriedades possíveis de se especificar em OWL são: Simétrica (*owl : SymmetricProperty*), Inversa (*owl : inverseOf*), Transitiva (*owl : TransitiveProperty*), Funcional (*owl : FunctionalProperty*) e Funcional Inversa (*owl : InverseFunctionalProperty*).

Restrições também podem ser impostas sobre as propriedades. Uma restrição é um tipo especial de descrição das classes. As restrições podem ser do tipo valor (*allValuesFrom*, *someValuesFrom* e *hasValue*) ou do tipo cardinalidade (*Cardinality*, *maxCardinality* e *minCardinality*).

Fragmento 2.4: Exemplo de propriedades (atributos).

```
< owl : DatatypeProperty rdf : ID = "é_pública" >
< rdfs : domain rdf : resource = "#Instituicao" / >
< rdfs : range rdf : resource = "http : //www.w3.org/2001/XMLSchema#boolean" / >
< /owl : DatatypeProperty >
```

3. Instâncias

Além disso, outro elemento básico das ontologias são as instâncias. Estas são introduzidas ao declararmos como sendo membros de uma classe. Toda instância em OWL é membro da classe *owl : Thing*. No Fragmento 2.5 as instâncias *carnaval* e *Barroco* pertencem respectivamente às classes *Cultura_Popular* e *Época*.

Fragmento 2.5: Exemplo de instâncias.

```
< Cultura_Popular rdf : ID = "carnaval" / > < Época rdf : ID = "Barroco" / >
```

Tendo em vistas as linguagens de marcação de ontologias explanadas nesta subseção, uma das primeiras etapas deste trabalho foi à procura por ontologias em língua portuguesa. A princípio não havíamos estabelecido limites de formato, já que o repositório OntoLP não possuía recursos. Porém, a implementação das métricas considera ontologias em OWL.

2.2.2 Levantamento de Ontologias em Português

Ontologias em língua portuguesa são recursos escassos. A estratégia utilizada para encontrá-las foi fixar a linguagem e o idioma (ontologias no formato OWL descritas em português). Foram

realizadas buscas no Google ⁵ especificando o tipo de consulta, por exemplo “*domínio filetype:owl*”, no Swoogle ⁶, motor de busca de ontologias, documentos, termos e dados publicados na *Web* e em páginas de projetos e/ou grupos de pesquisa que trabalham com o assunto, através do contato direto via *e-mail*.

Como resultado dessa pesquisa e consulta à comunidade, encontramos ontologias de diferentes domínios em OWL e em português, tais como: ecologia, nanotecnologia, arte, currículo lattes, emoção, privacidade de dados, rede de conhecimento científico, smartphone, música e equivalência de estímulos. Hoje contamos com aproximadamente 25 ontologias; nos idiomas português, inglês e português/inglês (multi-idioma); nos formatos Frames, RDF e OWL.

2.3 Métricas

Segundo Sure e Vrandecic [SUR07] métricas são necessárias para se avaliar ontologias durante as fases de construção e de aplicação, possibilitando uma compreensão rápida e simples a respeito do que está sendo modelado por estas estruturas, facilitando o controle de sua futura evolução. Nos últimos anos, um número significativo de iniciativas sobre métricas para ontologias têm sido sugeridas [LOZ04, CRO05, YAO05, YAN06, GAN06, TAR7].

Devido ao crescente desenvolvimento de aplicações baseadas em ontologias e a falta de métodos que ajudem a procurar ontologias e a justificar as escolhas surgiu o OntoMetric [LOZ04]. Este *framework* de métricas fornece uma longa lista de métricas em cinco dimensões, como ferramentas, linguagens, conteúdos, metodologias e custos, do domínio da engenharia de ontologias.

No trabalho Cross e Pal [CRO05] um conjunto de métricas de qualidade foi desenvolvido e implementado em um *plugin* para o editor de ontologias Protégé. Nele qualquer ontologia especificada em linguagens como RDFS ou OWL podem ser analisadas qualitativamente.

Uma proposta de modo a avaliar ontologias do ponto de vista da sua complexidade e evolução pode ser encontrada em [YAN06].

Semelhante às métricas de *software*, espera-se através das métricas de ontologia dar algumas dicas para desenvolvedores de ontologias. Por exemplo, ajudá-los a projetar ontologias, melhorar a qualidade das já existentes, prevenir e reduzir a necessidade de uma posterior manutenção [YAO05].

Em [GAN06] três tipos principais de medidas para avaliação das ontologias são identificadas: estrutural, que são ontologias tipicamente representadas como grafos, funcional, que relaciona o uso da ontologia e de seus componentes, e usabilidade que depende do nível de anotação da ontologia considerada.

Já a ferramenta OntoQA [TAR7] avalia ontologias relacionadas a um determinado conjunto de termos e as classifica de acordo com um conjunto de métricas as quais capturam diferentes aspectos

⁵<http://www.google.com.br/>

⁶<http://swoogle.umbc.edu/>

destas estruturas.

Na literatura, trabalhos sumarizando iniciativas sobre métricas para ontologias não foram encontrados. Por esse motivo, neste trabalho elaboramos uma revisão sistemática a respeito de métricas para ontologias.

As métricas candidatas (obtidas do conjunto de artigos resultantes da revisão sistemática) e as métricas efetivamente implementadas (padronizadas e descritas na fase de implementação do protótipo) serão apresentadas respectivamente nos Capítulos 3 e 4.

3. Revisão Sistemática

Nesse capítulo abordamos o assunto revisão sistemática, prática comum na Ciência da Saúde. Todavia, esta metodologia é considerada recente na Ciência da Computação, o primeiro trabalho que explana a condução de uma revisão sistemática em Engenharia de *Software* é datado de 2004 [KIT04].

Ao contrário da revisão informal da literatura, na revisão sistemática o pesquisador segue um protocolo de revisão pré-estabelecido. Este é bem definido e rigoroso de forma que outros profissionais possam reproduzi-lo. Enfim, uma revisão sistemática visa estabelecer um processo formal para conduzir este método de investigação, evitando a introdução de eventuais tendenciosidades, erros ou desvios sistemáticos do estudo, que se distanciam da verdade [BIO05].

Segundo [KIT04] uma revisão sistemática pode ser conduzida em três etapas principais, Figura 3.1: planejamento da revisão (1), condução da revisão (2) e publicação dos resultados (3). Na fase de planejamento é definido o protocolo da revisão, que descreve, entre outras coisas, o propósito da revisão e os procedimentos que serão adotados. Durante a condução da revisão é feita a busca por trabalhos relevantes para o objeto de estudo, a seleção das publicações e a extração dos dados de cada uma das publicações selecionadas. Na fase de publicação dos resultados, as considerações são descritas em relatórios ou artigos.

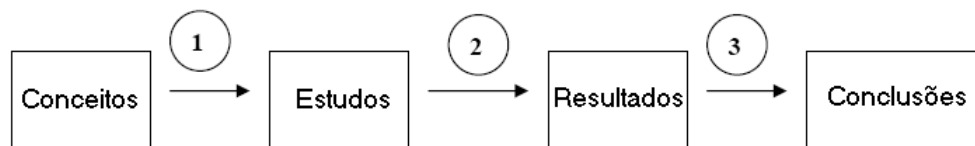


Figura 3.1 – Etapas da revisão sistemática. Fonte: adaptado [BIO05]

3.1 Planejamento da Revisão

Nesta fase definimos o protocolo, o qual especifica os métodos que serão utilizados para realizar uma determinada revisão sistemática [KIT04]. Neste trabalho tomamos como base o modelo proposto por [BIO05], o qual encontra-se disponível de maneira adaptada no Apêndice A.

O motivo que nos levou a elaborar esta revisão sistemática foi a necessidade de se evidenciar o estado da arte no tópico métricas para ontologias.

O propósito da revisão sistemática dessa dissertação consiste em realizar um levantamento de trabalhos sobre métricas de reuso, avaliação e *ranking* para ontologias a fim de sumarizar as publicações encontradas e, definir procedimentos como formulação da questão e seleção das fontes.

3.1.1 Formulação da Questão

A definição da questão (ou questões) de pesquisa é o elemento mais crítico de uma revisão sistemática. É utilizada basicamente para construir as *strings* de busca e para determinar os dados que serão extraídos a partir de cada estudo preliminar [BIO05].

A formulação da questão é composta pelos seguintes itens: questão foco, qualidade e amplitude da questão.

Questão Foco

O objetivo geral desta revisão é resumir os principais aspectos dos trabalhos sobre métricas para ontologias. Tendo em vista os trabalhos selecionados pretendemos identificar os seguintes itens:

- Identificar os propósitos dos trabalhos selecionados.
- Identificar se ferramentas foram desenvolvidas ou não.
- Identificar as ontologias utilizadas (linguagem e domínio).
- Identificar a categorização, o cálculo e a descrição das métricas.

Qualidade e Amplitude da Questão

A qualidade e amplitude da questão de pesquisa permitem descrever sua sintaxe (problema, questão, palavras-chave e sinônimos) e sua semântica (intervenção, controle, efeito, medição dos resultados, população, aplicação e projeto experimental).

- Problema

Em razão do surgimento do projeto *Web Semântica* o desenvolvimento de aplicações baseadas em ontologias tem crescido. Em vista disso, torna-se visível a necessidade do uso de métricas que auxiliem na busca por estas estruturas. Nesta revisão pretendemos identificar trabalhos sobre métricas inserido no contexto de reuso, *ranking* e avaliação de ontologias.

- Questão

Com base na questão foco e no problema, identificamos a seguinte questão de pesquisa: **Quais são as características encontradas nos trabalhos sobre métricas para ontologia ?**

- Palavras-chave e sinônimos

As palavras-chave utilizadas foram: *metrics*, *ontology*, *reuse*, *ranking* e *evaluate*. É importante ressaltar que a busca não foi restrita à subáreas da Ciência da Computação.

- Intervenção

A intervenção corresponde a identificação das métricas expostas nos trabalhos selecionados.

- Controle

O controle é inexistente, já que não existe um conjunto de dados de referência antes da realização desta revisão.

- Efeito

Os efeitos esperados ao final da revisão sistemática compreendem a ampliação do conhecimento, a identificação do estado da arte sobre métricas para ontologias e a identificação de um subconjunto relevante, implementável de métricas.

- Medição dos resultados

A medição dos resultados é realizada através da contagem do número de iniciativas encontradas.

- População

Observamos artigos publicados em periódicos e anais de conferências relacionados ao objeto de estudo. As restrições referentes a data inicial e final das publicações corresponde ao intervalo de janeiro de 2004 até fevereiro de 2009.

- Aplicação

Algumas métricas contidas nos resultados obtidos (artigos selecionados) serão implementadas e inseridas no portal de ontologias OntoLP. Além disso, um artigo será escrito para que outros profissionais e pesquisadores possam se beneficiar dos resultados.

- Projeto experimental

Não foram utilizados métodos de análise estatística para a avaliação dos resultados deste trabalho. Logo, a definição do projeto experimental não se aplica.

3.1.2 Seleção das Fontes

De acordo com o modelo adotado, a seleção das fontes inclui os seguintes itens: definição dos critérios de seleção, idioma dos estudos, identificação e seleção das fontes após avaliação.

Definição dos Critérios de Seleção

Os critérios definidos para selecionar as fontes foram: disponibilidade para consultar os artigos na *Web*, presença de mecanismos de busca com suporte a palavras-chave e ao operador *and*, base de dados atualizada, veículo de publicação com certo grau de confiabilidade reconhecido pelo Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Caso não atenda algum dos critérios citados a fonte candidata não é selecionada.

Idioma dos Estudos

Os estudos deverão estar escritos em língua inglesa, já que conferências e periódicos internacionais fazem uso deste idioma.

Identificação e Seleção das Fontes após Avaliação

Inicialmente foram identificadas as seguintes bibliotecas digitais: IEEEExplore, ACM Digital Library, CiteSeerX Library, SpringerLink, ScienceDirect e GoogleScholar.

- Métodos de pesquisa nas fontes selecionadas

A Tabela 3.1 apresenta as fontes selecionadas e suas respectivas URLs. Todas as fontes identificadas foram selecionadas. As particularidades de cada mecanismo de busca são descritas abaixo.

Tabela 3.1 – Fontes selecionadas.

Fonte	URL
IEEE	http://ieeexplore.ieee.org
ACM	http://portal.acm.org
CiteSeerX	http://citeseerx.ist.psu.edu
SpringerLink	http://www.springerlink.com
ScienceDirect	http://www.sciencedirect.com
Google Scholar	http://scholar.google.com.br

O IEEEExplore possui quatro tipos de busca *Basic*, *Advanced*, *Author* e *CrossRef*. Utilizamos o módulo de busca *Advanced*, onde foi possível inserir o operador *and* entre palavras-chave e limitar as buscas por ano de publicação através da aba *Select date range*.

A biblioteca digital da ACM permite que a busca seja feita em *The ACM Digital Library* ou em *The Guide*. Optamos pela *The ACM Digital Library* por esta possuir apenas trabalhos publicados na

base de dados da ACM. Nesta biblioteca também utilizamos o módulo *Advanced Search*, indicamos a faixa de tempo e inserimos operadores *booleanos* entre os termos utilizados na pesquisa.

Na biblioteca CiteseerX apenas os campos *Text* e *Range Criteria* da busca *Advanced* foram utilizados, e correspondem respectivamente às *strings* de busca e ao intervalo em anos das publicações.

A SpringerLink através do link *more options* possibilitou, assim como nas outras fontes, realizar uma busca por publicações em uma faixa de tempo e inserir operadores lógicos entre as expressões, como *and* por exemplo.

No ScienceDirect utilizamos a busca *Advanced*, na qual o campo *Term(s)* possibilitou a inserção de termos e o campo *Date(s)* data ou intervalo de datas no qual as publicações estão inseridas.

Também realizamos as buscas através do módulo *Advanced* no GoogleScholar, este possui funcionalidade semelhante às das outras fontes já mencionadas.

- *Strings* de busca

A partir das palavras-chave *metrics*, *ontology*, *reuse*, *ranking* e *evaluate* elaboramos 4 tipos de buscas, uma mais geral com os termos *metrics* e *ontology*, e três mais específicas, relacionadas a métricas com a finalidade de reusar (*reuse*), classificar (*ranking*) e avaliar (*evaluate*) as ontologias. Na Tabela 3.2 apresentamos o número de artigos retornados para cada *string* de busca em cada fonte selecionada (no período de janeiro de 2004 até fevereiro de 2009).

Tabela 3.2 – Retorno para cada *string* de busca em cada fonte selecionada.

String de busca	IEEE	ACM	CiteSeerX	Springer	ScienceDirect	GoogleScholar	Total
metrics AND ontology	64	838	1021	2075	1127	15200	20325
metrics AND ontology AND reuse	5	215	234	508	207	5150	6319
metrics AND ontology AND ranking	5	275	394	556	257	6040	7527
metrics AND ontology AND evaluate	14	460	824	1219	551	12700	15768

3.2 Condução da Revisão

3.2.1 Seleção dos Estudos

Uma vez que as fontes são definidas é preciso descrever o processo e os critérios de seleção dos estudos. A seleção dos estudos pode ser dividida em duas partes: definição dos estudos (definição dos critérios de inclusão e exclusão dos estudos, definição dos tipos de estudos e identificação dos procedimentos adotados para a seleção dos estudos) e execução da seleção (seleção inicial dos estudos, avaliação da qualidade dos estudos e revisão da seleção).

Definição dos Estudos

- Critérios de inclusão e exclusão dos estudos

São os critérios pelos quais os estudos serão avaliados, ou seja, se devem ser selecionados ou não em função do contexto da revisão sistemática. Foram desconsiderados os estudos que não estavam no formato de artigo completo, que não estavam relacionados a métricas para ontologias e cujo conteúdo não estava relacionado a métricas para possibilitar o reuso, ranking e avaliação de ontologias.

- Tipos de estudos

Os tipos de estudos preliminares que vão ser selecionados durante a revisão sistemática devem contemplar a aplicação de métricas propostas em ontologias de domínio.

- Procedimentos adotados para a seleção dos estudos

Como dito anteriormente após execução das buscas obtivemos o retorno apresentado na Tabela 3.2. Para reduzir o número de artigos a serem avaliados optou-se por selecionar como amostra inicial as 20 primeiras ocorrências mais relevantes de cada fonte e para cada *string* de busca. A amostra inicial (removendo as repetições) e as etapas (filtros) pelas quais estes artigos passaram bem como seus retornos podem ser visualizados na Tabela 3.3.

Tabela 3.3 – Retorno para cada fonte selecionada em cada etapa.

Fonte	Amostra Inicial	Primeira etapa	Segunda etapa	Terceira etapa
IEEE	29	9	4	3
ACM	55	4	3	0
CiteSeerX	39	18	10	1
Springer	60	10	5	0
ScienceDirect	49	4	1	1
Google Scholar	50	22	15	4
Total(sem repetições)	246	47	25	6

1. Primeira etapa: leitura do título e das palavras-chave.
2. Segunda etapa: leitura do resumo (*abstract*), introdução e conclusão.
3. Terceira etapa: leitura integral dos artigos restantes.

Execução da Seleção

Na execução da seleção o objetivo é registrar o processo de seleção preliminar de estudos, elaborar relatórios sobre os estudos obtidos e os resultados da sua avaliação. Esta pode ser dividida em três partes: seleção inicial dos estudos, avaliação da qualidade dos estudos e revisão da seleção. Dos 246 artigos selecionados 199 foram excluídos após leitura do título e das palavras-chave. Foram lidos de forma parcial (resumo, introdução, conclusão) 47 artigos e de forma integral 25 artigos. Apenas 6 deles foram selecionados para extração de informação (Tabela 3.4).

Tabela 3.4 – Apresentação dos artigos selecionados.

Fonte	Título
Google Scholar	Ontometric: A Method to Choose the Appropriate Ontology [GOM04]
Google Scholar, IEEE	Metrics for Ontologies [CRO05]
Google Scholar	Cohesion Metrics for Ontology Design and Application [YAO05]
IEEE	Evaluation Metrics for Ontology Complexity and Evolution Analysis [YAN06]
Google Scholar, Springer, CiteSeerX	Modelling Ontology Evaluation and Validation [GAN06]
IEEE	Ontology Evaluation and Ranking using OntoQA [TAR7]

- Seleção inicial dos estudos

A pesquisa em si foi executada e todos os estudos obtidos foram listados para posterior avaliação.

- Avaliação da qualidade dos estudos

A avaliação da qualidade dos estudos foi realizada a partir da verificação dos critérios de inclusão e exclusão dos estudos e dos tipos de estudos que os artigos correspondem.

- Revisão da seleção

A seleção de estudos devem ser revisadas para garantir que a qualidade dos estudos e a não eliminação dos artigos relevantes. Uma prática recomendável seria a utilização de revisores independentes, mas neste trabalho não contamos com revisores independentes.

3.2.2 Extração de Informação

Após os estudos serem selecionados começa o processo de extração de informações relevantes. Nele são descritos os critérios de inclusão e exclusão de informações, o formulário de extração de dados e a execução das extrações.

Critérios de Inclusão e Exclusão de Informações

As informações extraídas dos estudos deveriam abranger as descrições e/ou fórmulas de métricas para ontologias. Considerando o conjunto de ontologias em língua portuguesa no formato OWL utilizado neste trabalho, estas métricas devem ser condizentes para possibilitar em um momento posterior suas implementações. Informações como por exemplo propósito dos trabalhos selecionados e o desenvolvimento ou não de ferramentas (framework, plugin) também foram incluídas.

Formulário de Extração dos Dados

O formulário de extração de dados corresponde a diferentes aspectos divididos em relação ao trabalho, as ontologias e as métricas. Quanto ao trabalho podemos elencar os propósitos dos trabalhos selecionados e o desenvolvimento ou não de ferramentas. Já ao tratarmos das ontologias buscamos por linguagem de descrição de ontologia e por domínio. E finalmente quanto às métricas exploramos suas categorizações, quantidades, presença de cálculos (fórmulas matemáticas) e descrições textuais.

Execução das Extrações

A execução das extrações se deu através de uma nova leitura dos artigos procurando preencher o formulário de extração de dados.

3.3 Publicação dos Resultados

Apresentação dos Resultados em Tabelas

Os resultados obtidos na execução das extrações podem ser visualizados nas Tabelas 3.5, 3.6 e 3.7.

Na Tabela 3.5 os propósitos elencados visam de forma geral medir as ontologias sob diferentes enfoques, como por exemplo: complexidade, evolução e qualidade.

Podemos observar na Tabela 3.6 que a maioria dos trabalhos utiliza a linguagem OWL como entrada nas ferramentas, protótipos desenvolvidos.

