



# Construindo redes de ontologias heterogêneas

*Luma Oliveira Lombello*

*Julio Cesar dos Reis*

Technical Report - IC-19-06 - Relatório Técnico  
September - 2019 - Setembro

UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.  
O conteúdo deste relatório é de única responsabilidade dos autores.

# Construindo redes de ontologias heterogêneas

Luma Oliveira Lombello  
Julio Cesar dos Reis

Instituto de Computação, Universidade Estadual de Campinas, Campinas, SP, Brasil.

2019

## Resumo

Ontologias são artefatos para a representação de conhecimento, nos quais conceitos e suas inter-relações são descritos dentro de um domínio. Redes de ontologias são caracterizadas pela existência de ontologias heterogêneas em um mesmo espaço conceitual. Uma ontologia formal ou estruturada é uma proposta de representação determinada e geralmente fixada *a priori*, já uma ontologia “leve” é flexível, podendo ser mais suscetível em sofrer mudanças nas inter-relações ao longo do tempo. A conexão entre diferentes modelos de ontologias é relevante para a integração de dados e a troca de informação entre sistemas. Embora existam técnicas para relacionar automaticamente ontologias criando correspondências entre termos de significados similares, um modelo que faça a interconexão entre ontologias heterogêneas ainda é um tópico de pesquisa em aberto. Este relatório descreve resultados que permitem conceitos especificados em ontologias de modelos distintos serem relacionados um com o outro. Estudamos como ontologias formais e bem estruturadas da *Web Semântica* podem ser conciliadas com a proposta de ontologias “leves” menos estruturadas. Obtivemos um metamodelo em eBNF que permite conciliar ontologias estruturadas descritas em OWL com ontologias “leves” descritas por triplas em RDF *Fuzzy*.

## 1 Introdução

Existem muitas definições para o conceito de ontologia, desde filosóficas até técnicas. Porém, é possível englobar seus muitos significados ao dizer que ontologia é uma forma de representar conhecimento, por meio de um vocabulário que descreve objetos e suas relações [5] em um domínio. Dessa forma, a computação adota ontologias para o uso em *Web Semântica*, bancos de dados, processamento de linguagens naturais, engenharia de *software* e inteligência artificial [17].

Na computação, a categorização de dados é essencial, tanto para que exista uma relação de compreensão entre humanos e máquinas, quanto para que haja um maior empoderamento do usuário via a interação com sistemas interativos. Nesse contexto, ontologias estruturadas expressam relações formais entre conceitos e podem apoiar diálogo entre computadores e pessoas. Mais especificamente, ontologias são essenciais para a descrição formal da semântica de dados desenvolvendo uma papel essencial em sistemas de recuperação e integração de informação.

Por outro lado, ontologias “leves” visam representar o conhecimento de maneira mais flexível, de forma que usuários possam expressar conceitos sem a preocupação em desenvolver relações formais entre eles, ou usar uma linguagem de descrição de ontologias. Entretanto, há ainda a necessidade de interconectar ambos tipos de ontologias, visando uma melhor interoperabilidade entre essas propostas de representação do conhecimento.

Ontologias “leves” podem ser definidas como uma forma de categorização centrada no usuário, ou seja, como um sistema de organização baseado em marcadores e palavras-chave. A diferença entre esse modelo e outros sistemas de organização do conhecimento já existentes, como ontologias estruturadas e formais da *Web Semântica*, consiste na ausência de um vocabulário específico e bem delimitado para a aplicação de *software*. Raramente, a proposta de uso de *tags* possibilita a criação de relações hierárquicas bem definidas entre as palavras-chave [15].

Nessa perspectiva, um processo de *matching*, que visa detectar correspondência entre conceitos provenientes de diferentes ontologias, se torna ainda mais desafiador no contexto de ontologias heterogêneas. O mapeamento de termos de significados relacionados é dificultado, uma vez que as técnicas atuais de correspondência entre cadeias de caracteres não estão adaptadas à informalidade de termos definidos por usuários classificando objetos. A flexibilidade de um vocabulário não restrito permitido pelas ontologias “leves” traz grandes desafios para a conciliação de modelos.

Apesar disso, a interligação entre diferentes modelos de ontologias é primordial em um cenário de tecnologias como extensão dos indivíduos [11]. Em diversos contextos, há a urgência de uma ontologia “leve”, que possa acompanhar a dinamicidade dos comportamentos emergentes dos indivíduos e se reestruturar conforme as necessidades do contexto de interação, por exemplo. Ao mesmo tempo que para ocorrer transferência de dados entre sistemas de forma significativa e com raciocínio lógico automatizado é preciso uma ontologia relativamente estável, com uma estrutura hierárquica fixa, além de relações de domínio bem definidas.

Devido ao uso de *tagging*, a literatura na área de interoperabilidade entre ontologias apresenta um grande espaço de possibilidades de pesquisa, já que as técnicas existentes até então não abrangem a flexibilidade permitida pelos novos meios de sistemas de categorização.