Além disso, é importante salientar que há interseção entre alguns trabalhos, logo, existem repetições de métricas. Na Tabela 3.7 apresentamos a quantidade de métricas em cada um dos trabalhos. Nesta mesma tabela podemos destacar que metade dos trabalhos possui a categoria estrutura.

Estes pontos foram levados em consideração no momento de construção do protótipo do presente trabalho.

Tabela 3.5 – Extração de informações gerais sobre os trabalhos.

Referência	Propósito	Ferramenta
[GOM04]	Permitir que os usuários meçam a aptidão de ontologias existentes perante requisitos de um determinado sistema.	OntoMetric
[CRO05]	Fornecer ferramentas que avaliem automaticamente características e qualidade de uma ontologia.	OntoCAT
[YAO05]	Examinar a qualidade fundamental da coesão que se refere às ontologias.	
[YAN06]	Analizar qualidade, razão e correlação de conceitos e relações, para avaliar as ontologias do ponto de vista de complexidade e evolução.	
[GAN06]	Fornecer uma lista provisória de medidas qualitativas e quantitativas para a avaliação de ontologias.	
[TAR7]	Avaliar ontologias relacionadas a um determinado conjunto de condições e classifica-las segundo um conjunto de métricas, as quais captura diferentes aspectos de ontologias.	OntoQA

Tabela 3.6 – Extração de informações gerais sobre as ontologias.

Referência	Linguagem
[GOM04]	
[CRO05]	RDFS, OWL
[YAO05]	OWL
[YAN06]	
[GAN06]	OWL
[TAR7]	RDFS, OWL

Tabela 3.7 – Extração de informações gerais sobre as métricas.

Referência	Categorização	Quantidade	Cálculo	Descrição
[GOM04]	ferramenta, metodologia, linguagem, custo, conteúdo	139	não	sim
[CRO05]	tamanho intensional e extensional, estrutura intensional e extensional	68	sim	sim
[YAO05]	coesão	3	sim	sim
[YAN06]	primitivas e complexas	10	sim	sim
[GAN06]	tipo, estrutural, funcional e usabilidade	20	não	sim
[TAR7]	estrutura e instância	9	sim	sim

Comentários Finais

Enfim, das 249 métricas obtidas nos trabalhos resultantes desta revisão sistemática 93 métricas foram analisadas. No OntoMetric [GOM04] apenas analisamos as métricas categorizadas na dimensão conteúdo (27 métricas). No OntoCAT [CRO05] as métricas de tamanho intensional e parte das métricas de estrutura intensional (eliminamos aquelas que utilizavam a WordNet) foram examinadas

(37 métricas). No trabalho [GAN06] apenas analisamos as métricas categorizadas na dimensão estrutural (7 métricas). Em [YAO05], [YAN06] e [TAR7] todas as métricas foram analisadas, os trabalhos contém respectivamente 3, 10 e 9 métricas.

Os critérios de escolha das métricas foram à frequência no conjunto de artigos, aplicabilidade nas ontologias em língua portuguesa selecionadas, retornos numéricos e abrangentes. Após realizar filtros tendo em vista os critérios expostos, das 93 métricas ficamos com um subconjunto de 20 métricas.

Este subconjunto foi dividido em três grupos:

- Grupo 1 que corresponde a sete métricas que realizam a contagem de elementos básicos (classes, propriedades e indivíduos) de uma ontologia, utilizadas nos cálculos dos demais grupos;
- Grupo 2 que corresponde a cinco métricas que levam em consideração a profundidade e a largura da estrutura, hierarquia da ontologia (máxima, média e razão entre máxima e média);
- Grupo 3 que corresponde a oito métricas que executam cálculos como médias simples e desvio padrão da combinação entre os elementos básicos (média de indivíduos por classes e desvio padrão de indivíduos por classes).

No próximo capítulo iremos explorar a implementação destas 20 métricas no contexto do presente trabalho.

4. Protótipo

Neste capítulo descrevemos o protótipo desenvolvido após realização da revisão sistemática. É importante salientar que o sistema foi construído separadamente e posteriormente inserido no portal de ontologias OntoLP.

4.1 Bibliotecas Utilizadas

Para desenvolver o protótipo optamos por APIs, as quais permitem que ferramentas trabalhem em um nível adequado de abstração. A OWL API [HOR07] é voltada principalmente para representação OWL-DL. Já a JFreeChart é uma biblioteca que permite a geração de uma variedade de gráficos para uso em aplicações.

A OWL API encontra-se em sua segunda versão. A versão 1.0 foi desenvolvida como parte do projeto WonderWeb ¹ ao passo que a versão 2.0 foi desenvolvida como parte dos projetos Co-ODE ² e TONES ³. Outro fator interessante é que a mesma é utilizada nos editores de ontologias SWOOP, Protégé 4, OntoTrack e nos *reasoners* Pellet, DIP. Além disso, trata-se de código aberto de licença GPL .

A JFreeChart também é uma biblioteca de código aberto, suporta gráficos de setores (2D e 3D), gráficos de barras (horizontal e vertical, regular e empilhadas), gráficos de linha, gráficos de dispersão, gráficos de séries temporais, dentre outros.

Enfim, tendo em vista as características expostas utilizamos estas duas APIs para manipular as ontologias em OWL disponíveis no portal OntoLP e para criar os gráficos com o resultado das métricas aplicadas nestas ontologias.

4.2 Ontologias Utilizadas

Das ontologias do portal selecionamos 10 que estão em português e que possuem sintaxe OWL para serem utilizadas nos experimentos. Deste conjunto 3 foram geradas em trabalhos de conclusão de curso, 2 em dissertações, 2 em projetos e 3 em trabalhos submetidos ao *I Portuguese and Multi-lingual Ontologies Workshop*⁴. Abaixo descrevemos detalhadamente cada uma delas.

¹<http://wonderweb.semanticweb.org/>

²<http://www.co-ode.org/>

³<http://www.tonesproject.org/>

⁴<http://www.inf.pucrs.br/ontolp/oficina.php>

Ecologia

O projeto “BLOC-Eco” [ZAVO07] da USP cujos integrantes são Claudia Zavaglia, Sandra Aluísio, Maria das Graças Volpe Nunes, Acsa Elizabeth Marques Ferreira, Deni Yuzo Kasama, Maria Amélia Quiozini, Maria Luíza Oliveira Garcia, Monique Lopes Ferraresi, Leandro Henrique Mendonça de Oliveira, Luiz Carlos Ribeiro Junior, apresenta como foco principal o desenvolvimento de uma ontologia do domínio da ecologia. São disponibilizados o *corpus* de ecologia, a lista de termos extraída manualmente deste *corpus* via critério semântico e validados por um especialista da área (lista de referência).

Nanotecnologia

Esta ontologia foi desenvolvida no contexto da dissertação de mestrado “Estruturação do conhecimento e relações semânticas: uma ontologia para o domínio da nanociência e nanotecnologia” [KAS09] da UNESP. Apresenta uma estrutura conceitual do domínio da nanociência e nanotecnologia, em língua portuguesa.

Arte

O projeto “OnAIR” [TRI05] da USP cujos integrantes são Renata Wassermann, Fabio Kon, Paula Braga, Raphael Cóbe, Christian Paz-Trillo, Carlos Eduardo Atencio Torres, Esdras Bispo Junior, envolve responder perguntas feitas em linguagem natural por usuários a partir de consultas a uma base de sistemas de recuperação de vídeo. Uma coleção de vídeos do domínio da arte contemporânea obtidos através de entrevistas com uma especialista do domínio foi utilizado e uma ontologia deste domínio foi construída a partir da fala da artista entrevistada.

Currículo Lattes

Na dissertação de mestrado “Populando ontologias através de informações em HTML - o caso do currículo lattes” [CAS08] da USP foi construída a ontologia do domínio Currículo Lattes. Para popular e para posteriormente gerar relatórios foram utilizados como fonte informação os currículos da Plataforma Lattes.

Emoção

No trabalho de conclusão de curso “PersonalImp: ferramenta para construção da personalidade e de cursos de ação para agentes improvisacionais voltados a interfaces” [WIL07] da PUCRS o objetivo é representar e gerenciar emoções em agentes de interface que possam ser aplicados em ambientes

colaborativos. Para a construção da ontologia conceitos relacionados a modelos de emoções foram buscado na área da psicologia.

Privacidade de Dados

A ontologia *Privacy* está inserida no trabalho “Privacidade e responsabilização: representação do domínio” [SIL09] da PUCRS em parceria com a HP cujos integrantes são Douglas da Silva, Mírian Bruckschen, Paulo Bridi, Roger Granada, Alexandre Agustini, Renata Vieira e Caio Northfleet. Apesar da mesma encontra-se em fase de desenvolvimento, uma parte significativa do domínio, conceitos relacionados a privacidade e responsabilização, já foram modelados.

Rede de Conhecimento Científico

O trabalho “Um Modelo para Redes de Conhecimento Científico” [PAI09] da USP, desenvolvido por Daniel C. Paiva, Marcos L. Mucheroni e Marcio L. Netto, propõe a união de Redes Sociais, Redes de Conhecimento Científico e as Redes de Co-citação. Na ontologia construída são incorporados os relacionamentos entre autores, incluindo áreas temáticas e, a partir de artigos, as co-citações que dão dinâmica maior às redes.

Smartphone

No trabalho de conclusão de curso “Ontologia de domínio para análise de blogs” [MER07] da Feevale a ontologia do domínio *smartphone* foi construída e tem como objetivo apoiar a realização de pesquisas e auxiliar na análise dos conteúdos de *blogs*. Ela foi modelada, populada e avaliada através da execução de experimentos de inferência.

Música

No trabalho de conclusão de curso “Educação Musical à Distância utilizando Ontologias” [BOF05] da Feevale a ontologia do domínio música foi construída e trata-se de uma ontologia que engloba a história da música, incluindo gêneros musicais, obras e autores. Com base nesta ontologia foi desenvolvido um *quiz* (sistema de perguntas e respostas) onde a pessoa que realiza uma pergunta recebe dicas semânticas do tipo gênero, obra e autor.

Equivalência de Estímulos

No trabalho “Stimuli: uma ontologia para paradigma de Equivalência de Estímulos” [BEL09] da UFSCar cujos integrantes são Rodrigo E. Bela, Marilde T. P. dos Santos e Mauro Biajiz o objetivo foi desenvolver uma ferramenta de ensino com base no paradigma de equivalência de estímulos

(programas de ensino baseados neste paradigma têm influenciado positivamente as pesquisas sobre educação na psicologia), por este motivo precisou-se construir a ontologia *Stimuli*.

4.3 Métricas Implementadas

As fórmulas das métricas implementadas, retiradas dos artigos selecionados na revisão sistemática, tiveram suas estruturas padronizadas como exposto no Capítulo 2 através da 6-tupla $O := \{C, P, A, H^C, prop, att\}$. Onde utilizamos: C para conceitos, P para relações, A para relações que são atributos, $H^C(C_i, C_j)$ para hierarquia de conceitos onde C_i é uma subclasse de C_j , $prop$ para função que relaciona os conceitos de modo não taxonômico e att para função que relaciona os conceitos com valores literais (*string*). A base de conhecimento também deve ser padronizada para isso utilizamos os elementos: I conjunto de instâncias e $inst(C)$ função classes instanciadas [TAR05, TAR7].

Além disso utilizamos $Count$ para contagem, $SuperClass$ para a função superclasse, $SubClass$ para a função subclasse, Max para máximo ($MaxDepth$ e $MaxWidth$), $Depth$ para profundidade, $Width$ para largura, Avg para média ($AvgDepth$, $AvgMaxMed$, $AvgWidth$, $AvgAttClass$, $AvgPropClass$, $AvgPClass$, $AvgC_iC$, $AvgC_iC_j$, $AvgCIC$, $AvgIC$) e $StdDev$ para desvio padrão ($StdDevIC$). Outro fator que devemos ressaltar é que todas as métricas apresentadas resultam em valores numéricos.

1. Número de classes raiz: Indica o número de classes que não possuem superclasses [YAO05, CRO05] (Apêndice C Figura C.1).

$$Count(C_j) \tag{4.1}$$

$$\text{onde: } SuperClass(C_j) = 0$$

2. Número de classes folha: Indica o número de classes que não possuem subclasses [YAO05, CRO05, YAN06] (Apêndice C Figura C.2).

$$Count(C_i) \tag{4.2}$$

$$\text{onde: } SubClass(C_i) = 0$$

3. Número total de classes: Indica o número total de classes definidas. O termo conceito também é utilizado no lugar de classes [GOM04, CRO05, YAN06] (Apêndice C Figura C.3).

$$Count(C) \tag{4.3}$$

4. Número de propriedades (atributos): Indica o número de propriedades (atributos), denominado de *DatatypeProperty* em OWL. Representa as relações entre classes e tipos [CRO05] (Apêndice C Figura C.4).

$$Count(att) \quad (4.4)$$

5. Número de propriedades (relações): Indica o número de propriedades (relações), denominado de *ObjectProperty* em OWL. Representa as relações entre classes [CRO05] (Apêndice C Figura C.5).

$$Count(prop) \quad (4.5)$$

6. Número total de propriedades: Indica o número total de propriedades definidas. Somatório de atributos e relações [GOM04, CRO05, YAN06] (Apêndice C Figura C.6).

$$Count(P) \quad (4.6)$$

7. Número total de instâncias: Indica o número total de instâncias. As instâncias representam objetos no domínio de interesse. Instâncias também são conhecidas como indivíduos (Apêndice C Figura C.7).

$$Count(I) \quad (4.7)$$

8. Profundidade máxima: Indica o tamanho do caminho mais longo na hierarquia da ontologia [GOM04, CRO05, YAN06] (Apêndice C Figura C.8).

$$MaxDepth(H^C) = Max(Depth(C_j)) \quad (4.8)$$

9. Profundidade média: Indica a razão entre o somatório da profundidade de todos os caminhos pelo número total de caminhos [CRO05, GAN06, YAO05] (Apêndice C Figura C.9).

$$AvgDepth(H^C) = (\sum_j Depth(C_j)) / Count(C_i) \quad (4.9)$$

10. Média da profundidade máxima pela profundidade média: Indica a agregação das classes e a coesão da ontologia [YAN06] (Apêndice C Figura C.10).

$$AvgMaxMed(H^C) = MaxDepth(H^C) / AvgDepth(H^C) \quad (4.10)$$

11. Largura máxima: Indica o maior número de classes vizinhas nos níveis da ontologia [CRO05] (Apêndice C Figura C.11).

$$MaxWidth(H^C) = Max(Width(C_k)) \quad (4.11)$$

onde: $C_k = \text{classescomsubclasses}$, $Width(C_k) = Count(SubClass(C_k))$

12. Largura média: Indica a razão entre o somatório da largura de todas as classes com subclasses e o número total de classes com subclasses [CRO05](Apêndice C Figura C.12).

$$AvgWidth(H^C) = \sum_k (Width(C_k)) / Count(C_k) \quad (4.12)$$

13. Média do número de propriedades (atributos) pelo número total de classes: Indica a média simples entre propriedades (atributos) pelo número total de classes [CRO05](Apêndice C Figura C.13).

$$AvgAttClass = Count(att) / Count(C) \quad (4.13)$$

14. Média do número de propriedades (relações) pelo número total de classes: Indica a média simples entre propriedades (relações) pelo número total de classes [CRO05, YAN06] (Apêndice C Figura C.14).

$$AvgPropClass = Count(prop) / Count(C) \quad (4.14)$$

15. Média do número de propriedades pelo número total de classes: Indica o grau de conectividade médio das classes [CRO05] (Apêndice C Figura C.15).

$$AvgPClass = Count(P) / Count(C) \quad (4.15)$$

16. Média do número de classes folha pelo número total de classes: Indica a média simples entre o número de classes folha pelo número total de classes [CRO05, YAN06, TAR7] (Apêndice C Figura C.16).

$$AvgC_iC = Count(C_i) / Count(C) \quad (4.16)$$

17. Média do número de classes folha pelo número de classes raiz: Indica a média simples entre o número de classes folha pelo número de classes raiz [YAN06] (Apêndice C Figura C.17).

$$AvgC_iC_j = Count(C_i) / Count(C_j) \quad (4.17)$$

18. Média do número de classes populadas pelo número total de classes: Indica a riqueza da base de conhecimento [TAR7] (Apêndice C Figura C.18).

$$AvgCIC(H^C) = Count(inst(C)) / Count(C) \quad (4.18)$$

19. Média do número de instâncias pelo número total de classes: Indica a média simples entre o número total de instâncias pelo número total de classes (Apêndice C Figura C.19).

$$AvgIC(H^C) = Count(I) / Count(C) \quad (4.19)$$

20. Desvio padrão do número total de instâncias pelo número total de classes: Indica como as instâncias estão espalhadas entre as classes de uma estrutura [TAR7] (Apêndice C Figura C.20).

$$StdDevIC(H^C) = (\sum_i (Count(inst(C))_i - AvgIC(H^C))) / Count(I) \quad (4.20)$$

As métricas média do número total de instâncias pelo número total de classes e número total de instâncias foram calculadas e utilizadas em uma terceira métrica, o desvio padrão do número total de instâncias pelo número total de classes, não sendo diretamente referenciadas nos trabalhos elencados.

O conjunto de trabalhos obtidos na revisão dividem as métricas selecionadas em intensional de tamanho e de estrutura [CRO05]; coesão [YAO05]; dimensão estrutura [GAN06]; dimensão conteúdo [GOM04]; primitiva e complexa [YAN06]; estrutura e instância [TAR7].

Por tratar-se de ontologias de diferentes domínios, estratégias como por exemplo matriz de correlação e classificadores não foram utilizados para agrupar as métricas implementadas. Além disso, cada iniciativa apresenta divisões distintas referentes as métricas, definidas em função dos objetivos, propósitos do trabalho.

Sendo assim, subdividimos neste trabalho as métricas em grupos, onde: Grupo 1 representa simplesmente a contagem de elementos básicos de uma ontologia (classes, propriedades e instâncias); Grupo 2 representa a profundidade e a largura da estrutura; Grupo 3 representa o cálculo de médias simples e de desvio padrão a partir de combinações dos elementos básicos.

4.4 Protótipo no Portal OntoLP

No sistema temos como entrada as ontologias no formato OWL e como saídas um vetor com os valores obtidos através do cálculo das métricas, outro vetor com os nomes das métricas e os gráficos de cada ontologia utilizada neste experimento (Apêndice B). As Figuras B.1,B.2,B.3,B.4, B.5, B.6, B.7, B.8, B.9 e B.10, dos seguintes domínios são apresentadas: ecologia, nanotecnologia, arte contemporânea, currículo lattes, emoção, privacidade de dados, rede de conhecimento científico, smartphone, música e equivalência de estímulos.

O portal foi desenvolvido utilizando linguagens como HTML, CSS , PHP e Java Script, possui uma interface simples e de fácil manipulação.

A partir da inserção do protótipo no Portal OntoLP, as métricas podem ser visualizadas por meio de tabelas contendo nome, valor e sua documentação/descrição (Figuras 4.1) e gráficos (Figuras 4.2).

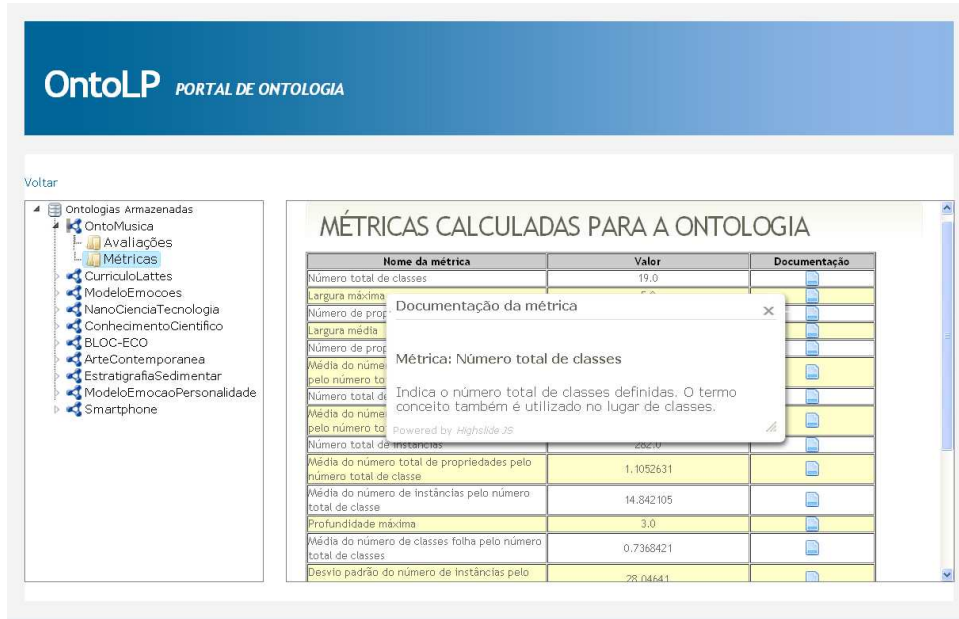


Figura 4.1 – Portal OntoLP descrição das métricas.

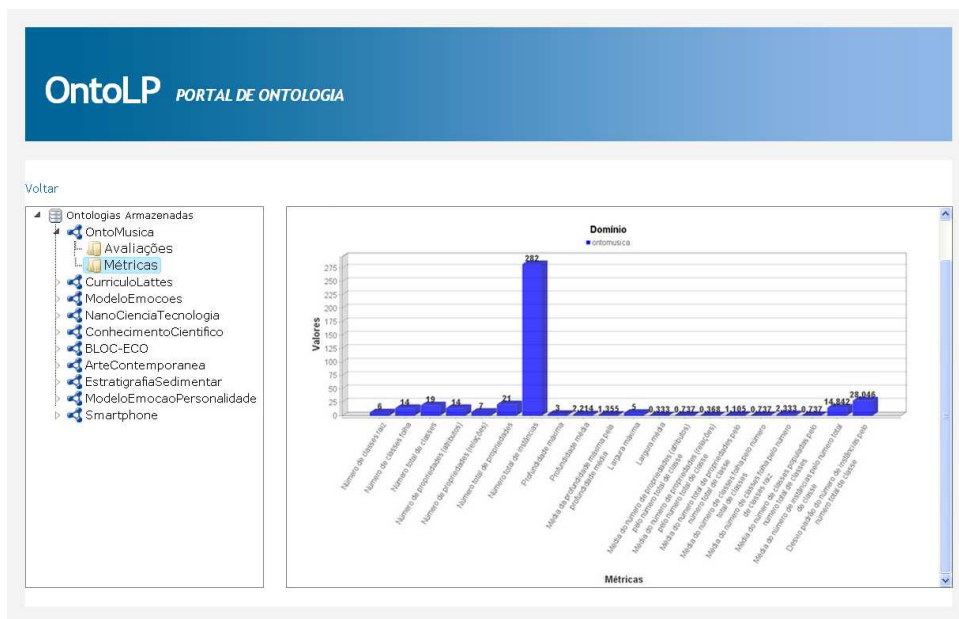


Figura 4.2 – Portal OntoLP gráfico das métricas.

5. Avaliação

Neste capítulo detalhamos a elaboração e a aplicação de um questionário utilizado como método de avaliação dos resultados das métricas implementadas.

5.1 Elaboração do Questionário

Questionários são instrumentos de pesquisa que visam avaliar opiniões, objetivos, preferências e crenças de pessoas de forma rápida e simples [WAI07]. Eles têm vantagens sobre outros tipos de pesquisas, pois são baratos, exigem pouco esforço do autor da pergunta, podem ser aplicados via telefone, correio ou email, e muitas vezes apresentam respostas padronizadas que facilitam a compilação de dados [OAT06]. Sua utilização envolve os seguintes itens: elaboração das perguntas e respostas, amostragem da população, avaliação das respostas e análise dos resultados [WAI07].

Na elaboração das perguntas alguns cuidados devem ser tomados, normalmente é aconselhável clareza da pergunta (a pergunta deve conter o mínimo de informações possíveis de modo a se tornar compreensível e que não deixe margem para dúvidas), adequação da linguagem utilizada na pergunta (deve possuir uma linguagem simples, buscando sempre o entendimento da pergunta por quem está sendo entrevistado, independentemente de seu grau de instrução) e objetividade da pergunta (deve ser curta e direta de modo a não causar cansaço no pesquisador e no respondente) [GUN03].

Já a elaboração das respostas diz respeito aos tipos de questões fechadas e/ou abertas. Para uma pesquisa inicial ao não se conhecer a abrangência ou a variabilidade das possíveis respostas, são necessárias perguntas abertas. Uma vez que se conhecem os tópicos geralmente mencionados pelos respondentes a cerca de uma dada temática, especialmente quando existem muitos respondentes ou pouco tempo, deve-se usar perguntas fechadas [GUN03].

Uma prática comum é usar a escala de Likert, na qual as questões são escritas como afirmativas e o respondente deve escolher uma das alternativas, por exemplo: discordo totalmente, discordo, neutro, concordo, concordo totalmente [GUN03]. Esta escala é muito utilizada no levantamento de atitudes, opiniões e avaliações. Na escala nominal utiliza-se número ou símbolos para identificar pessoas, objetos ou categorias, por exemplo: país, grau de instrução, estado civil e sexo.

No questionário elaborado (Apêndice C) utilizamos questões fechadas e abertas. Como questões abertas podemos elencar as questões 1b, 1c, 3, 5, 10. E como questões fechadas podemos citar aquelas que apresentam as seguintes escalas: escala nominal (questões 1a, 1d, 2, 4) e escala Likert (questões 6, 7, 8, 9).

É válido salientar que sempre convém realizar um estudo piloto (pré-teste) para verificar se e como as perguntas estão sendo entendidas pelo público-alvo. Neste questionário amostra da população corresponde aos usuários, pesquisadores e engenheiros de ontologias. Na fase de pré-teste obtivemos

17 retornos, e foi possível detectar que mais de um campo foi selecionado nas questões 6, 7, 8 e 9 e que outros campos foram deixados em branco. Na fase final obtivemos 30 retornos (Apêndice D).

Outro fator que merece ser mencionado é a utilização dos dados do questionário, usualmente estes são apenas reportados. Além disso, a distribuição das respostas de cada questão (ou das questões mais relevantes) deve ser descrita. E apesar da tabulação dos dados ser uma tarefa árdua ela deve ser feita ao final da pesquisa, possibilitando a avaliação das respostas e a análise dos resultados.

Ferramentas para automatizar o processo de construção de questionários têm sido elaboradas, são exemplos: questionpro¹ e surveymonkey².

5.2 Aplicação do Questionário

O objetivo do questionário elaborado foi verificar a opinião de um grupo de pesquisadores, engenheiros e usuários de ontologias sobre a utilidade das métricas na inspeção de ontologias disponibilizadas em bibliotecas especializadas, como no portal OntoLP.

Na fase de pré-teste (Apêndice D.1), os questionários (no formato impresso) foram aplicados nos participantes do evento sobre ontologias, *I Portuguese and Multi-lingual Ontologies Workshop*, que ocorreu no mês de setembro de 2009 em São Carlos. Após, algumas decisões foram tomadas, tais como: reformulação de questões e inserção de novas questões de maneira a complementar e suprir as necessidades da pesquisa.

Na fase final (Apêndice D.2), o questionário sofreu alterações e foi aplicado no formato digital. Para facilitar o processo de tabulação dos dados as questões respondidas foram salvas em XML. Deixamos este disponível pelo período de um mês (6 de dezembro 2009 até 6 de janeiro 2010).

Os retornos do questionário final mostraram-se satisfatórios, uma vez que 76,67% diz já ter construído ontologias e 83,34% diz já ter buscado por ontologias descritas nos idiomas inglês, português e espanhol. Sendo o português o segundo mais citado, perdendo apenas para o inglês (questões 2, 4 e 5). Outro fator é a amostra, a qual é constituída de 14 doutores, 11 mestres, 2 especialistas e 3 graduados (questão 1). Os domínios indicados são variados (questão 3).

Nas perguntas referentes a utilidade das métricas implementadas agrupadas ou divididas em grupos (questões 6, 7, 8 e 9) na inspeção das ontologias disponíveis no portal OntoLP o retorno concordo e concordo totalmente prevaleceu perante os demais.

Conforme Figura 5.1 15 (a) e 16 (b) pessoas concordaram totalmente, 13 (c) e 12 (d) pessoas concordaram.

¹<http://www.questionpro.com/>

²<http://www.surveymonkey.com/>

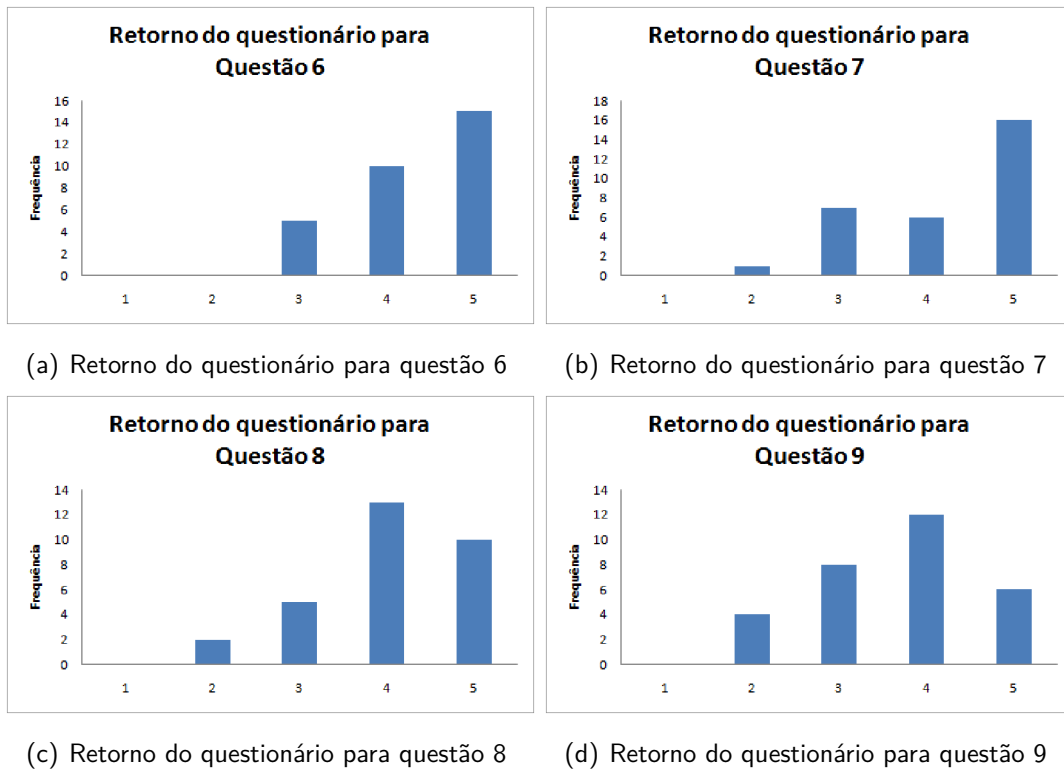


Figura 5.1 – Retorno para as questões relativas às métricas implementadas.

As respostas indicam que os usuários consideram as métricas úteis e ajudam na compreensão dos recursos disponibilizados. Observamos também que a análise das métricas de forma isolada, ou seja, subdivididas nos grupos 1, 2 e 3, não se mostraram tão satisfatórias se comparadas com o agrupamento de todas elas.

Observações (questão 10) feitas pelos respondentes como questionamentos a respeito da falta de métricas mais específicas relacionadas às relações presentes (número de tipos de relações, média de relações por tipo de relação, desvio padrão de relações por tipo de relação, onde tipo de relações seriam “is-a”, “part-of”) e também do interesse da apresentação de valores percentuais (percentual de classes que não possuem superclasses) se encaixam nos trabalhos futuros, listados no Capítulo 6.

Visto que um número significativo de pessoa, 25 respondentes, buscam por recurso desta natureza podemos considerá-los importantes tanto para o portal quanto para a comunidade.

6. Considerações Finais

Nesse capítulo expomos as contribuições e os trabalhos futuros desta dissertação.

6.1 Contribuições

Consideramos como contribuições desta dissertação a revisão sistemática elaborada sobre métricas para ontologias, a aplicação das métricas selecionadas nas ontologias em língua portuguesa e a inserção do sistema de extração de métricas no portal OntoLP. Uma motivação para esse trabalho foi buscar maneiras de informar a comunidade sobre ontologias disponíveis em repositórios específicos.

A aplicação de métricas nas ontologias em língua portuguesa é proposital, já que possibilita uma visão geral do está sendo desenvolvido em português. É importante ressaltar que estas métricas poderiam ser aplicadas em ontologias de qualquer idioma, não sendo projetadas para uma única língua. Porém até o presente momento elas são aplicáveis a ontologias descritas somente no formato OWL.

A apresentação dos valores obtidos foi feita de duas maneiras, tabelas e gráficos. E com a finalidade de avaliarmos o quanto as métricas implementadas são informativas, elaboramos um questionário e aplicamos o mesmo a potenciais usuários, engenheiros e pesquisadores da área de ontologias. Através dos resultados obtidos foi possível atestar a utilidade das métricas para ontologias no Portal OntoLP, como apresentado no Capítulo 5.

É importante ressaltar que este trabalho está inserido no projeto intitulado “Pesquisa e desenvolvimento de ontologias para a língua portuguesa - OntoLP” que contempla organizar e realizar um levantamento de recursos existentes, desenvolver um portal de informações sobre ontologias em língua portuguesa e publicações relacionadas a estas ontologias.

Neste contexto, em especial, no ambiente do Portal OntoLP, as métricas como já apresentadas têm um papel importante na busca por ontologias e conseqüentemente no reuso, uma vez que a busca faz parte do processo de reuso.

6.2 Trabalhos Futuros

A partir desta dissertação foi possível identificar os seguintes trabalhos futuros:

- Expansão do conjunto de ontologias em português: para darmos continuidade a este trabalho as métricas poderiam ser aplicadas em uma amostra maior, permitindo uma análise mais fiel do que vem sendo elaborado ao longo dos anos. Outro fator que merece ser lembrado é a

demorada tarefa de buscar por recursos deste gênero, por isso a aplicação em somente 10 ontologias;

- Expansão do conjunto de métricas implementadas: aumentar o conjunto de métricas implementadas para dar mais informações ao usuário, no entanto uma posterior seleção pode mostrar-se necessária para evitar que o usuário tenha acesso a informações desnecessárias, repetidas ou conflitantes;
- Considerar avaliações qualitativas e não apenas quantitativas: além de métricas quantitativas considerar avaliações qualitativa. Como exemplo podemos citar as observações descritas por um dos respondentes do questionário apresentado no Capítulo 5, o qual sugere a identificação dos tipos de relações existentes na ontologia “is-a”, “part-of”;
- Criação de novas métricas: sugerir novas métricas que complementem as informações obtidas nas métricas atuais, sejam elas baseadas na métricas atuais ou baseadas em novas teorias;
- Criação de novas interfaces além da *Web*: por fim criar novas interfaces de extração de métricas como por exemplo *plugin* para editores de ontologias, para navegadores *Web* (*Web Semântica*) e GUI .

Após realização dos experimentos foi possível constatar que as ontologias presentes no portal (considerando-se as 10 ontologias totalmente em língua portuguesa) são de estrutura simplificada pois representam em sua maioria hierarquias formais com instâncias. Ontologias construídas em nosso idioma são recentes se comparamos como ontologias em inglês, mas se fazem necessárias principalmente nos domínios governo eletrônico, cultura brasileira (folclore, música) e saúde coletiva.

Bibliografia

- [ANT08] Antoniou, G. e Harmelen, F. “A Semantic Web Primers”. MIT Press, 2008, Segunda Edição, 268p.
- [BER01] Berners-Lee, T.; Hendler, J. e Lassila, O. “The Semantic Web”. *Scientific American*, vol. 284, Maio 2001, pp. 34–43.
- [BIO05] Biolchini, J.; Mian, P. G.; Natali, A. C. C. e Horta, G. Travassos. “Systematic Review in Software Engineering”. Relatório Técnico, COPPE/UFRJ, 2005.
- [BOF05] Boff, R. E. “Educação Musical à Distância Utilizando Ontologias”. Trabalho de Conclusão do Curso de Ciência da Computação, Feevale, 2005, 87p.
- [BEL09] Bela, R. E.; Santos, M. T. P. e Biajiz, M. “Stimuli: Uma Ontologia para Paradigma de Equivalência de Estímulos. In: First Portuguese and Multi-lingual Ontologies Workshop, 2009, 2p.
- [CAS08] Castano, A. C. “Populando Ontologias Através de Informações em HTML - O Caso do Currículo Lattes”. Dissertação de Mestrado, Programa de Pós-Graduação em Ciência da Computação, USP, 2008, 90p.
- [CRO05] Cross, V. e Pal, A. “Metrics for ontologies”. In: Annual Meeting of the North American Fuzzy Information Processing Society, 2005, pp. 448–453.
- [GAN06] Gangemi, A.; Catenacci, C.; Ciaramita, M. e Lehmann, J. “Modelling ontology evaluation and validation”. In: The Semantic Web: Research and Applications, 2006, pp. 140–154.
- [GOM04] Gómez-Pérez, A.; Fernández-López, M. e Corcho, O. “Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web”. Springer, Quarta Edição, 2004, 421p.
- [GRU95] Gruber, T. R. “Toward Principles for the Design of Ontologies used for Knowledge Sharing”. *International Journal of Human Computer Studies*, vol. 43, 1995, pp. 907–928.
- [GUA98] Guarino, N. “Formal Ontology in Information Systems”. In: First International Conference on Formal Ontology in Information Systems, 1998, pp. 3–15.
- [GUN03] Gunther, H. “Como Elaborar um Questionário”. Capturado em <http://www.psi-ambiental.net/pdf/01Questionario.pdf>, 2003
- [HOR07] Horridge, M.; Bechhofer, S. e Noppens, O. “The OWL API”. In: OWL: Experiences and Directions Third International Workshop, 2007, 9p.

- [IEE90] IEEE "IEEE Standard Glossary of Software Engineering Terminology". Relatório Técnico, IEEE, 1990.
- [KAS09] Kasama, D. Y. "Estruturação do Conhecimento e Relações Semânticas: Uma Ontologia para o Domínio da Nanociência e Nanotecnologia". Dissertação de Mestrado, Programa de Pós-Graduação em Estudos Lingüísticos do Instituto de Biociências, Letras e Ciências Exatas, UNESP, 2009, 178p.
- [KIT04] Kitchenham, B. "Procedures for Performing Systematic Reviews". Relatório Técnico, Keele University, 2004.
- [LAS01] Lassila, O e McGuinness, D. L. "The Role of Frame-Based Representation on the Semantic Web". *Linkoping Electronic Articles in Computer and Information Science*, vol. 6, 2001, pp. 78–87.
- [LOZ04] Lozano-Tello, A. e Gómez-Pérez, A. "Ontometric: A Method to Choose the Appropriate Ontology". *Journal of Database Management*, vol. 15, 2004, pp. 1–18.
- [MER07] Mertins, F. "Ontologia de Domínio para Análise de Blogs". Trabalho de Conclusão do Curso de Ciência da Computação, Feevale, 2007, 94p.
- [MAE02] Maedche, A. e Zacharias, V. "Clustering Ontology-Based Metadata in the Semantic Web". In: 6th European Conference on Principles of Data Mining and Knowledge Discovery, 2002, pp. 383–408.
- [NOY01] Noy, N. F. e McGuinness D. L. "Ontology Development 101: A Guide to Creating your First Ontology". Relatório Técnico, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics, 2001.
- [OAT06] Oates, B. J. "Researching Information Systems and Computing". Sage Publications, Primeira Edição, 2006, 341p.
- [PIN00] Pinto, H. S. e Martins, J. P. "Reusing Ontologies". In: AAAI 2000 Spring Symposium Series, Workshop on Bringing Knowledge to Business Processes, 2000, pp. 77–84.
- [PAI09] Paiva, D. C.; Mucheroni, M. L. e Netto, M. L. "Um Modelo para Redes de Conhecimento Científico". In: First Portuguese and Multi-lingual Ontologies Workshop, 2009, 2p.
- [RUS04] Russell, S. e Norvig, P. "Inteligência Artificial". Editora Campus, Segunda Edição, 2004, 1040p.
- [SIL09] Silva, D.; Bruckschen, M.; Bridi, P.; Granada, R.; Agustini, A.; Vieira, R. e Northfleet, C. "Privacidade e Responsabilização: Representação do Domínio". In: First Portuguese and Multi-lingual Ontologies Workshop, 2009, 2p.