Este relatório apresenta conceitos e técnicas utilizadas no desenvolvimento de ontologias “leves”. Nossa investigação estuda técnicas para conciliá-las com entida-

des ontológicas estruturadas. Exploramos a transformação de dados em uma matriz probabilística (representando espaços ontológicos como ontologias “leves”) para um modelo de dados em RDF *fuzzy* (*Resource Description Framework*). Alcançamos o objetivo de representar ontologias “leves” por meio de um processo de triplificação RDF *fuzzy*. Adicionalmente, investigamos técnicas para se conciliar ontologias heterogêneas, via um metamodelo formalizado pela linguagem eBNF, que permite o relacionamento entre sistemas de organização do conhecimento rígidos e estruturados com sistemas de organização de conhecimento “leves” e versáteis. Viabilizamos um *Ontospace* em que é possível navegar e recuperar dados. Mais especificamente, reportamos os seguintes resultados:

- Modelagem de uma ontologia “leve” por meio do desenvolvimento de uma matriz probabilística representando espaços ontológicos, e sua transposição para triplas RDF *fuzzy*;
- Desenvolvimento de um metamodelo em eBNF capaz de suportar a interconexão de dados provindos de ontologias estruturadas em OWL e ontologias “leves” representadas em *fuzzy* RDF;
- Aplicação do nosso processo em um ambiente educacional para crianças, no qual o cenário foi apoiado por um robô mBot [3], programado para executar um conjunto de ações que imitam expressões emocionais humanas;

O restante deste relatório está organizado da seguinte forma: na seção 2 apresenta trabalhos relacionados; a seção 3 descreve a proposta de rede de ontologias via metamodelos; a seção 4 relata a avaliação conduzida enquanto a seção 5 discute os resultados; a seção 6 descreve as conclusões e os trabalhos futuros.

## 2 Revisão da Literatura

A literatura apresenta diversas técnicas de alinhamento entre ontologias estruturadas, entretanto ainda há vários desafios quando se trata de técnicas automáticas de correspondência entre ontologias heterogêneas.

Os primeiros passos dados na direção da integração de *bookmarks* e *tags* com ontologias estruturadas baseiam-se em RDF e RDFS (*RDF-Schema*). Esses modelos são conjuntos de especificações desenvolvidos pelo *World Wide Web Consortium* (W3C) que tem por objetivo representar informações, por meio do intercâmbio de dados, capazes de definir uma ontologia, ou descrever um esquema de conhecimento. Elas promovem uma estrutura semântica para ontologias processáveis pelas máquinas, que contém as informações necessárias para um sistema computacional, e suporte para interoperabilidade entre estruturas de diferentes comunidades, por meio de classes, valores e propriedades dos esquemas [9].

Entretanto, o avanço promovido por RDF e RDFS não foi o suficiente, e houve o surgimento da OWL, também criada pela W3C. OWL é uma linguagem padronizada que pode ser usada para superar o problema da interoperabilidade semântica, por suportar uma larga variedade de aplicações inteligentes baseadas na *Web*. A OWL provê uma linguagem fundamentada em lógica de descrição para expressar semânticas relativamente ricas visando definir ontologias *Web*. Essas ontologias têm sido largamente exploradas para apoiar o processamento de linguagens naturais e consistente o suficiente com os padrões e arquiteturas modernas da *Web* [12].

Um exemplo do uso combinado entre esses modelos RDF e OWL pode ser visto na iniciativa denominada *Web Of Data*, essa iniciativa cria conteúdo na *Web* descrevendo semanticamente dados de forma estruturada, publicados por fontes distribuídas. As informações podem ser convertidas em conjuntos de dados RDF de acordo com os princípios de *Linked Data* [2]. RDF refere-se a um modelo de descrição de dados adequado para publicar dados estruturados intercambiáveis na *Web*. Isso promove ontologias interpretáveis por máquina, que contêm informações necessárias e formais para suportar a interoperabilidade entre sistemas computacionais [9].

Outra maneira estudada de aproximar ontologias heterogêneas baseia-se na superposição de conteúdo entre duas ontologias [8]. Essa forma de conexão se dá por meio de uma ontologia de mais alto nível, que consegue relacionar duas de nível mais baixo. Ou seja, existem conceitualizações similares entre as diferentes ontologias, mas que estão em vocabulários e disposições muito distintas, então cria-se uma terceira ontologia (meta-modelo) que apresente termos mais genéricos e capazes de interligar os conceitos similares presentes em ambas.

Uma das abordagens na área é a interconexão de ontologias heterogêneas apresentada com o envolvimento da inteligência humana. Primeiramente, coleta-se uma grande quantidade de informações sobre *tagging*, por exemplo, por meio de ferramentas *Web* como *del.icio.us* e *flickr*, em seguida avalia-se esses dados criando relações entre os objetos, os autores e as *tags*. Sistemas de *tagging* podem ser considerados até certo ponto como uma ontologia “leve”.

Kaipainen et al. [13] propôs esse conceito de ontologia “leve” como um meio de conceituar domínios dinâmicos, cuja representação do conhecimento raramente dependerá de um vocabulário controlado definido por especialistas em um domínio. Esses autores definem que uma ontologia “leve” declara um conjunto de metadados para descrever um domínio por meio de propriedades espaciais de um conceito. Em outras palavras, usuários definem termos do domínio e assinalam pesos normalizados para priorizar os termos naquele espaço conceitual. Por exemplo, imagine que o usuário assinale que para uma imagem na *Web* o conceito de “natureza” e o valor “0.7”, no sentido de quanto ele acredita que aquele conceito se aplique à imagem. Em nossa proposta de pesquisa, visamos representar ontologias “leves” via modelos com RDF que exploram o conceito de lógica nebulosa (*fuzzy*) [10].

Paralelamente a isso, pode-se associar recursos lexicais como *Google*, *Wikipedia* e

*WordNet* aos conjuntos de *tags* criados. Essa associação permite promover a substituição de *tags* por conceitos e homônimos. O uso da *WordNet*, por exemplo, pode desenvolver o mapeamento de palavras com significados similares (sinônimos). O último passo é o mapeamento entre as *tags* e os dados já existentes em ontologias. Esse modelo gerado é por fim apresentado a comunidade esperando que ela possa comprovar a qualidade dos resultados e continuar auxiliando na criação desses mapeamentos, caso percebam a falta de informações [7].

Umer e Mundy [18], por exemplo, estudaram um processo de integração de ontologias baseado em três cenários diferentes: reutilização de ontologias, mapeamento de ontologias e fusão de ontologias. Sua abordagem comparou conceitos de ontologia com base em sua sintaxe, semântica e relação entre conceitos. Da mesma forma, Nguyen e Truong [16] investigaram a integração de ontologias *fuzzy*, abordando questões relacionadas a várias estruturas difusas de conceitos e relacionamentos. Os autores propuseram uma abordagem em duas etapas com base no cálculo da distância dos conjuntos de dados envolvidos e aplicaram um algoritmo de consenso para obter a integração. Alves e Santach è [1] propuseram a Ontologia *Folksonomizada* como uma abordagem de fusão considerando ambas as direções de *folksonomias* para ontologias e vice-versa. Bonacin *et al.* [4] forneceu os primeiros passos para obter um metamodelo para apoiar a interoperabilidade entre ontologias "leves", ontologias de OWL e ontologias de semiótica organizacional.

Nossa análise da literatura apontou vários desafios para a reconciliação entre modelos de ontologia e dados gerados por usuários (expressando diferentes maneiras de representar o conhecimento). A maioria dos estudos existentes enfatizou a integração de ontologias descritas na mesma língua ou foram considerados modelos que compartilham posições filosóficas semelhantes.

### 3 Rede de ontologias via metamodelos

No nosso estudo, buscamos conciliar ontologias "leves" e estruturadas. Para isso, fizemos um estudo da construção de ontologias "leves", por meio de um processo de transformação de dados em uma matriz de probabilidades para triplas em RDF *fuzzy*. Depois, formalizamos um metamodelo em eBNF para conciliar informações provenientes das ontologias "leves" com ontologias estruturadas na linguagem OWL (cf. Figura 1).

#### 3.1 Ontologias "leves"

Resumidamente, a construção de ontologias "leves" se deu em dois passos: primeiro desenvolvemos uma matriz de probabilidades, e depois passamos os dados armazenados nessa matriz para triplas em *fuzzy* RDF.

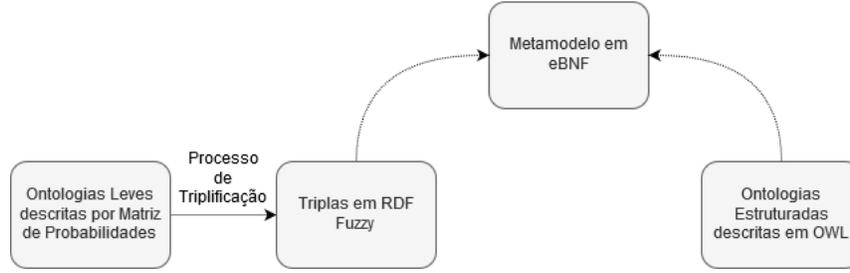


Figura 1: Proposta usando Metamodelo em eBNF.

### 3.1.1 Matriz de Probabilidades

No desenvolvimento da matriz, definimos que cada linha representa uma dimensão na ontologia e cada coluna representa um conceito no domínio. As definições conceituais são caracterizadas por valores, variando de 0 a 1, em que 0 significa que a probabilidade/crença de uma relação entre um item conceitual e uma dimensão ontológica é nula e 1 significa a crença de que essa relação sempre ocorre. Neste sentido, o valor associado a cada intersecção da matriz expressa quanto se acredita que uma dada dimensão ontológica está associada a um dado do domínio.

Usando categorias de carro como exemplo, a Tabela 1 ilustra a distribuição de probabilidades (simulada). Este exemplo expressa que *audiTT* está mais fortemente relacionado a um *Carro Esporte* e *Fiat Panda* a um *Carro Compacto*.