- [SPE07] Speretta, M. e Gauch, S. "Automatic Ontology Identification for Reuse". In: IEEE/WIC/ACM International Conference on Web Intelligence, 2007, pp. 419–422.
- [SUR07] Sure, Y. e Vrandecic, D. "How to Design Better Ontology Metrics". In: 4th European Conference on The Semantic Web, 2007, pp. 311–325.
- [SMI04] Smith, M. K.; Welty, C. e McGuinness, D. L. "OWL Web Ontology Language Guide". Capturado em: <http://www.w3.org/TR/owl-guide/>, Fevereiro 2004.
- [WIL07] Wilkens, R. S e Moraes, M. C. "Proposta de uma Ontologia para Representar e Gerenciar Emoções em Agentes". In: Workshop e Escola de Sistemas de Agentes para Ambientes Colaborativos, 2007, 10p.
- [TAR7] Tartir, S. e Arpinar, I. B. "Ontology Evaluation and Ranking Using OntoQA". In: First IEEE International Conference on Semantic Computing, 2007, pp. 185–192.
- [TAR05] Tartir, S.; Arpinar, I. B.; Moore, M.; Sheth, A. P. e Aleman-Meza, B. "OntoQA: Metric-Based Ontology Quality Analysis". In IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources, 2005, 9p.
- [TRI05] Trillo, C. D. P. "Recuperação de Vídeos Indexados por Conceitos". Dissertação de Mestrado, Programa de Pós-Graduação em Ciência da Computação, USP, 2005, 84p.
- [WEL01] Welty, C. e Guarino, N. "Supporting Ontological Analysis of Taxonomic Relationships". *Data and Knowledge Engineering*, vol. 39, 2001, pp. 51–74.
- [WAI07] Wainer, J.; Kowaltowski, T. e Breitman, K. "Método de Pesquisa Quantitativa e Qualitativa para a Ciência da Computação". Capturado em: <http://www.ic.unicamp.br/~wainer/papers/metod07.pdf>, Janeiro 2007.
- [YAO05] Yao, H.; Orme, A. M. e Eitzkorn, L. "Cohesion metrics for ontology design and application". *Journal of Computer Science*, vol. 1, 2005, pp. 107–113.
- [YAN06] Yang, Z.; Zhang, D. e Ye, C. "Evaluation Metrics for Ontology Complexity and Evolution Analysis". In: *IEEE International Conference on e-Business Engineering*, 2006, pp. 162–170.
- [ZAVO07] Zavaglia, C.; Aluísio, S.; Nunes, M. G. V e Oliveira, L. "Estrutura Ontológica e Unidades Lexicais: uma Aplicação Computacional no Domínio da Ecologia". In: 5º Workshop em Tecnologia da Informação e da Linguagem Humana, 2007, pp. 1575–1584.

Apêndice A. Protocolo Elaborado

1. Formulação da Questão

1.1 Questão-foco

1.2 Qualidade e amplitude da questão

- Problema
- Questão
- Palavras-chave e sinônimos
- Intervenção
- Controle
- Efeito
- Medição dos resultados
- População
- Aplicação
- Projeto experimental

2. Seleção das fontes

2.1 Definição dos critérios de seleção das fontes

2.2 Idioma dos estudos

2.3 Identificação e seleção das fontes após avaliação

- Método de pesquisa das fontes
- String de busca

3. Seleção dos estudos

3.1 Definição dos estudos

- Critérios de inclusão e exclusão dos estudos
- Definição do tipo de estudo
- Procedimentos adotados para a seleção dos estudos

3.2 Execução da seleção

- Seleção inicial dos estudos
- Avaliação da qualidade dos estudos
- Revisão da seleção

4. Extração de informações

4.1 Critérios de inclusão e exclusão de informações

4.2 Formulário de extração dos dados

4.3 Execução da extração

5. Sumarização dos resultados

5.1 Apresentação dos resultados em tabelas

5.2 Comentários finais

Apêndice B. Resultado da Aplicação das Métricas

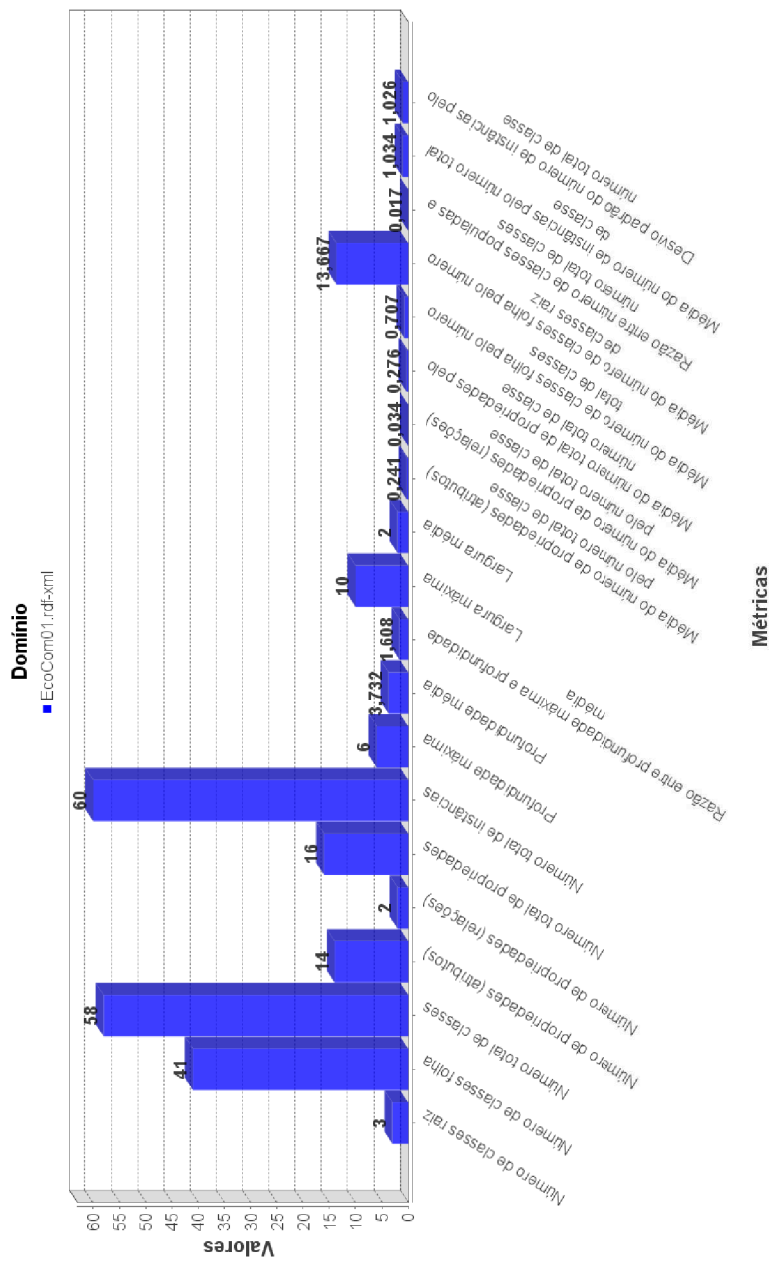


Figura B.1 – Ontologia de Ecologia.

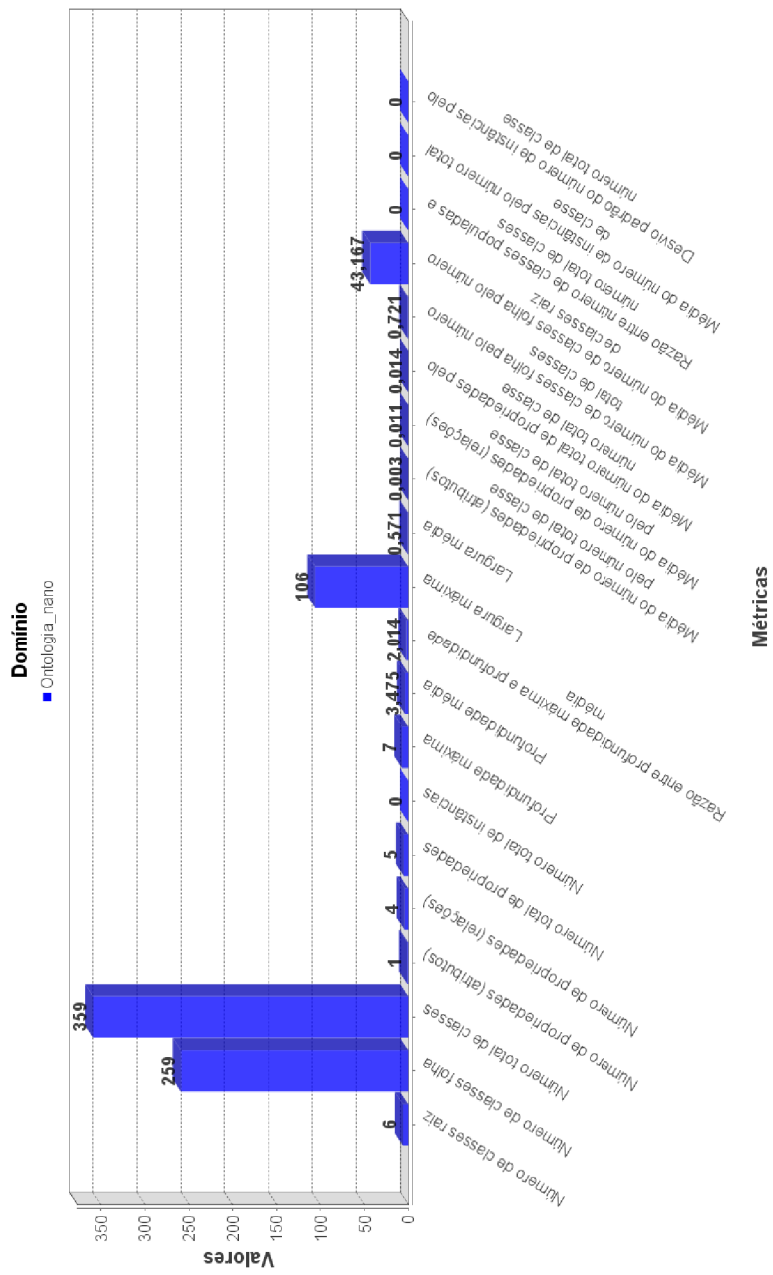


Figura B.2 – Ontologia de Nanotecnologia.

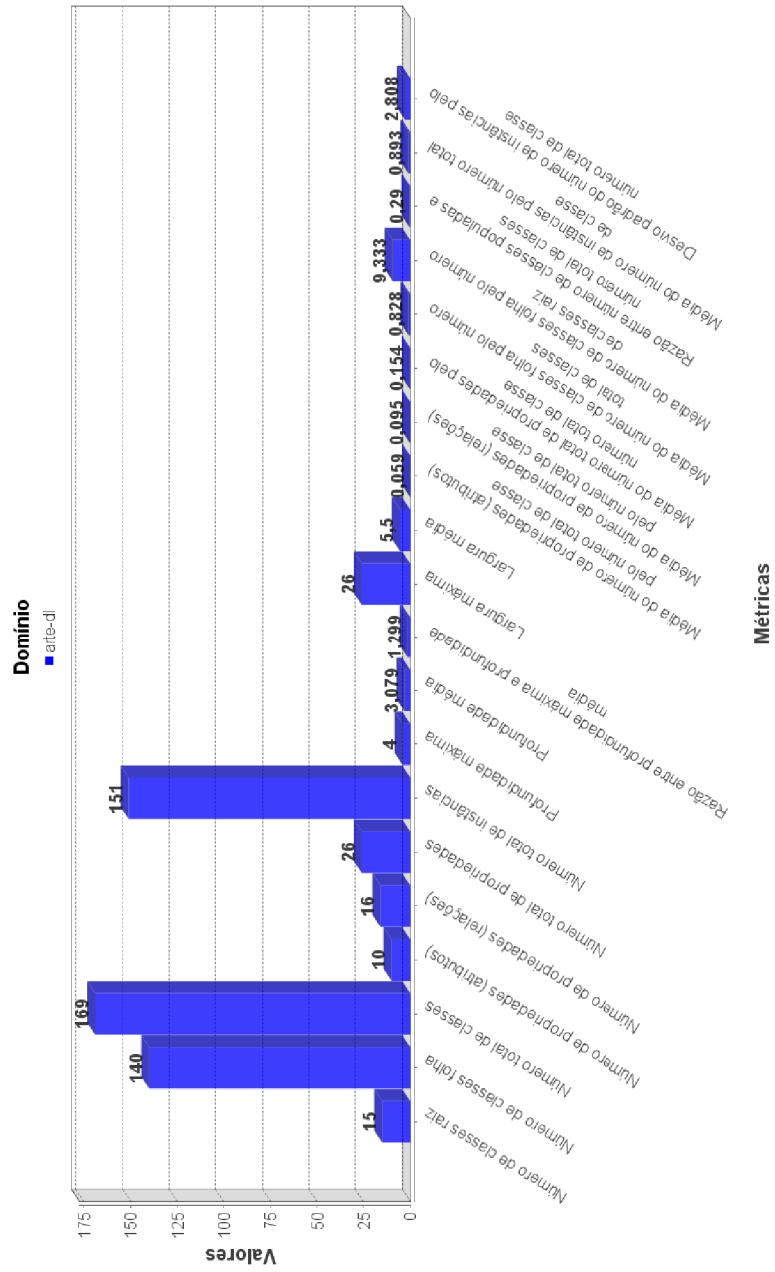


Figura B.3 – Ontologia de Arte Contemporânea.

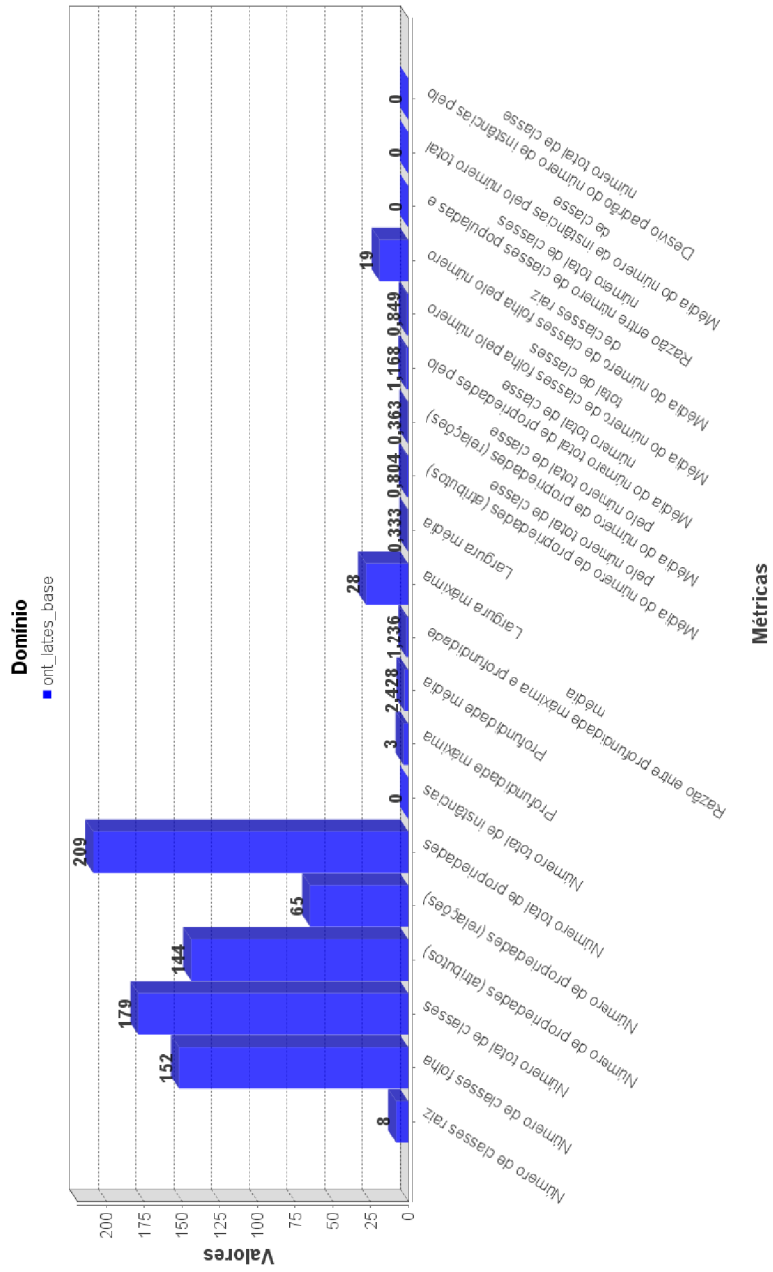


Figura B.4 – Ontologia de Currículo Lattes.

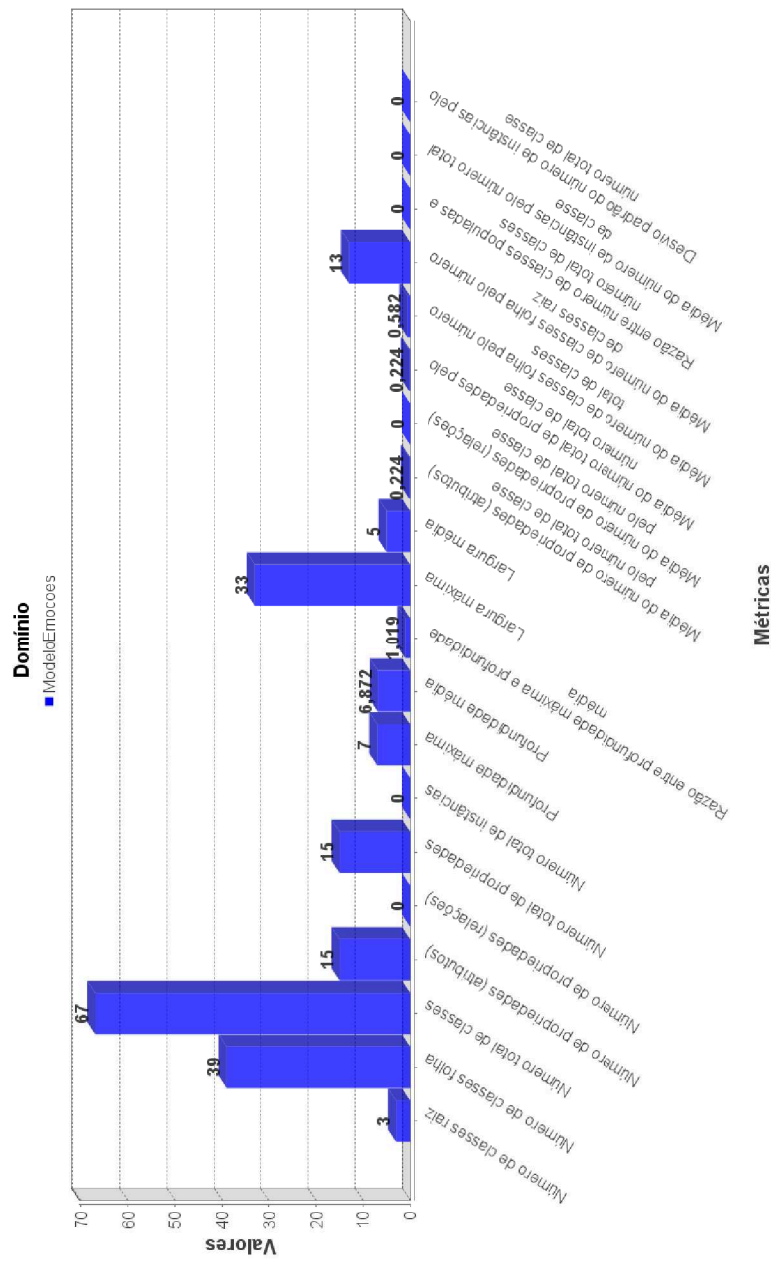


Figura B.5 – Ontologia de Emoção.