<i>tipoDeCarro</i>	<b>audiTT</b>	<b>Fiat Panda</b>
<b>Esportivo</b>	0.80	0.20
<b>Compacto</b>	0.10	0.90

Tabela 1: Exemplo de uma matriz de probabilidades.

Nossa implementação simplificada de ontologia "leve" é capaz de representar e aumentar dinamicamente o número de relacionamentos a serem expressos. Ela permite a inserção de novas dimensões ontológicas e novos itens de domínio, bem como a remoção dos existentes. Essa exploração de múltiplas perspectivas permite que cada propriedade seja levada em conta em um grau escolhido pelo usuário.

A implementação de uma matriz de probabilidades começa com a definição de conceitos de domínio e dimensões ontológicas. Esses conceitos são inseridos em vetores de conceitos e dimensões ontológicas. Inicialmente, todos os valores da matriz são inicializados com uma probabilidade padrão de 1.

A Figura 2 apresenta uma visão geral dos métodos de execução implementados para criar e atualizar a matriz. Inicialmente, há a seleção de um dos conceitos do domínio (*cf.* A na Figura 2); com base nisso, o segundo método verifica se o conceito selecionado já está presente na matriz (*cf.* B na Figura 2); se necessário, o processo

explora um algoritmo para incluir novas dimensões e conceitos ontológicos com valores padrão (cf. C na Figura 2); se o conceito já está presente, fornecemos um método para selecionar de acordo com as probabilidades uma das dimensões ontológicas da matriz (cf. D na Figura 2). Nesse método, essas dimensões com probabilidades mais altas têm mais chance de serem selecionadas.

Essa dimensão escolhida é usada para fornecer uma resposta ao sistema. A interpretação desta resposta pelos usuários é explorada para atualizar a matriz de probabilidades com base em suas atribuições. O *feedback* pode ser obtido de várias formas, por exemplo, no escopo do nosso cenário de avaliação (apresentado na seção 4), realizamos um sistema de votação (cf. E na Figura 2). Perguntamos aos usuários quais conceitos eles acreditam estar mais conectados a uma dimensão ontológica selecionada; se a maioria votou no conceito que foi inicialmente selecionado, temos um *resultado positivo* (significando uma coerência entre o conceito e a dimensão ordenada), caso contrário, temos um *resultado negativo*.

O resultado positivo dispara para aumentar o valor de probabilidade da dimensão até certo grau (um parâmetro do sistema). Nesse sentido, as probabilidades da dimensão ontológica para um determinado conceito são atualizadas de acordo com um fator de adaptação (ou função) (cf. F in Figure 2). A seleção de dimensões ontológicas e os ajustes de acordo com o *feedback* coletado permitem a convergência para uma representação de crença de acordo com o ambiente e as construções sociais.

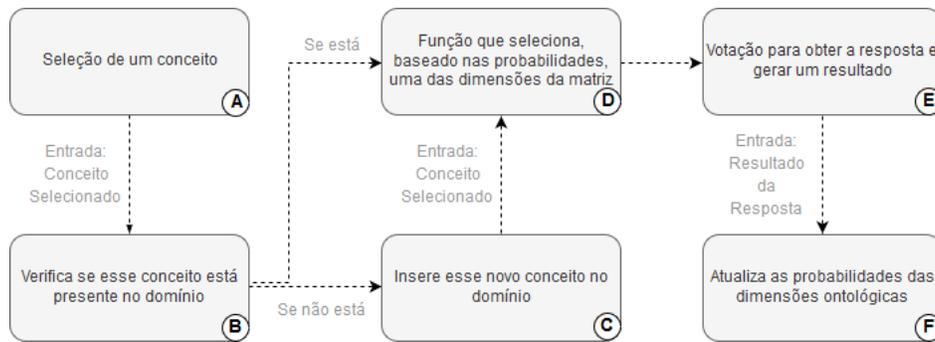


Figura 2: Fluxo para criação e atualização da matriz de probabilidades.

O algoritmo 1 apresenta o pseudo-código do método para atualizar os valores de probabilidade das dimensões ontológicas para um dado conceito. A seguir, explicamos os parâmetros do algoritmo.

- *IncPerc*: Fator de Adaptação;
- *Mat<sub>ip</sub>*: Matriz de probabilidades;
- *OntoDimL*: Vetor de todas as dimensões ontológicas;

- *Dim*: Posição da dimensão ontológica selecionada;
- *ItenL*: Vetor de todos os conceitos do domínio;
- *Iten*: Posição do conceito do domínio selecionado;
- *x*: Número de 0 a 100 para uma função definida  $f(x)$  que aumenta (ou diminui) o fator de adaptação;

---

**Algorithm 1:** Ajuste de probabilidades de dimensões ontológicas para um conceito do domínio.

---

**Require:**  $IncPerc, Mat_{ip}, OntoDimL, Dim, ItenL, Iten, x$

- 1:  $IncPerc \leftarrow f(x)$
  - 2: **if**  $Result == true$  **then**
  - 3:    $Mat_{ip}[OntoDimL.pos(Dim), ItenL.pos(Iten)] =$   
        $Mat_{ip}[OntoDimL.pos(Dim), ItenL.pos(Iten)] * IncPerc$
  - 4: **else**
  - 5:    $Mat_{ip}[OntoDimL.pos(Dim), ItenL.pos(Iten)] =$   
        $Mat_{ip}[OntoDimL.pos(Dim), ItenL.pos(Iten)] / IncPerc$
  - 6: **end if**
  - 7: **return**  $Mat_{ip}$
- 

Um aspecto fundamental é a definição do fator de adaptação, que pode ser aumentado ou diminuído de acordo com o contexto de uso e convergência esperada das probabilidades de dimensões. A adaptação pode ser feita de acordo com diferentes estratégias (um parâmetro): (1) Usando um fator fixo, se obtivermos um *feedback* positivo, a probabilidade da relação entre o conceito e a dimensão ontológica selecionada aumenta de acordo com um fator fixo; valor diminui. (2) se obtivermos um *feedback* positivo, além de aumentar a probabilidade da dimensão ontológica selecionada, ela diminui o valor de probabilidade de outras dimensões ontológicas. Uma função matemática pode ser usada para definir os fatores de aumento e diminuição.

### 3.1.2 Processo de Triplificação

Inicialmente, definimos o conceito de triplas RDF, que representam relacionamentos entre recursos vinculados por um predicado usando uma propriedade formal definida. Um recurso pode ser qualquer coisa descrita no "mundo real", seja uma coisa física como um computador, ou um conceito como "teoria da relatividade". Em RDF, os recursos são representados como *Identificador Uniforme de Recurso* (URI). Neste sentido, uma tripla RDF refere-se a uma entidade de dados composta por sujeito, predicado e objeto definido na forma de  $t = (s, p, o)$  onde:

- **Sujeito:** ( $s$ ) é uma referência URI ou um nó em branco.

- **Predicado:** ( $p$ ) é uma referência URI como propriedade definindo características de um indivíduo em uma classe de ontologia.
- **Objeto:** ( $o$ ) é uma referência URI, ou uma literal, ou um nó em branco.

No desenvolvimento das triplas de *fuzzy* RDF, uma dada relação entre um conceito e uma dimensão da matriz requer 3 triplas. Elas compõem um subconjunto de dados RDF, que no contexto de *Linked Data* é a reunião de um número finito de triplas RDF em um domínio, definido aqui como  $\mathcal{R}' = (t_1, t_2, t_3)$ . No grafo RDF *fuzzy*, exploramos um nó conjuntivo para apoiar a construção dessas três triplas, pois isso nos permite representar melhor a natureza das entidades. Então conectamos cada dimensão ontológica com os conceitos de domínio e seus respectivos valores de crença das relações da seguinte forma:

- $t_1 = (\text{conceitoDoDomínio}, \text{predicadoDeEscolha}, \text{nóConj})$ ;
- $t_2 = (\text{nóConj}, \text{predicadoDeEscolha}, \text{dimensãoOntológica})$ ;
- $t_3 = (\text{nóConj}, \text{grau/crença}, \text{probabilidadeDaRelação})$ .

Nesta proposta, contamos com o conceito de lógica *fuzzy* [13] para expressar triplas RDF. Para isso, adicionamos valores de crença associados às triplas de RDF (sujeito, predicado e objeto). A inferência usando sujeitos e objetos inclui valores de crença entre 0 e 1. Esse valor representa o grau de interconexão entre duas entidades

O desenvolvimento de nosso subconjunto de dados RDF começa vazio ( $\mathcal{R}' = \emptyset$ ). Os itens da coluna (conceitos) na matriz tornam-se os sujeitos das primeiras triplas, e seu objeto refere-se a um nó conjuntivo. O procedimento adiciona a primeira tripla ( $\mathcal{R}' = \mathcal{R}' + t_1$ ). No desenvolvimento das próximas duas triplas, o nó conjuntivo permanece como sujeito; na segunda tripla, o objeto é a relação (dimensão) que queremos apresentar, sendo formada com as linhas da matriz tornando-se os objetos da nossa tripla; então a segunda tripla é adicionada ( $\mathcal{R}' = \mathcal{R}' + t_2$ ); na terceira tripla, mostramos o quanto acreditamos nessa relação, porque a probabilidade da matriz se torna o objeto. Em seguida, o procedimento insere a terceira tripla ( $\mathcal{R}' = \mathcal{R}' + t_3$ ).

Com as dimensões ontológicas e os conceitos do domínio já definidos, podemos escolher um termo preciso para expressar a relação que existe entre eles, e com isso o predicado que seja mais adequado na representação RDF. No exemplo do carro (Tabela 1), podemos escolher um predicado como: 'tipoDeCarro'. A Figura 3 apresenta um grafo fuzzy RDF como parte do resultado do processo de triplificação para a matriz de Tabela 1.