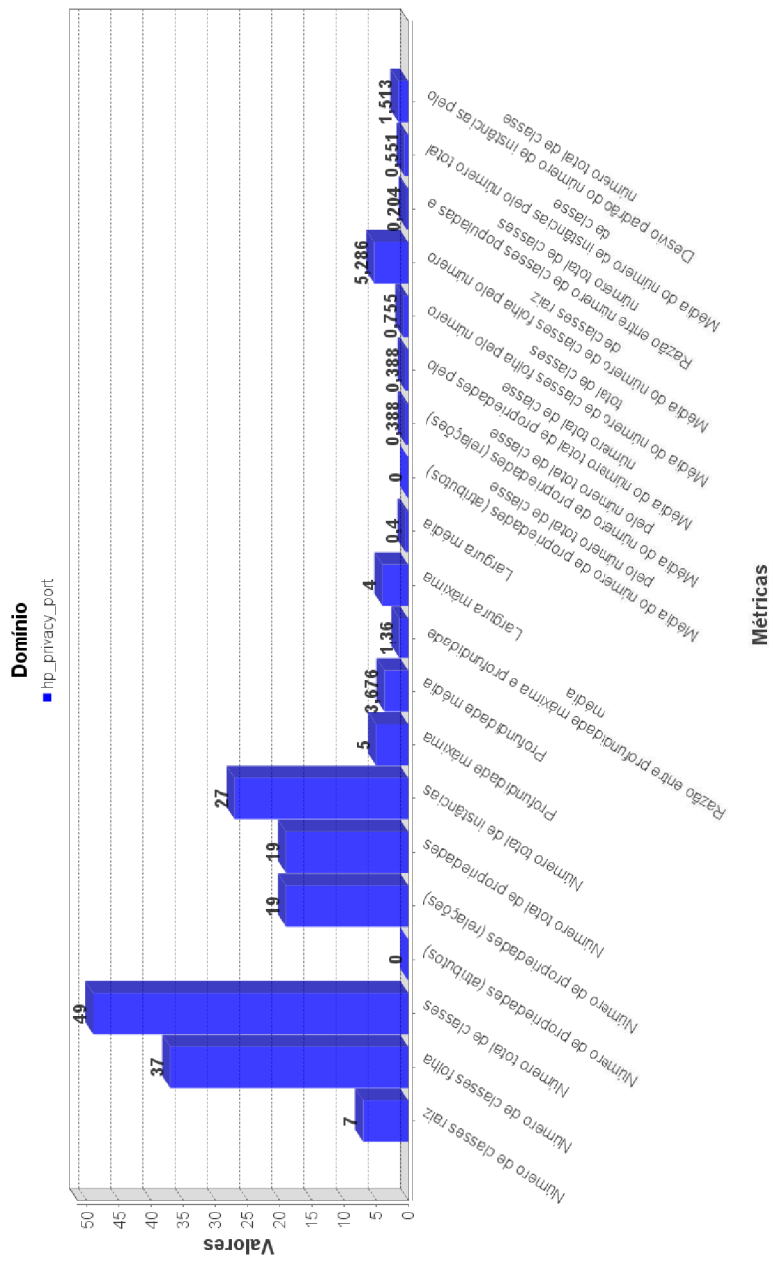


Figura B.6 – Ontologia de Privacidade de Dados.

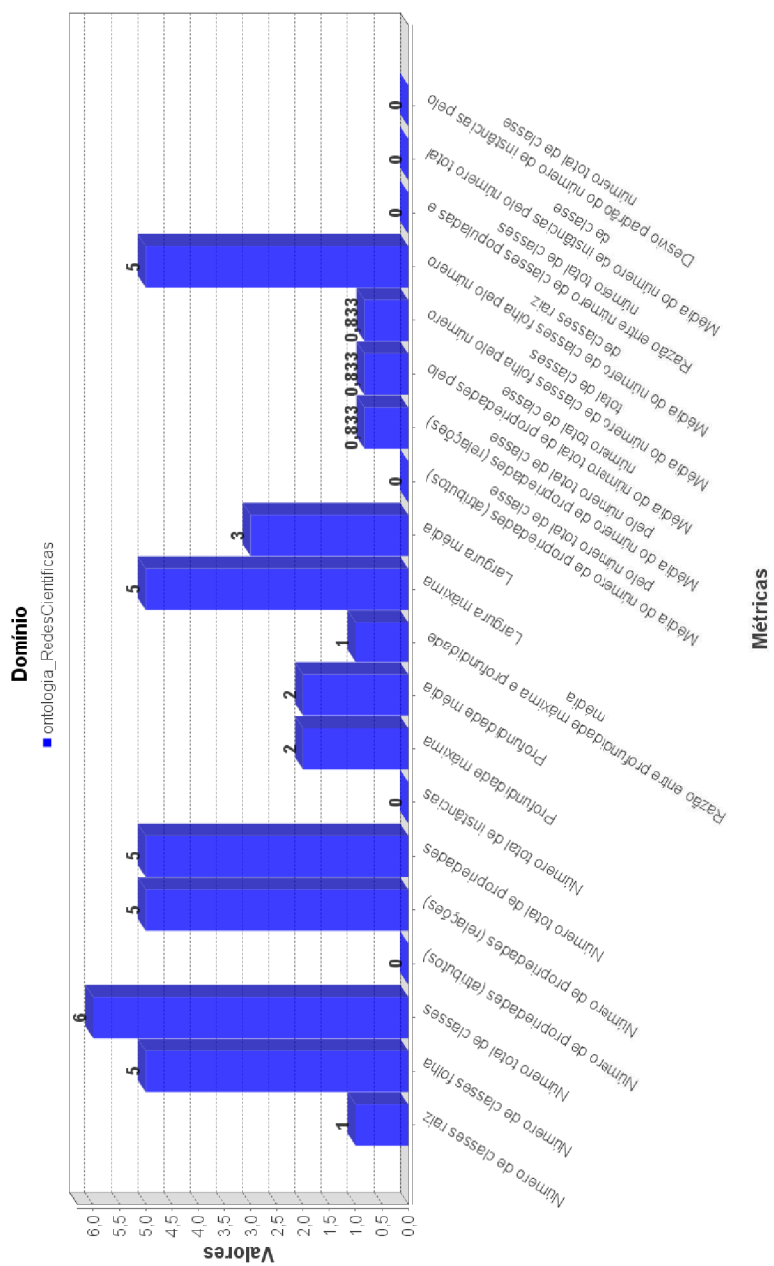


Figura B.7 – Ontologia de Rede de Conhecimento Científico (RCC).

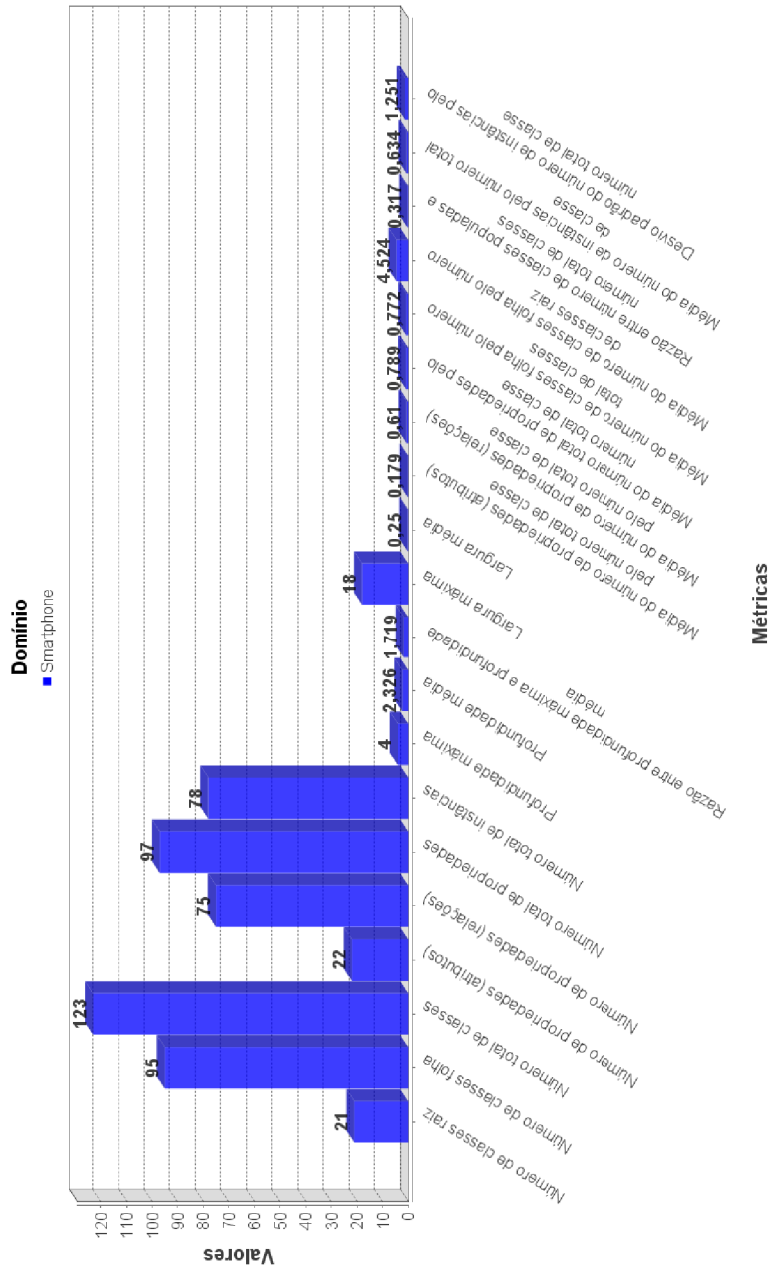


Figura B.8 – Ontologia de Smartphone.

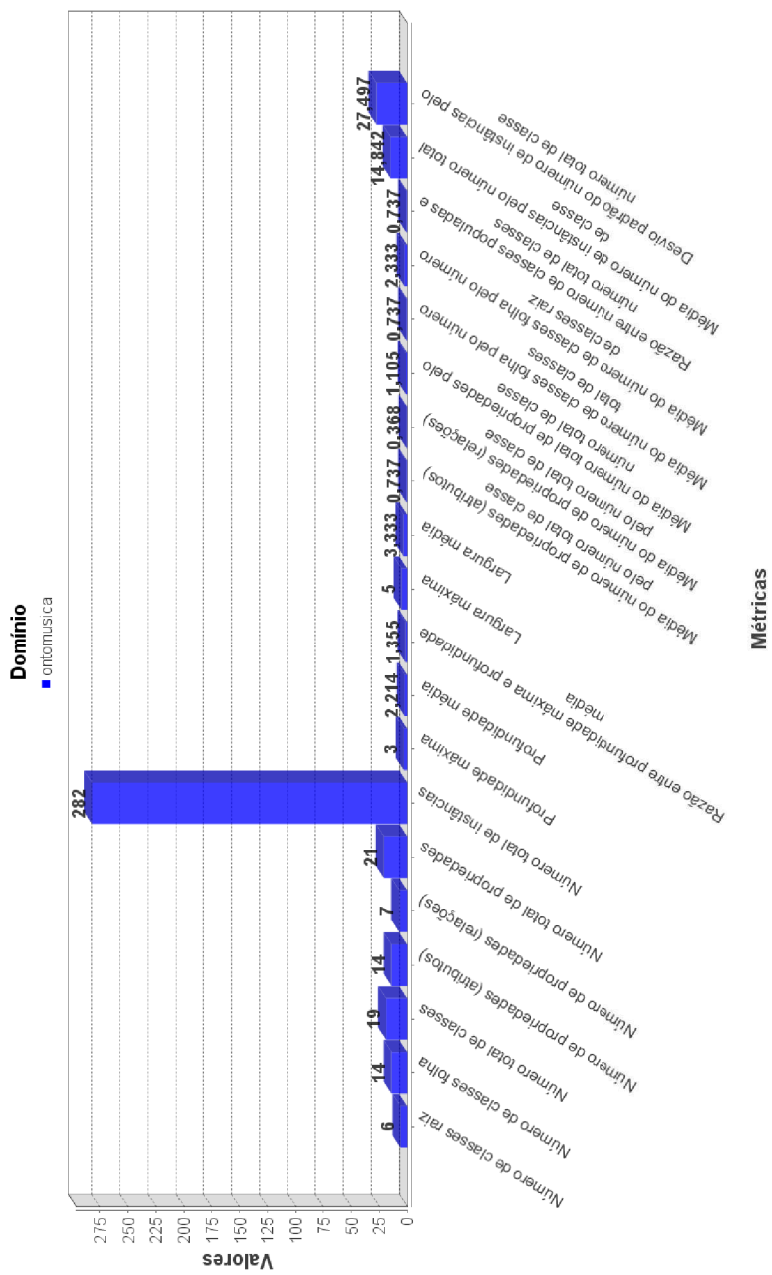


Figura B.9 – Ontologia de Música.

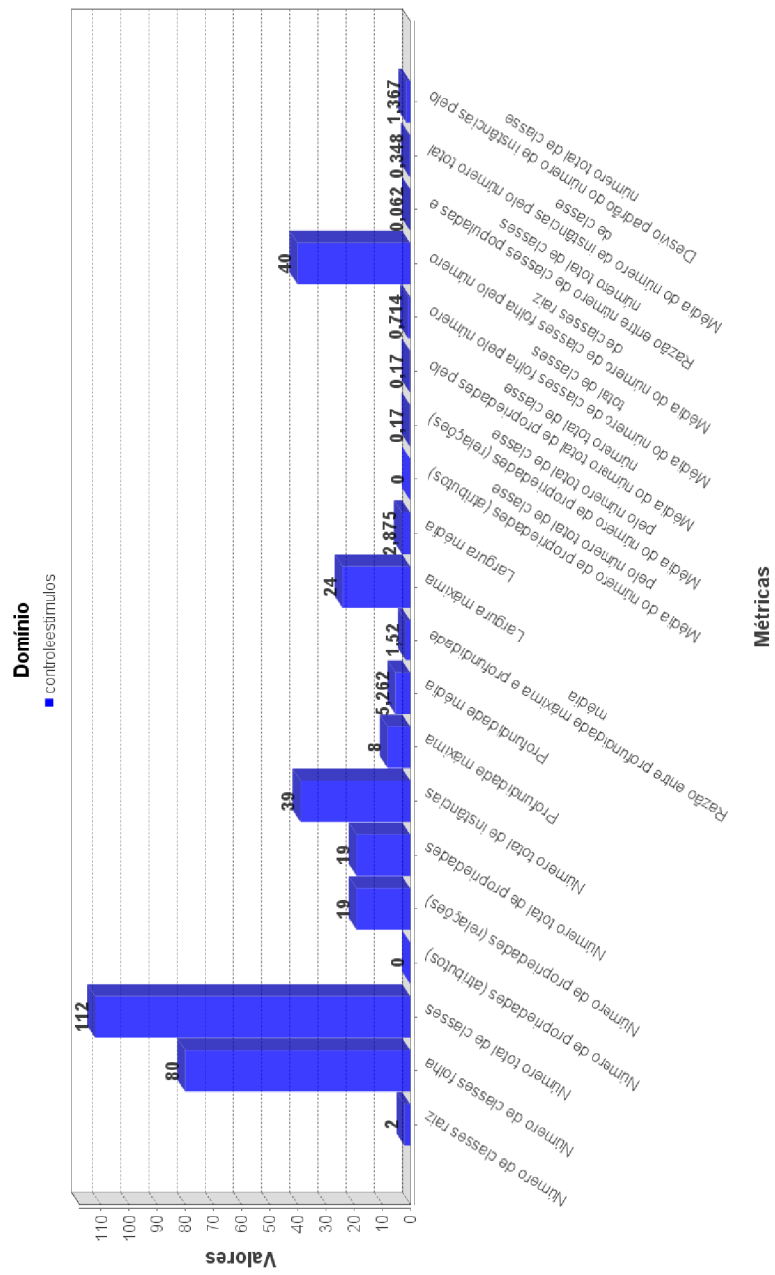


Figura B.10 – Ontologia de Equivalência de Estímulos (EE).

Tabela B.1 – Resultado da aplicação das métricas.

	Ecologia	Nanotecnologia	Arte	Currículo	Emoção	Privacidade	RCC	Smartphone	Música	EE
Número de classes raiz	3.0	6.0	15.0	8.0	3.0	7.0	1.0	21.0	6.0	2.0
Número de classes folha	41.0	259.0	140.0	152.0	39.0	37.0	5.0	95.0	14.0	80.0
Número total de classes	58.0	359.0	169.0	179.0	67.0	49.0	6.0	123.0	19.0	112.0
Número de propriedades (atributos)	14.0	1.0	10.0	144.0	15.0	0.0	0.0	22.0	14.0	0.0
Número de propriedades (relações)	2.0	4.0	16.0	65.0	0.0	19.0	5.0	75.0	7.0	19.0
Número total de propriedades	16.0	5.0	26.0	209.0	15.0	19.0	5.0	97.0	21.0	19.0
Número total de instâncias	60.0	0.0	151.0	0.0	0.0	27.0	0.0	78.0	282.0	39.0
Profundidade máxima	6.0	7.0	4.0	3.0	7.0	5.0	2.0	4.0	3.0	8.0
Profundidade média	3.732	3.475	3.079	2.428	6.872	3.676	2.0	2.326	2.214	5.262
Média da profundidade máxima pela profundidade média	1.608	2.014	1.299	1.236	1.019	1.360	1.0	1.719	1.355	1.520
Largura máxima	10.0	106.0	26.0	28.0	33.0	4.0	5.0	18.0	5.0	24.0
Largura média	2.0	0.571	5.5	0.333	5.0	0.4	3.0	0.25	3.333	2.875
Média do número de propriedades (atributos) pelo número total de classes	0.241	0.003	0.059	0.804	0.224	0.0	0.0	0.179	0.737	0.0
Média do número de propriedades (relações) pelo número total de classes	0.034	0.011	0.095	0.363	0.0	0.388	0.833	0.61	0.368	0.17
Média do número de propriedades pelo número total de classes	0.276	0.014	0.154	1.168	0.224	0.388	0.833	0.789	1.105	0.17
Média do número de classes folha pelo número total de classes	0.707	0.721	0.828	0.849	0.582	0.755	0.833	0.772	0.737	0.714
Média do número de classes folha pelo número de classes raiz	13.667	43.167	9.333	19.0	13.0	5.286	5.0	4.524	2.333	40.0
Média do número de classes populadas pelo número total de classes	0.017	0.0	0.29	0.0	0.0	0.204	0.0	0.317	0.737	0.062
Média do número total de instâncias pelo número total de classes	1.034	0.0	0.893	0.0	0.0	0.551	0.0	0.634	14.842	0.348
Desvio padrão do número de total instâncias pelo número total de classes	1.026	0.0	2.808	0.0	0.0	1.513	0.0	1.251	27.497	1.367

Apêndice C. Questionário Elaborado

Este questionário tem como objetivo verificar a opinião de um grupo de pesquisadores, engenheiros e usuários de ontologias sobre o uso de métricas para ontologias. Perguntamos se a informação dada pelas métricas é útil na inspeção das ontologias disponibilizadas em bibliotecas especializadas, como este portal. Estas métricas estão divididas nos seguintes grupos: **Grupo 1** contagem de elementos de uma ontologia (classes, propriedades e instâncias); **Grupo 2** profundidade e largura da estrutura; **Grupo 3** médias simples e desvio padrão.

O portal possui 10 ontologias em português descritas por 20 métricas. As métricas podem ser visualizadas através do link **métricas**. Mais informações sobre as mesmas estão disponíveis em **detalhes**.

Após visualização das métricas responda o questionário abaixo dando sua opinião sobre as métricas de uma maneira geral e em particular para cada grupo.

1. Dados pessoais

País:

Estado:

Instituição:

Grau de instrução:

2. Você já construiu ontologias?

Sim Não

3. Em qual(is) domínio(s)?

4. Você já buscou por ontologias na Web?

Sim Não

5. Em qual(is) idioma(s)?

Para as questões de 6 até 9 utilize os seguintes níveis de relevância: 1 não, discordo totalmente; 2 não, discordo; 3 neutro; 4 sim, concordo; 5 sim, concordo totalmente.

6. Você considera que as métricas, de uma maneira geral, são úteis na inspeção das ontologias disponibilizadas neste portal?

1 2 3 4 5

7. Você considera que as métricas do grupo 1 são úteis na inspeção das ontologias disponibilizadas neste portal?

1 2 3 4 5

8. Você considera que as métricas do grupo 2 são úteis na inspeção das ontologias disponibilizadas neste portal?

1 2 3 4 5

9. Você considera que as métricas do grupo 3 são úteis na inspeção das ontologias disponibilizadas neste portal?

1 2 3 4 5

10. Outras observações?

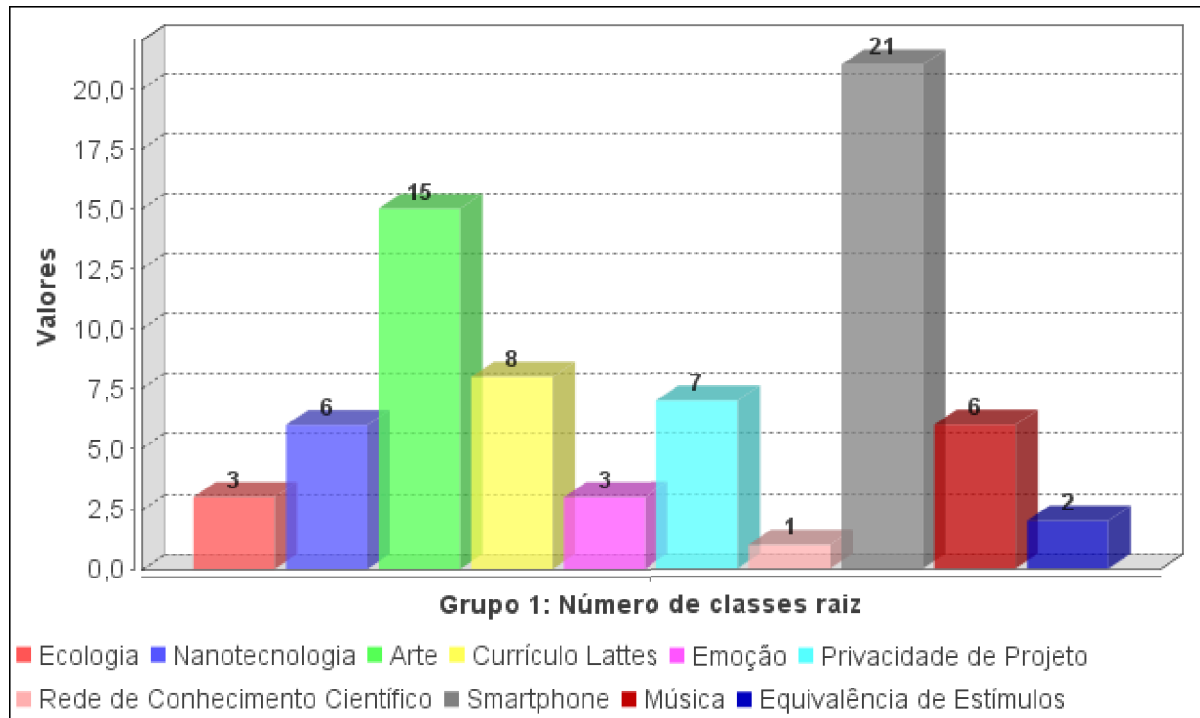


Figura C.1 – Número de classes raiz.