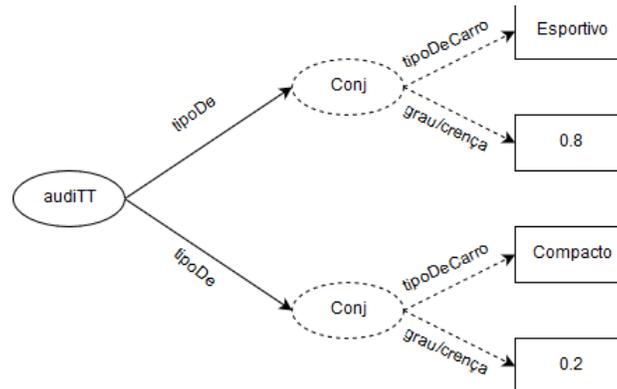


Figura 3: Grafo RDF *fuzzy* para o exemplo de categorias de carros.

### 3.2 Metamodelo em eBNF

Desenvolvemos um metamodelo para suportar a interconexão entre ontologias "leves" representadas em RDF *fuzzy* e ontologias estruturadas representadas em OWL. Para esse fim, escolhemos o uso da linguagem eBNF<sup>1</sup> que é utilizada na construção de metamodelos. Essa linguagem descreve a sintaxe de uma linguagem de programação através de um compilado de suas regras. A Tabela 2 apresenta as formas de controle da linguagem eBNF na criação das regras na compilação, ou seja, apresenta os caracteres que são a base da linguagem, para que possamos descrever componentes de outras linguagens. A Tabela 3 apresenta um exemplo do uso dessas regras na descrição de um inteiro.

<b>Sequência</b>	Os itens aparecem da esquerda para a direita A ordem dos itens é importante
<b>Escolha</b>	Itens alternativos são separados por   Um dos itens deve ser escolhido da lista A ordem dos itens não é importante
<b>Opção</b>	Itens opcionais são colocados dentro de [] Podem ser escolhidos ou descartados
<b>Repetição</b>	Itens que podem ser repetidos são colocados dentro de {} Podem ser escolhidos zero ou mais vezes

Tabela 2: Formas de controle da linguagem eBNF.

Com as regras da linguagem elaboradas, definimos quais são os objetivos do metamodelo em eBNF: (1) representar a conexão de entidades provenientes de modelos distintos, promovendo intercâmbio de informações em ontologias diferentes; e (2) assumir um papel neutro, com poucas premissas. Segundamente, identificamos os blocos que precisam compor o nosso metamodelo para que seja possível fazer a interconexão

<sup>1</sup><https://www.cl.cam.ac.uk/mgk25/iso-14977.pdf>

<b>sinal</b> $\leftarrow$ +   -									
<b>dígito</b> $\leftarrow$ 0   1   2   3   4   5   6   7   8   9									
<b>inteiro</b> $\leftarrow$ [ <i>sinal</i> ] <i>dígito</i> { <i>dígito</i> }									

Tabela 3: Exemplo de descrição em linguagem eBNF.

entre os modelos de ontologias desejados. Por fim, formalizamos esse metamodelo, de forma genérica, e depois o instanciamos no cenário da nossa pesquisa. A figura 4 explicita o metamodelo genérico, enquanto que a figura 5 apresenta o metamodelo aplicado à nossa pesquisa.

Na nossa proposta de solução (cf. Figura 4), utilizamos o conceito de Conexão, pois ele permite a criação de relações entre entidades de um modelo A e de um modelo B. Além disso, cada Conexão, tem uma descrição, por meio do elemento Descrição da Conexão, e é possível identificar a natureza de cada relação, através do Tipo de Conexão. Ademais, ao Tipo de Conexão, está associado um valor que é o Grau de Conexão Acreditado, variando entre 0 e 1.

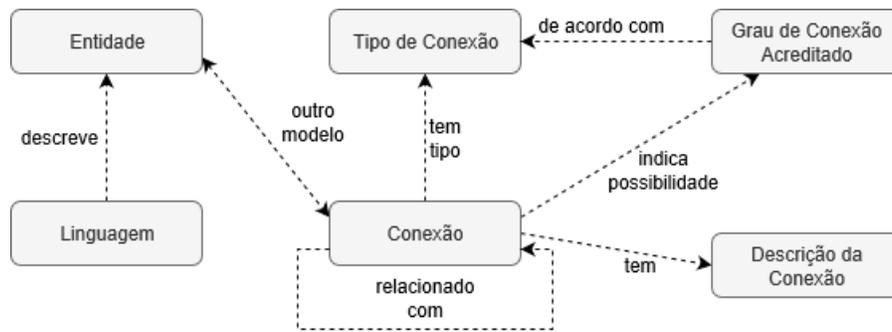


Figura 4: Representação do metamodelo utilizado como base para solução.