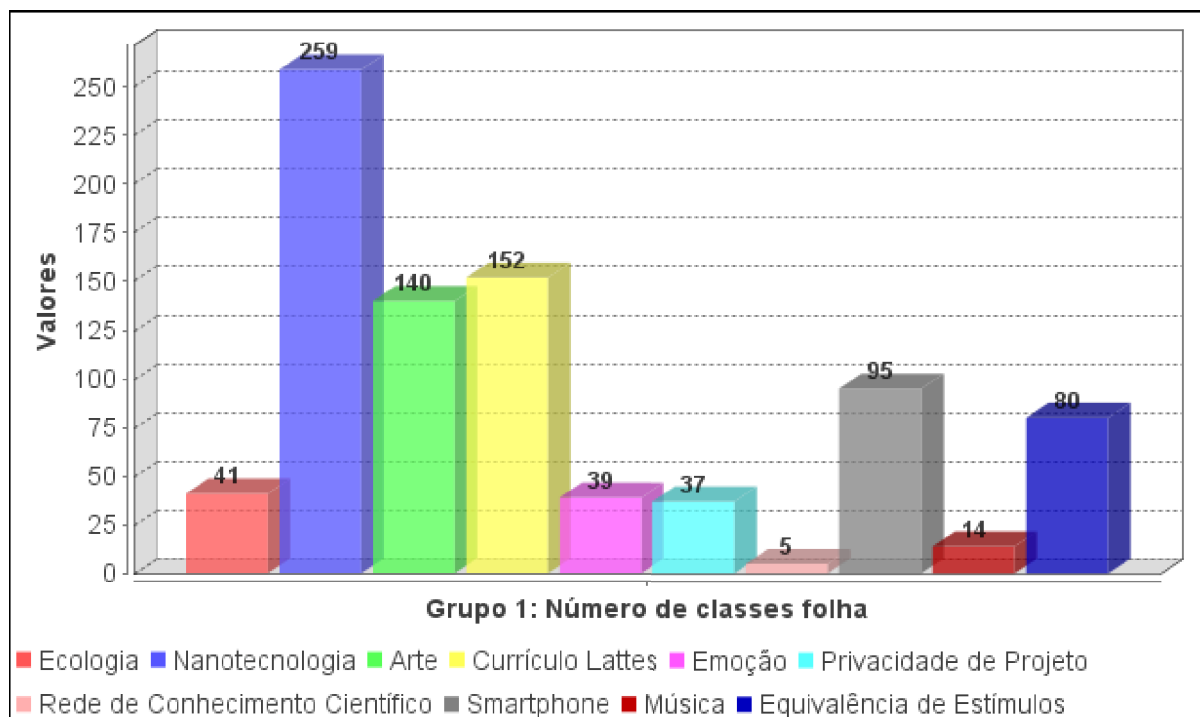


Figura C.2 – Número de classes folha.

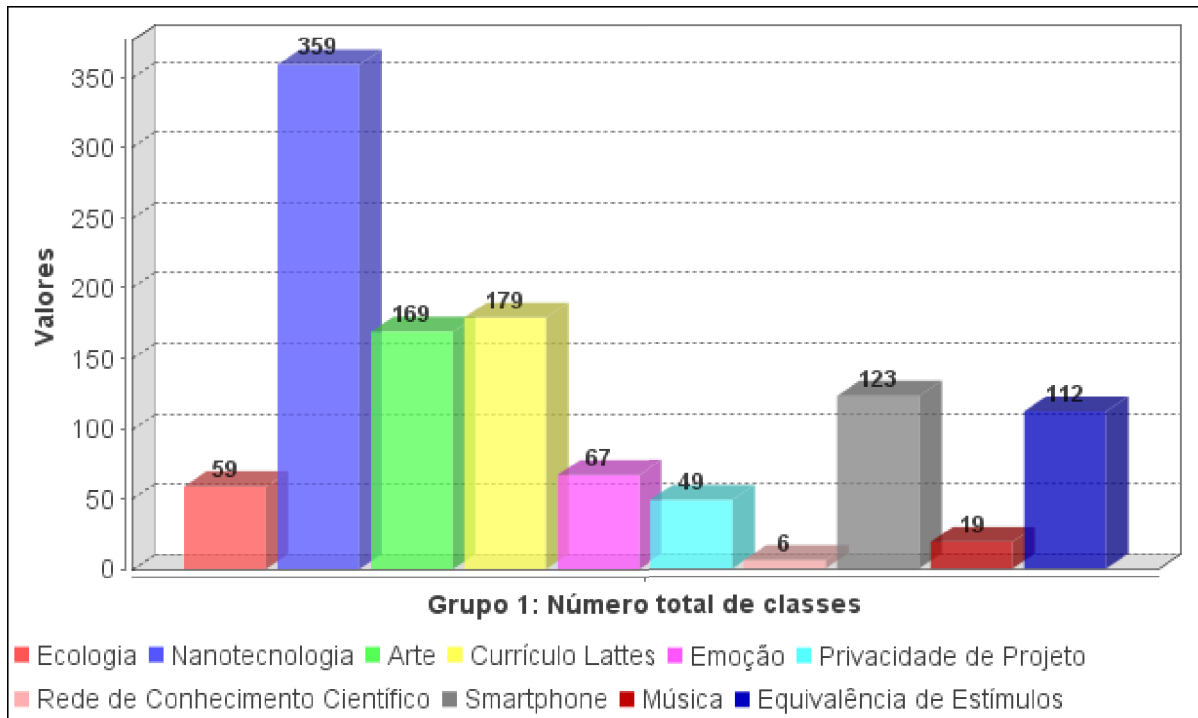


Figura C.3 – Número total de classes.

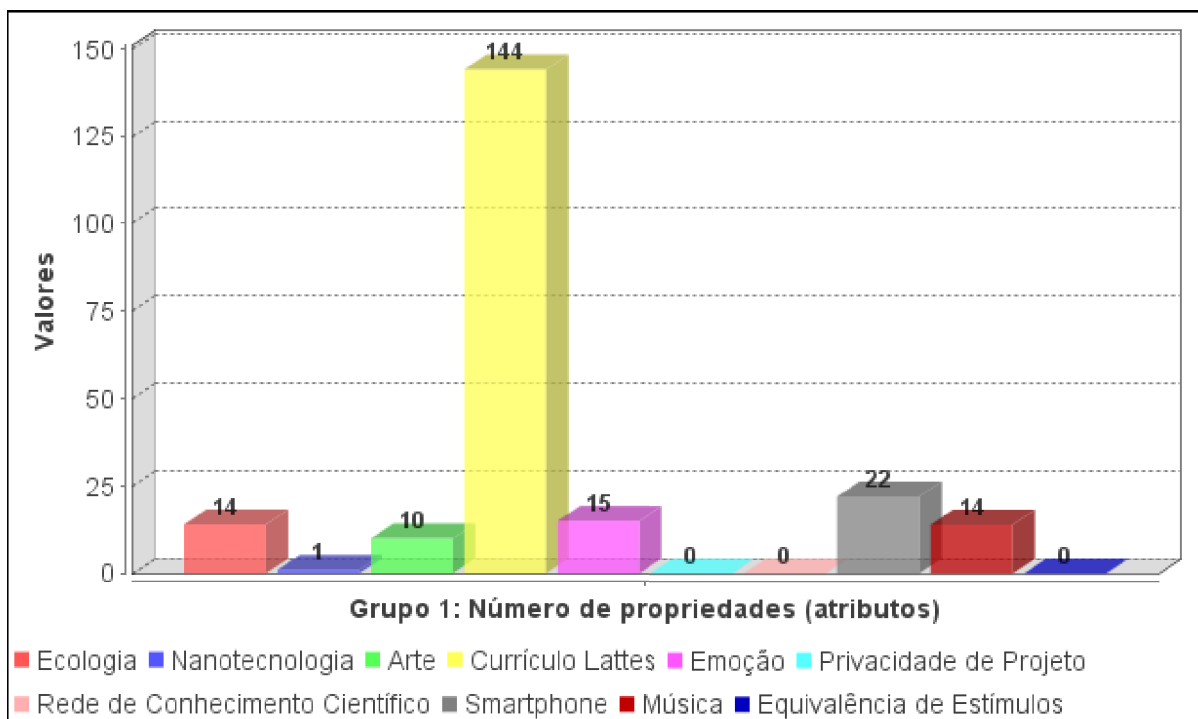


Figura C.4 – Número de propriedades (atributos).

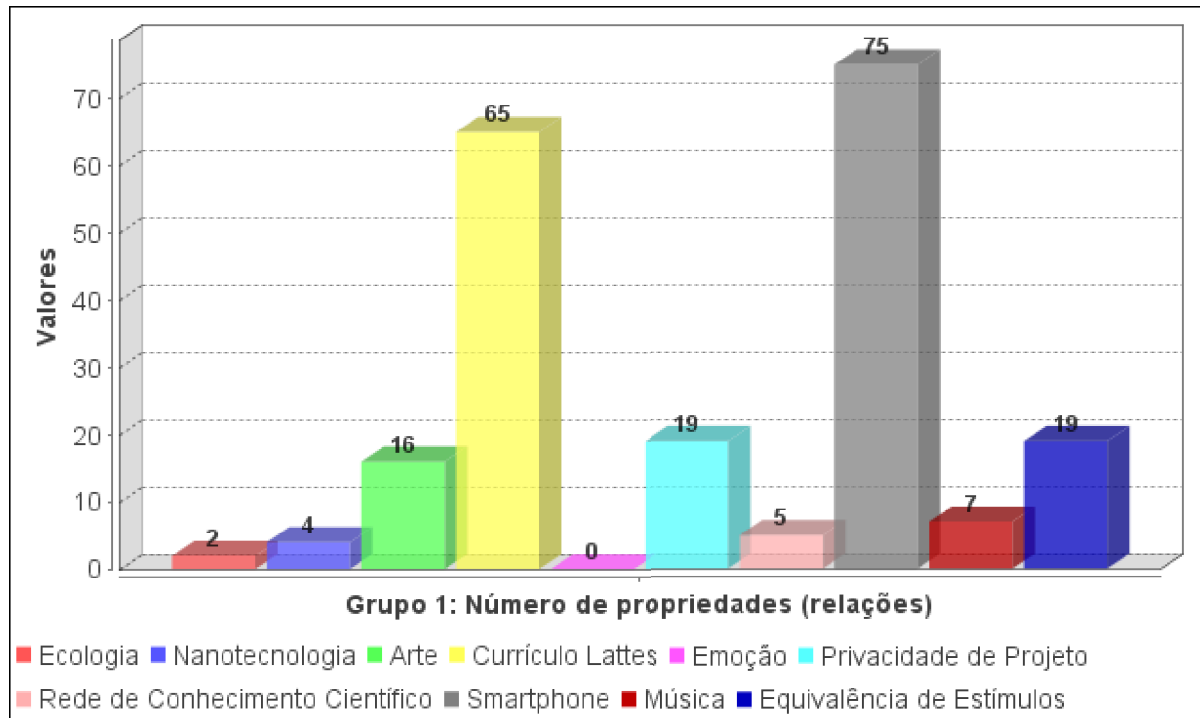


Figura C.5 – Número de propriedades (relações).

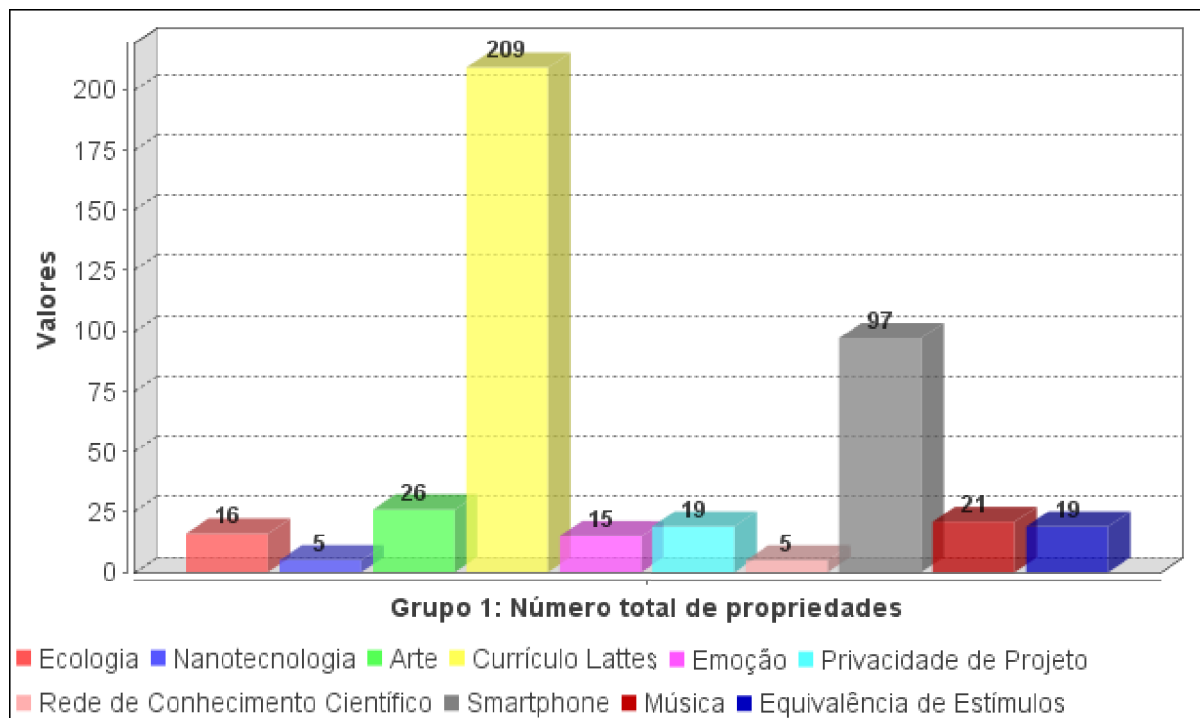


Figura C.6 – Número de propriedades.

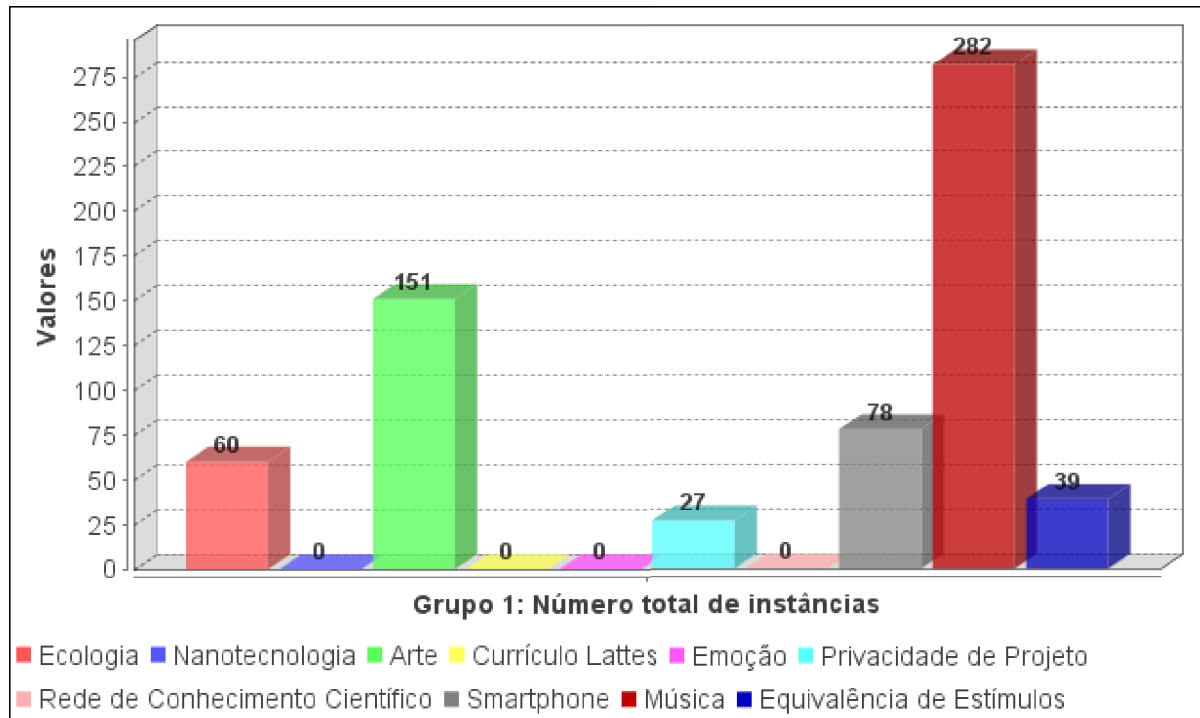


Figura C.7 – Número total de instâncias.

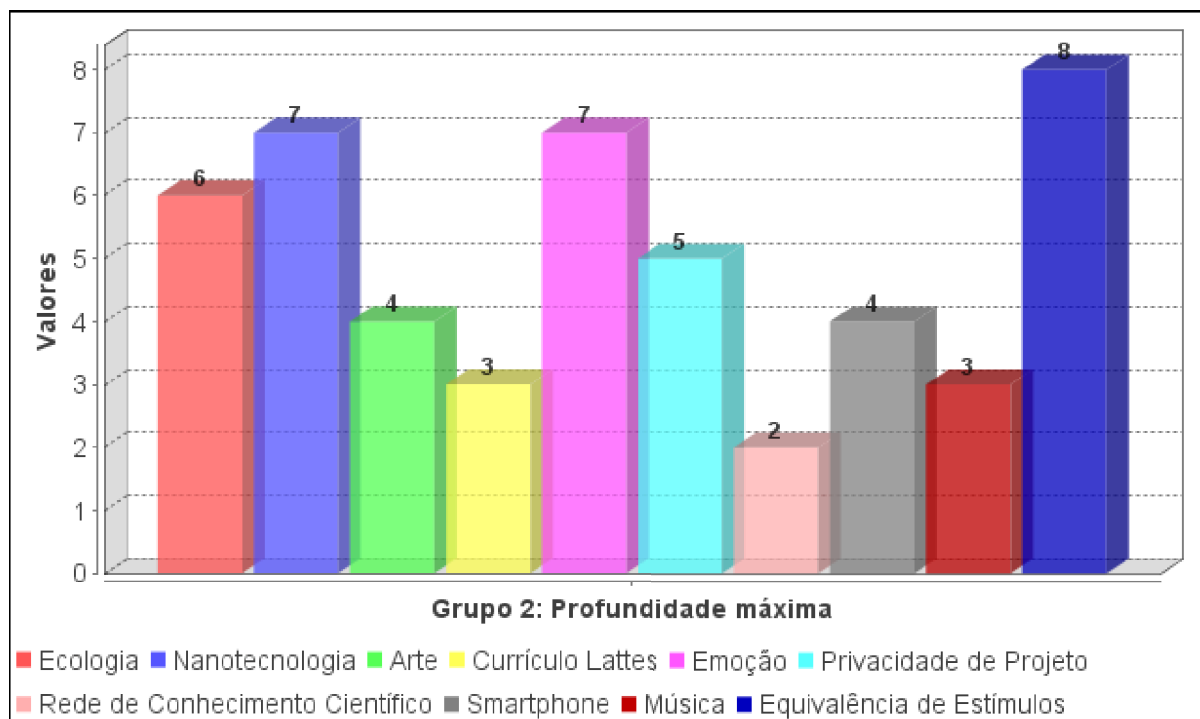


Figura C.8 – Profundidade máxima.

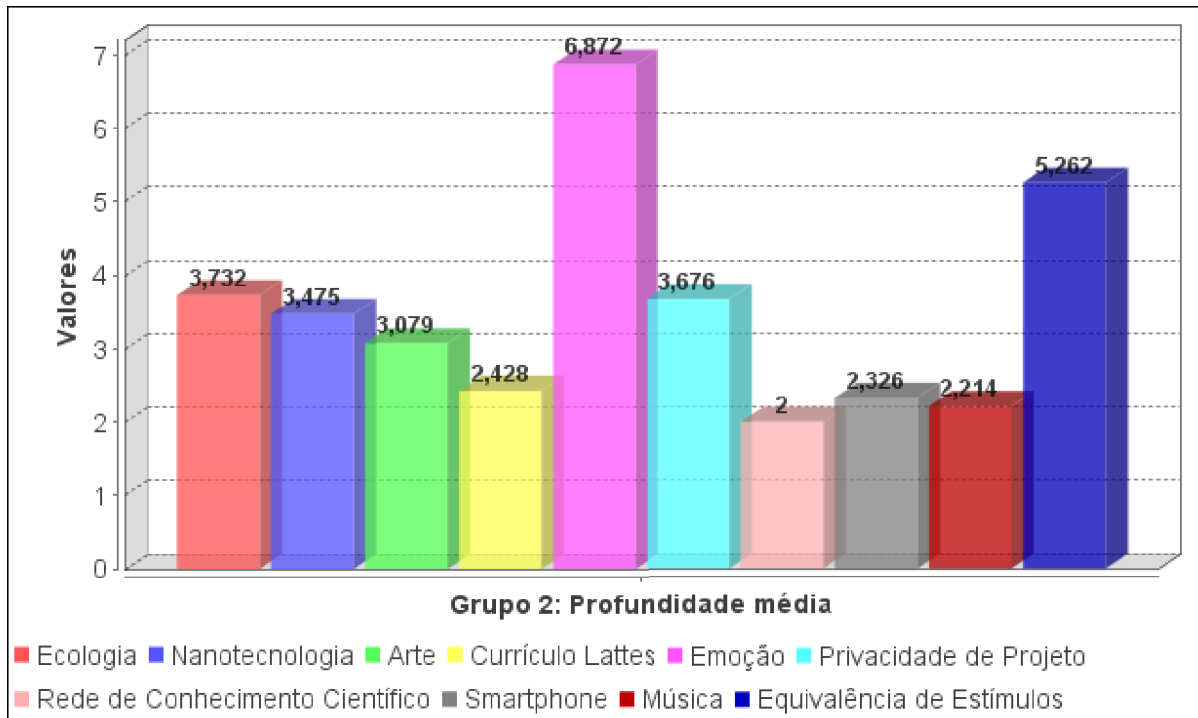


Figura C.9 – Profundidade média.

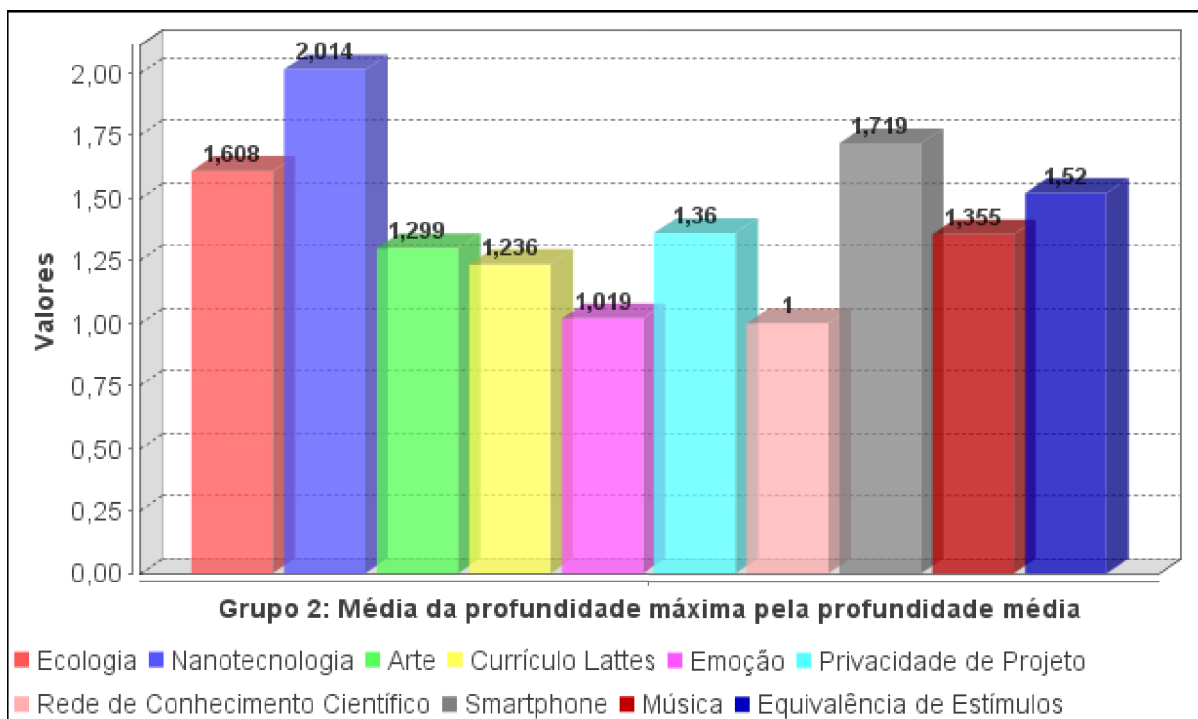


Figura C.10 – Média da profundidade máxima pela profundidade média.

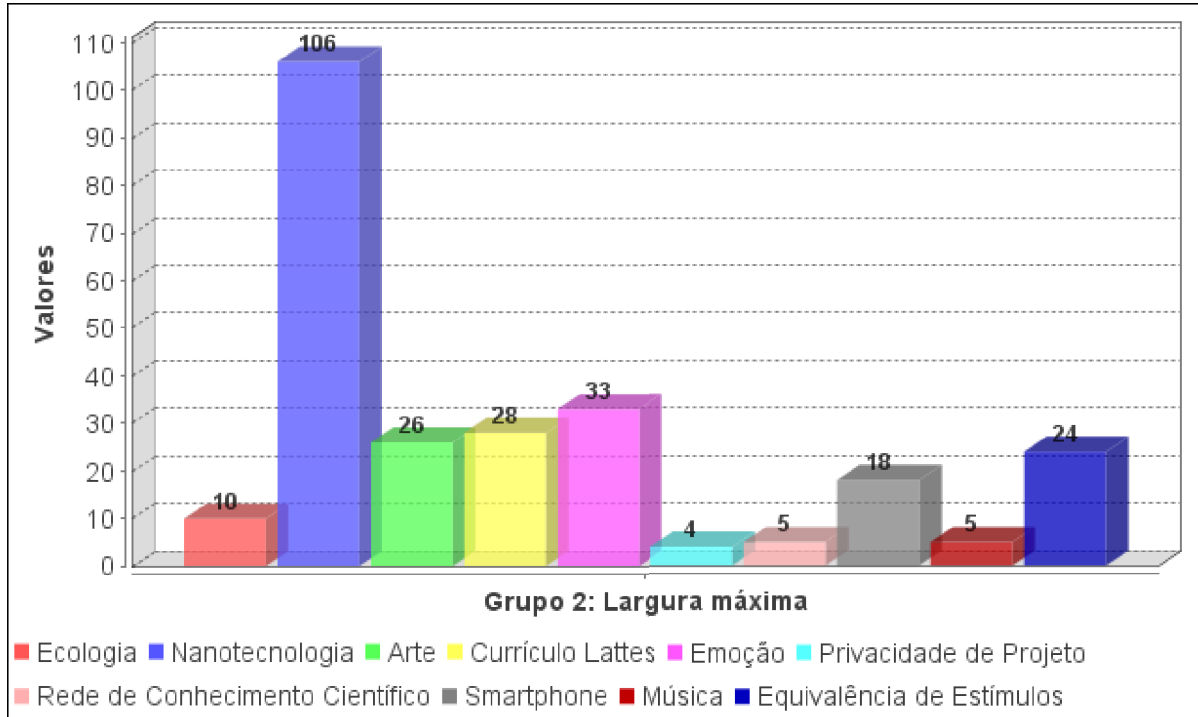


Figura C.11 – Largura máxima.

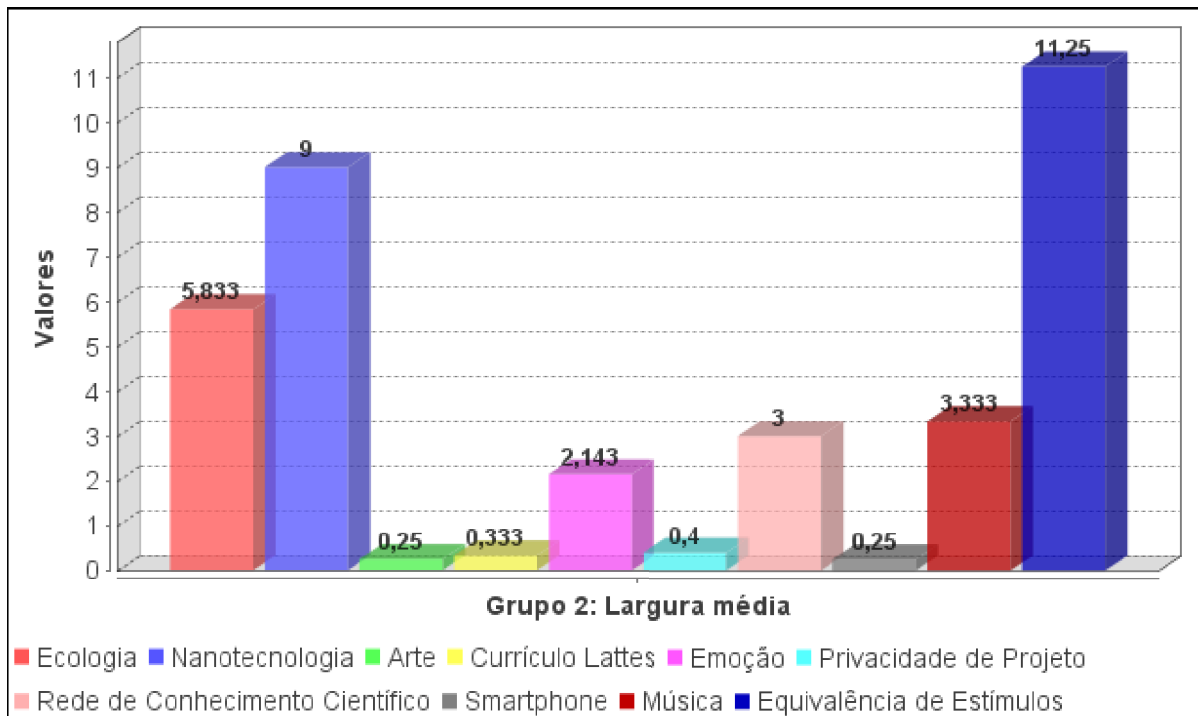


Figura C.12 – Largura média.

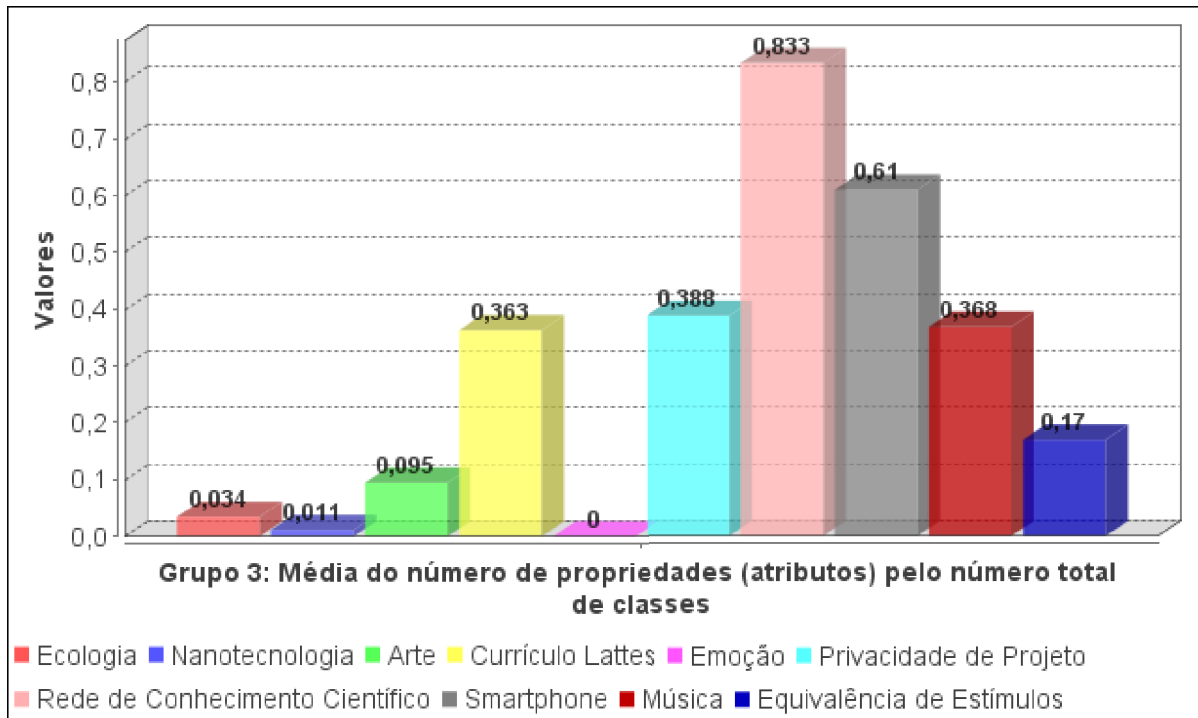


Figura C.13 – Média do número de propriedades (atributos) pelo número total de classes.

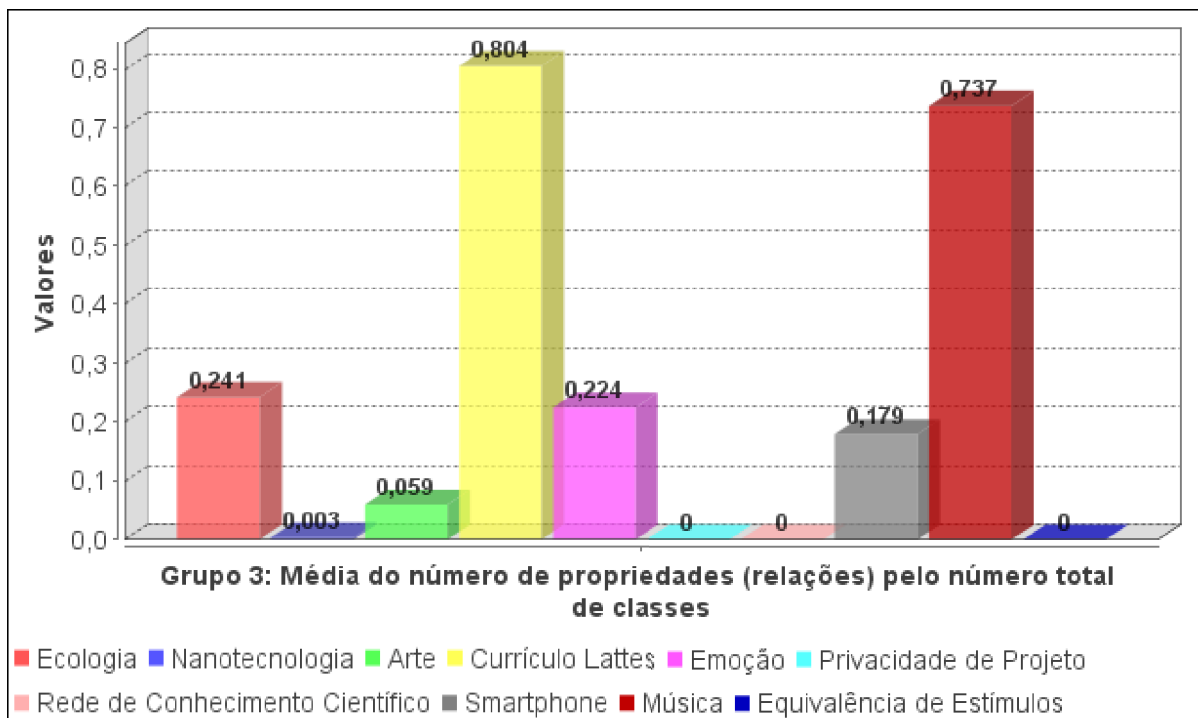


Figura C.14 – Média do número de propriedades (relações) pelo número total de classes.

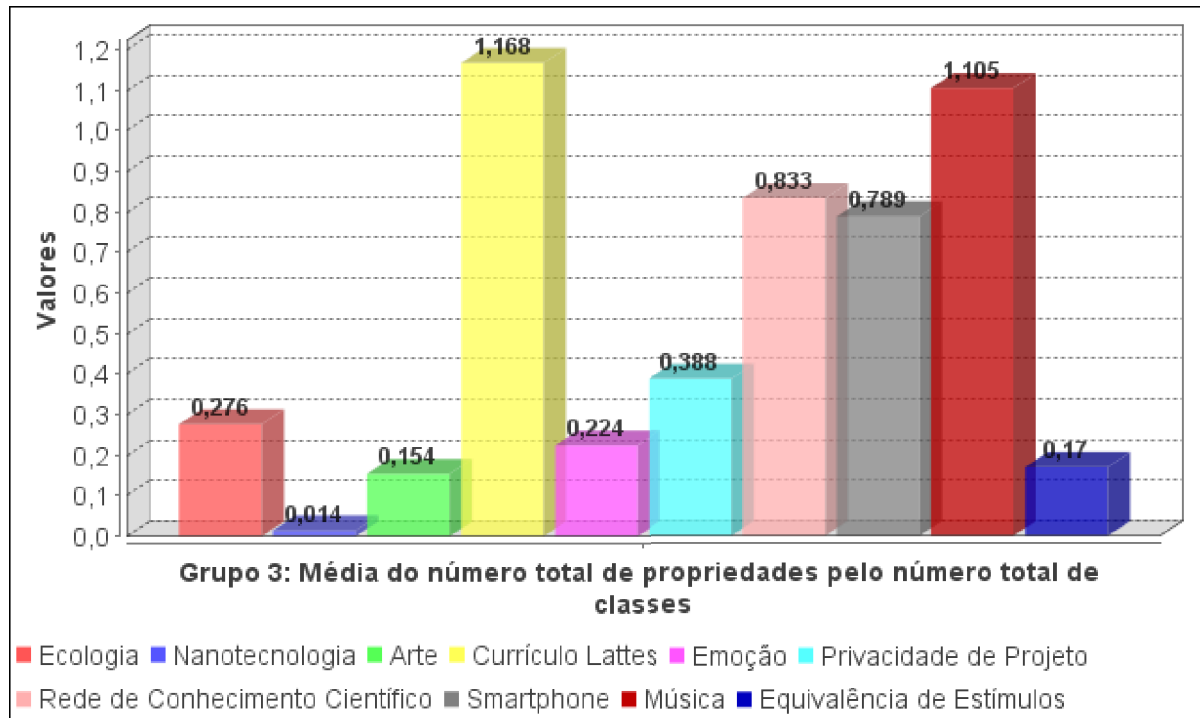


Figura C.15 – Média do número de propriedades pelo número total de classes.