A Figura 5 apresenta o metamodelo em eBNF instanciado para os modelos que visamos conciliar. Para o cenário da nossa pesquisa, as Linguagens utilizadas são RDF *fuzzy* e OWL, que estão descrevendo os seguintes conjuntos de Entidades, respectivamente: Sujeito, Predicado, Objeto, Tipo, Valor/Grau e Classe, Propriedade do Objeto, Informação da Propriedade, Instâncias. Uma Conexão é descrita por um único URI, e os Tipos de Conexão podem ser conjunção, disjunção e assertivo. O Grau de Conexão Acreditado descreve o grau de confiança entre a Conexão e seu Tipo, e a Descrição da Conexão nos permite descrever que tipo de relação está sendo determinada, por exemplo, parte de ou instância de.

Com o metamodelo devidamente descrito e instanciado, o Código 1 apresenta a formalização em eBNF que descreve as relações que são expressas na nossa rede de ontologias heterogêneas. Uma das maneiras avaliadas para a interconexão do metamodelo descrito em eBNF e as ontologias expressas em RDF *fuzzy* e OWL foi a

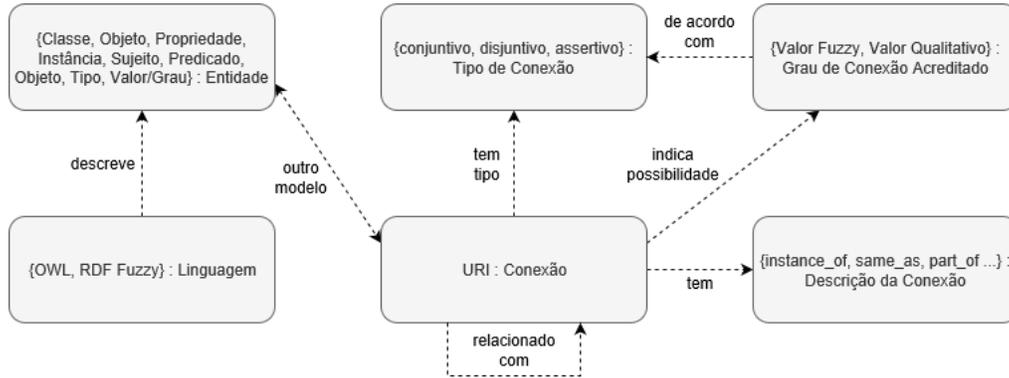


Figura 5: Representação do metamodelo instanciado utilizado como solução.

utilização de um *Parser* via uma classe em Java para interpretar a sintaxe de uma dada entrada e desenvolver relações entre esses dados. Neste trabalho, iniciamos a implementação do parser interpretador do metamodelo, mas estudos adicionais são necessários para uma avaliação refinada do mesmo.

Listing 1: Formalização do metamodelo utilizado como solução em eBNF.

```

OIDM ::= 'Oidm(' [ idmID ] { language } ')'
language ::= 'Language(' languageID langName { axiom } ')'
axiom ::= entity | connection
entity ::= 'Entity(' entityID entityName ')'
connection ::= 'Connection(' connectionID [connectionName]
connectionRestriction connectionDescription ')'
connectionRestriction ::= connectionType '(' connectionRestrictionID
entityID {entityID beliefConnectionWeight} ')'
connectionType ::= DisjointConnection | ConjunctiveConnection |
AssertConnection
connectionDescription ::= 'is a' | 'part of' | 'same as' | 'instance_of' |
beliefConnectionWeight ::= floatingPointNumber | valueScale
floatingPointNumber ::= digit {digit} ['. ' digit {digit} ]
digit ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
valueScale = '[' {plainLiteral ','}* plainLiteral ']'
langName ::= plainLiteral
entityName ::= plainLiteral
connectionName ::= plainLiteral
connectionTypeName ::= plainLiteral
oidmID ::= URReference
languageID ::= URReference

```

```

entityID ::= URireference
connectionID ::= URireference
connectionTypeID ::= URireference
conectionRestrictionID ::= UIRreference
/* reutilizado de OWL */
plainLiteral ::= lexicalForm | lexicalForm@languageTag
lexicalForm ::= as in RDF, a unicode string in normal form C
languageTag ::= as in RDF, an XML language tag

```

## 4 Avaliação

Apresentamos o uso da nossa rede de ontologias heterogêneas em um cenário de aplicação, integrando ontologias estruturadas em OWL e ontologias "leves" em RDF *fuzzy*, por meio do metamodelo formalizado em eBNF.

Os dados de entrada para a matriz foram informações extraídas de oficinas com crianças da Divisão de Educação Infantil e Complementar da UNICAMP, Brasil. Nessas *workshops*, estudamos como as crianças interpretam as ações de um robô mBot [3]. Aplicamos a nossa técnica de implementação de ontologias "leves" como forma de selecionar o comportamento do robô na dinâmica da oficina, *i.e.*, uma matriz representando uma propriedade 'expressa\_por', que relaciona as emoções às ações do robô expressando essas emoções. O sistema foi projetado para moldar o comportamento dos robôs de acordo com o *feedback* das respostas das crianças em sessões iterativas e, assim, a matriz é atualizada ao longo do tempo.