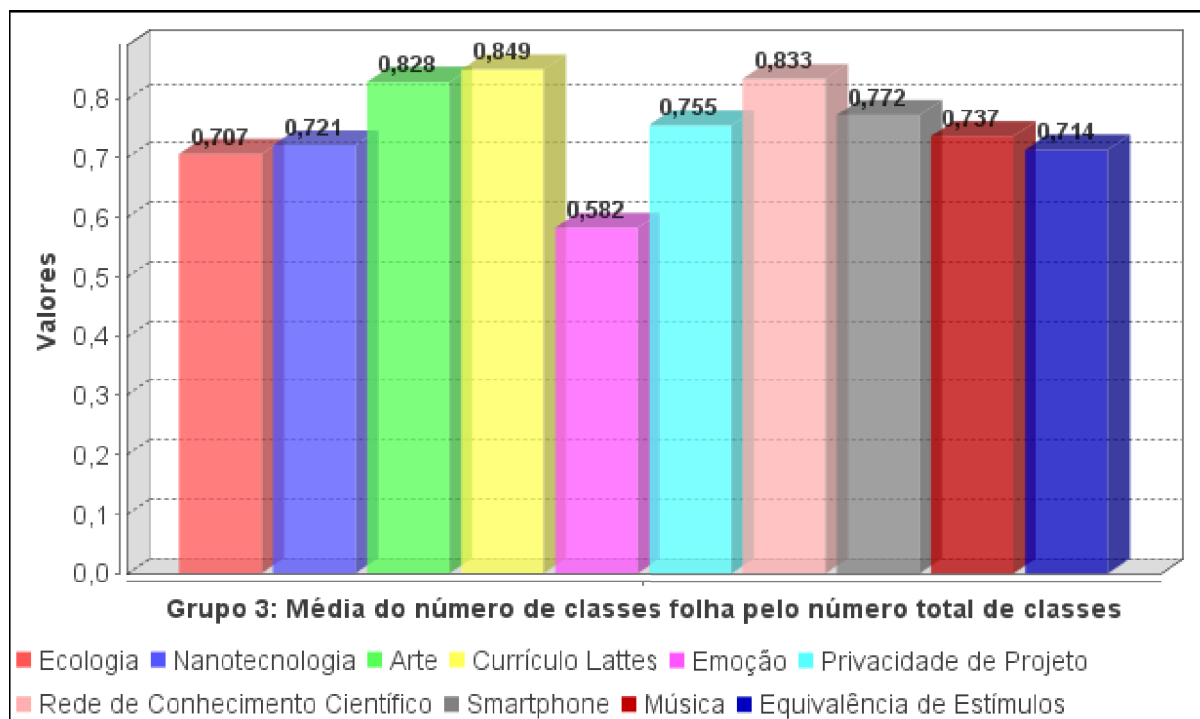


Figura C.16 – Média do número de classes folha pelo número total de classes.

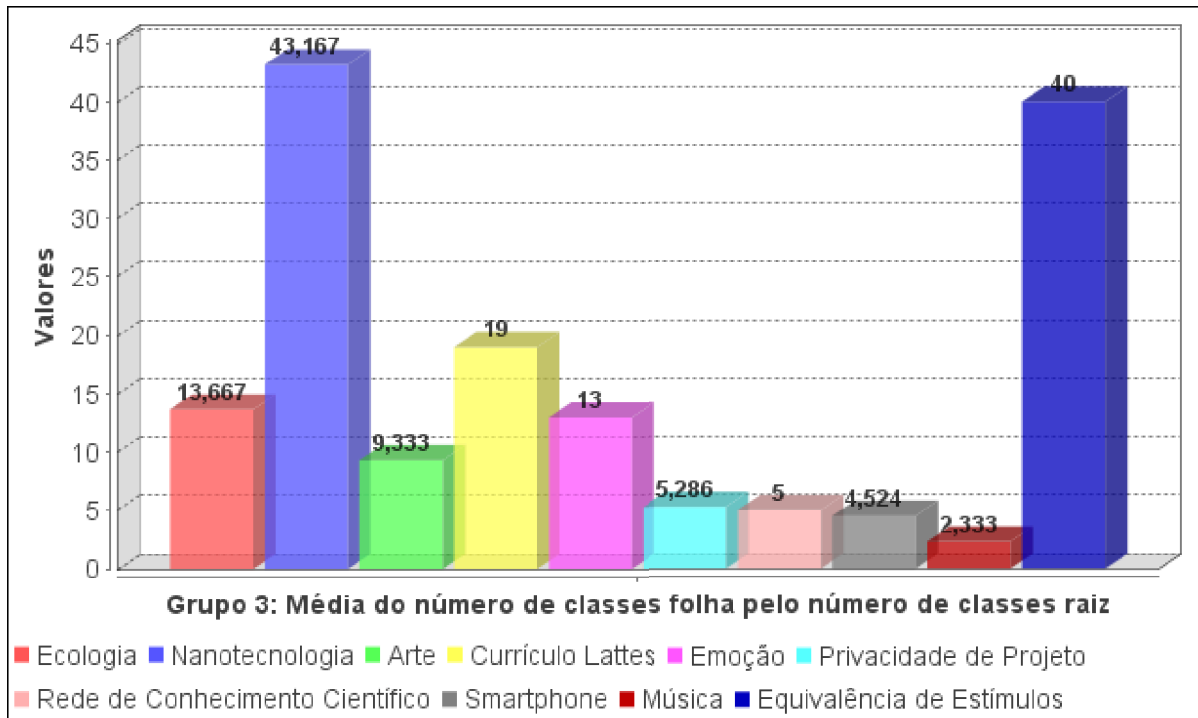


Figura C.17 – Média do número de classes folha pelo número de classes raiz.

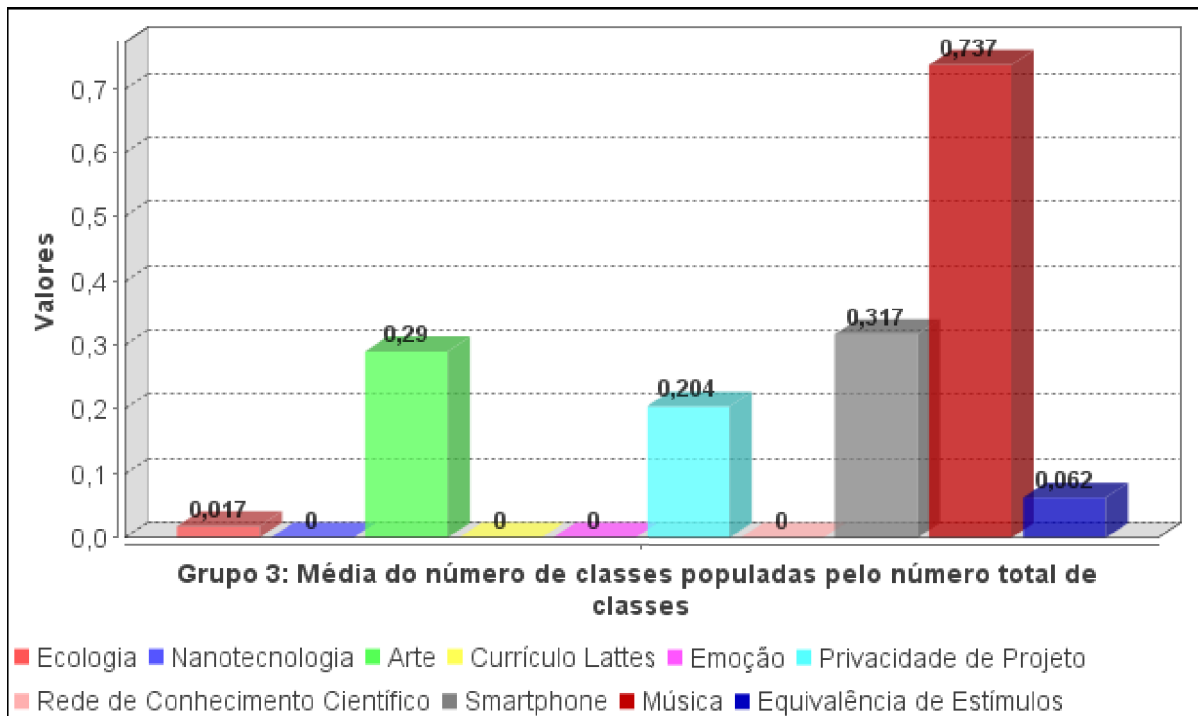


Figura C.18 – Média do número de classes populadas pelo número total de classes.

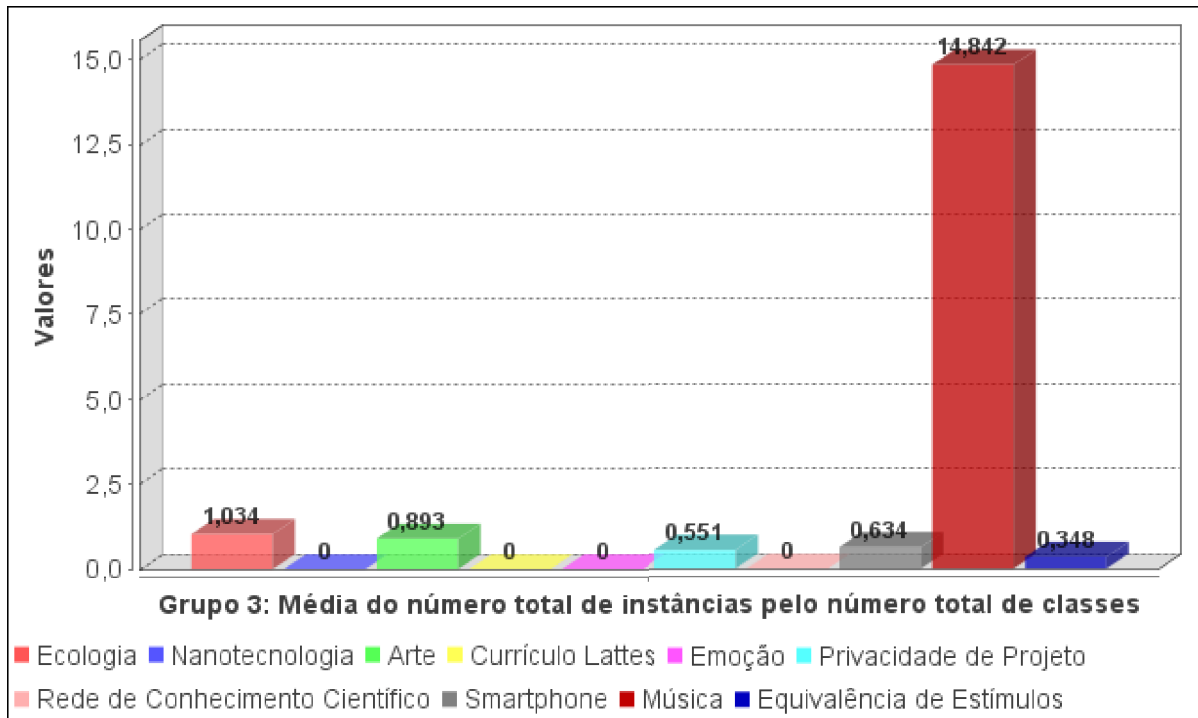


Figura C.19 – Média do número total de instâncias pelo número total de classes.

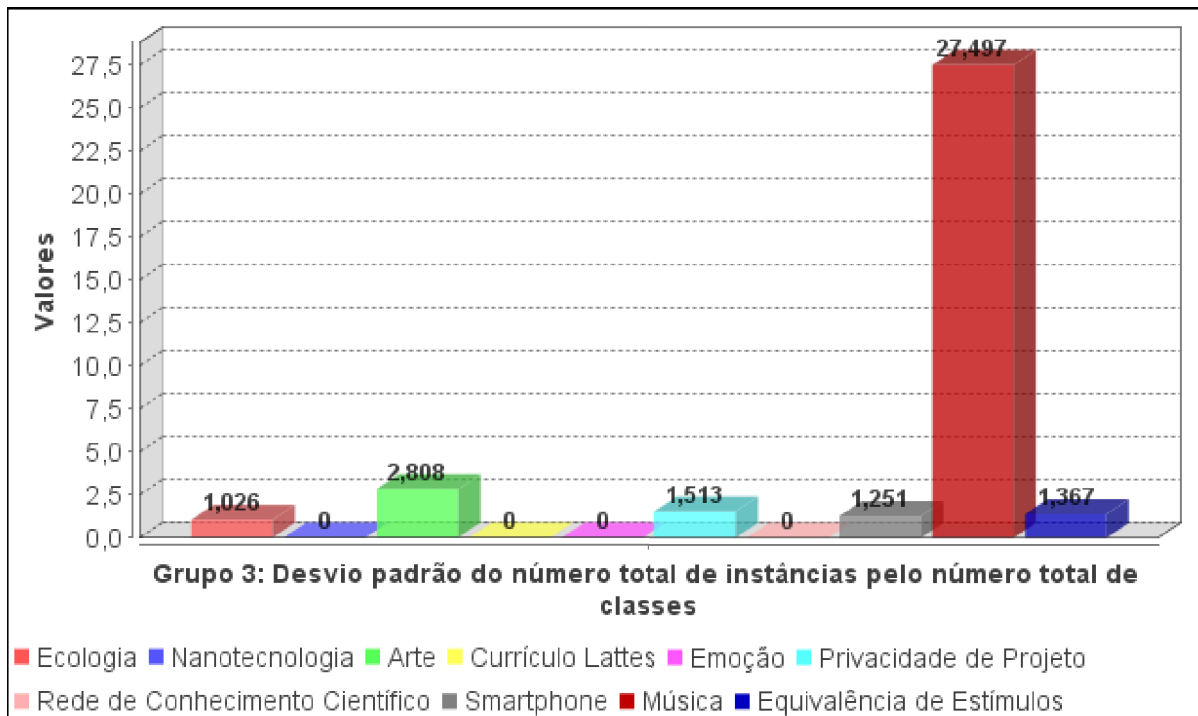


Figura C.20 – Desvio padrão do número de total instâncias pelo número total de classes.

Grupo 1

- Número de classes raiz: Indica o número de classes que não possuem superclasses.
- Número de classes folha: Indica o número de classes que não possuem subclasses.
- Número total de classes: Indica o número total de classes definidas. O termo conceito também é utilizado no lugar de classes.
- Número de propriedades (atributos): Indica o número de propriedades (atributos), denominado de DatatypeProperty em OWL. Representa as relações entre instâncias e tipos.
- Número de propriedades (relações): Indica o número de propriedades (relações), denominado de ObjectProperty em OWL. Representa as relações entre classes.
- Número total de propriedades: Indica o número total de propriedades definidas. Somatório de atributos e relações.
- Número total de instâncias: Indica o número total de instâncias definidas. As instâncias representam objetos no domínio de interesse. Instâncias também são conhecidas como indivíduos.

Grupo 2

- Profundidade máxima: Indica a profundidade máxima, representa o caminho mais longo na hierarquia.
- Profundidade média: Indica a profundidade média, representa a razão entre o somatório da profundidade de todos os caminhos e o número total de caminhos.
- Média da profundidade máxima pela profundidade média: Indica a agregação das classes e a coesão da ontologia.
- Largura máxima: Indica a largura máxima, representa o maior número de classes vizinhas, nos níveis da ontologia.
- Largura média: Indica a largura média, representa a razão entre o somatório da largura de todas as classes com subclasses e o número total de classes com subclasses.

Grupo 3

- Média do número de propriedades (atributos) pelo número total de classes: Indica a média simples entre propriedades (atributos) pelo número total de classes.
- Média do número de propriedades (relações) pelo número total de classes: Indica a média simples entre propriedades (relações) pelo número total de classes.
- Média do número de propriedades pelo número total de classes: Indica o grau de conectividade médio das classes.
- Média do número de classes folha pelo número total de classes: Indica a média simples entre o número de classes folha pelo número total de classes.
- Média do número de classes folha pelo número de classes raiz: Indica a média simples entre o número de classes folha pelo número de classes raiz.
- Média do número de classes populadas pelo número total de classes: Indica a riqueza da base de conhecimento.
- Média do número total de instâncias pelo número total de classes: Indica a média simples entre o número de instâncias pelo número total de classes.
- Desvio padrão do número total de instâncias pelo número total de classes: Indica como as instâncias estão espalhadas entre as classes de uma estrutura.

Apêndice D. Resultado da Aplicação do Questionário

UF	Instrução	Construção	Domínio	Busca	Idioma	Todas	Grupo 1	Grupo 2	Grupo 3
SP	doutorado	Sim	Permacultura, Comunicação	Sim	Português, Inglês	5	4	4	4
PE	doutorado	Sim		Sim	Português, Inglês, Alemão	5	3,4	1,2	5
RS	graduação	Sim	Privacidade	Sim	Inglês	5	5	3	3
RS	doutorado	Sim	Geologia	Sim	Inglês	4	1		3
SP	doutorado	Não		Sim	Inglês	5	2	3	4
PR	doutorado	Sim	Desenvolvimento de Projeto, Emoções	Sim	Português, Inglês	2	3	2	2
RS	especialização	Sim		Sim	Português, Inglês	5	3	2	2
SP	outro		Agronegócios	Sim	Português	4	4	3	5
MA	doutorado	Sim	Direito, Turismo	Sim	Inglês	4	4	4	4
	doutorado	Sim	Medicina	Sim	Português, Inglês, Espanhol	4	2	3	3
RS	mestrado	Sim	Localização, Trilhas, Educação, Privacidade	Sim	Português, Inglês	3	4	2	3
SP	mestrado	Não		Sim	Português	3	3	2	3
AM	mestrado	Sim	Teste de Software, Sistemas Operacionais	Sim	Inglês	5	4	4	3
MA	mestrado	Sim	Jurídico	Sim	Português, Inglês	3	3	2	4
SP	mestrado	Não		Sim	Português	3	4	2	3
PR	graduação	Sim	Direito	Sim	Português	4	3	3	4
RS	graduação	Sim	Pizza, Pessoa	Sim	Português, Inglês, Espanhol	4	5	3	5

Tabela D.1 – Resultado questionário (pré-teste).

Tabela D.2 – Resultado questionário (final).

UF	Instituição	Instrução	Construção	Domínio	Busca	Idioma	Todas	Grupo 1	Grupo 2	Grupo 3
MG	UFMG	doutorado	Sim	KOS, Arqueologia, Indústria petroléira	Sim	Inglês, Português	5	5	4	5
RJ	UNIRIO	doutorado	Sim	Processos de negócios e contexto, Sistemas colaborativos	Não		5	5	5	5
	Universidade Atlântica	doutorado	Sim	Universidade e Geográfico	Sim	Português, Inglês	5	5	5	5
SP	USP	doutorado	Não	Já trabalhei na extensão de uma ontologia da área de biologia (genética).	Sim	Inglês.	4	4	4	4
SP	USP	mestrado	Sim	Arte Contemporânea	Sim	Português, Inglês, Espanhol	4	5	4	2
RS	PUCRS	mestrado	Sim	localização e aprendizado	Sim	Inglês, Português	5	5	4	2
MG	UFJF	doutorado	Sim	Educação a Distância, Software Agropecuário, Células e Artefatos Científicos	Sim	Português, Inglês	4	5	2	2
RS	PUCRS	graduacao	Sim	Privacidade, Segurança e Acadêmico	Sim	Inglês, Português	5	5	4	4
SC	UFSC	doutorado	Sim	Transmissão de Energia elétrica, saúde, direito, etc	Sim	Inglês, Português	5	5	5	2
RJ		doutorado	Sim	Domínio de educação	Sim	Português	5	5	5	4
SP	Embrapa	doutorado	Sim	Ciência da Computação	Sim	Português, Inglês, Espanhol	4	3	4	3
RS	UFRGS, UniRitter, Ulbra	mestrado	Sim	Urbanismo	Sim	Inglês	4	4	3	3
RJ	PUC-Rio	mestrado	Não		Não		3	2	3	3
RS	UFRGS	mestrado	Sim	Geologia/Estratigrafia	Sim	Inglês	4	4	4	3
SP	USP	doutorado	Não		Sim	Português	4	3	5	3
SP	USP	graduacao	Não		Sim	Inglês	4	3	4	4
RS	PUCRS	especializacao	Sim	Turismo	Sim	Português, Inglês	4	4	4	4
MG	UFMG	doutorado	Sim	Organizações, medicina, plantas medicinais	Sim	Inglês	5	4	4	4
RJ	UFRJ	mestrado	Sim	serviços web (no escopo de uma dissertação de mestrado)	Sim	Inglês	3	3	3	3
	INRIA	doutorado	Sim	Bibliografico	Sim	Português, Inglês	5	5	5	5
RJ	IME	doutorado	Sim	Educação	Sim	Inglês, Português	5	5	5	4
RJ	UFF	graduacao	Não		Não		3	3	3	3
RJ	UFRJ	mestrado	Não		Sim	Inglês	4	5	5	4
RJ	UFRJ	especializacao	Não		Não		5	3	3	4
AM	FPF	mestrado	Sim	Teste de Software, Sistemas Operacionais e a interseção entre esses dois domínios.	Sim	Inglês, Português	5	5	5	5
RJ	PUC-Rio	doutorado	Sim	saúde, informática	Sim	Português, Inglês	5	5	5	5
RS	Feevale	mestrado	Sim	Smartphones e música.	Sim	Inglês, Português	3	3	2	4
RJ	UFRJ	mestrado	Sim	Previdência Social Brasileira	Sim	Português, Inglês	3	4	4	4
RS	PUCRS	mestrado	Sim	Categorias de entidades nomeadas	Não		5	5	4	4
PR	UTFPR	doutorado	Sim	Ontologias para realização de experimentos na área de turismo e de tarefas realizadas em computadores	Sim	Inglês	5	5	4	3