No total, 25 crianças, com idade entre 4 e 5 anos de idade, participaram deste estudo <sup>2</sup>. As crianças vieram de duas classes separadas, de manhã e à tarde — referenciadas a partir de agora como grupo 1 e grupo 2 — com respectivamente 13 e 12 alunos cada. Cada grupo tinha um professor diferente. O *workshop* adotou o seguinte procedimento (*cf.* [6] para obter detalhes):

- Uma das crianças, escondida das outras em uma caixa, escolhe uma emoção para expressar (seleciona um conceito - A na Figura 2); e, executa uma imitação representando essa expressão emocional na frente de uma câmera.
- O sistema reconhece essa emoção;
- Com base em nossa ontologia "leve", o sistema seleciona uma ação a ser realizada pelo robô (seleciona uma das dimensões ontológicas);
- O robô executa a ação na frente das crianças em uma área de audiência;

---

<sup>2</sup>aprovado pelo Comitê de ética em pesquisa 72413817.3.00000.5404

- As crianças tentam interpretar a emoção expressa pelo robô. Cada criança, em particular, vota em uma emoção (E na Figura 2), usando cartões *emoji* identificados com um RFID (identificação por radiofrequência).
- A matriz é atualizada de acordo com a resposta das crianças (F na Figura 2);

A matriz de probabilidades é a nossa solução para representar o conhecimento sobre as expressões emocionais e um conjunto de comportamentos que podem ser realizados pelo robô. A matriz de probabilidades representa as ações dos robôs (linhas) com expressões emocionais (colunas); Seis ações foram programadas no robô para a oficina; as linhas foram rotuladas de ação 1 a ação 6 (*cf.* Tabela 4) *e.g.*, caminhe para a frente e apresente os olhos piscando em um painel frontal. As colunas foram rotuladas com emoções que essas ações poderiam representar (felicidade, tristeza, surpresa, desprezo, raiva e nojo). Cada célula representa um valor ponderado de probabilidade do robô executar uma ação para uma determinada expressão emocional. A matriz fornece o comportamento flexível e de imprecisão ao comportamento do robô.

A matriz relacionou a probabilidade de uma determinada ação do robô representar uma emoção para uma dada situação. É importante mencionar que em uma ontologia "leve", essa relação não é dada *a priori*, esses valores podem variar de situação para situação. Ao longo da execução da dinâmica do *workshop* e das respostas de entrada das crianças, as probabilidades foram ajustadas representando a compreensão das crianças sobre a correlação entre os conceitos e as dimensões da ontologia.

Inicialmente, a matriz inteira foi inicializada para valorizar 1, o que significa que essa relação é totalmente desconhecida no início. A matriz foi atualizada durante todo o *workshop* com base nas respostas que expressaram a interpretação das crianças sobre as ações do robô. Por exemplo, após várias rodadas, a ação do robô de andar para a frente e apresentar olhos piscando estava relacionada à expressão de emoção *felicidade* pelas crianças. Essa ação estaria fortemente relacionada à felicidade de acordo com a matriz de probabilidade.

Tabela 4 apresenta a matriz final de probabilidades no final da aula da tarde (grupo 2), que obteve a melhor efetividade. Nesta configuração, o maior aumento de probabilidades foi 40% em caso de *feedback* positivo, e a maior diminuição foi 74% em caso de *feedback* negativo.

A etapa seguinte de nossa avaliação concerniu na transformação da ontologia "leve" em triplas de *fuzzy* RDF. Para tanto, aplicou-se o processo de triplificação à matriz final gerada durante o *workshop* (tabela 4).

As emoções foram traduzidos para sujeitos das triplas geradas. As ações do robô (dimensões) foram traduzidas para objetos com o valor de crença de cada relação. O predicado 'expressaEmoção' indica as propriedades representadas pela matriz. Foram criados seis subconjuntos de dados RDF compostos por 3 triplas no seguinte formato:

$$\mathcal{R}' = ((emoção_y, expressaPor, nóConj), (nóConj, expressaEmoção, ação_z), (nóConj, grauDeCrença, probDaRelação)).$$

	Desprezo	Felicidade	Raiva	Tristeza	Nojo	Surpresa
Ação 1	1.00	0.84	0.31	1.00	0.51	1.00
Ação 2	1.00	0.36	0.60	0.51	1.00	1.00
Ação 3	0.71	0.60	0.60	0.51	1.00	1.00
Ação 4	1.00	0.26	0.43	1.00	1.00	1.00
Ação 5	1.00	0.36	1.40	1.00	1.00	1.00
Ação 6	1.00	0.36	0.60	1.00	1.00	1.00

Tabela 4: Comportamento da matriz após 12 iterações

Abaixo, 2 apresenta um exemplo de triplas *fuzzy* RDF, descritas em turtle, geradas para a emoção de felicidade e ação 1 com base nos valores da matriz da oficina. A criação de todo o conjunto de dados RDF segue uma construção análoga para cada par ação e emoção.

Listing 2: Exemplo em turtle das triplas *fuzzy* RDF para Ação 1.

```
@prefix ex: <http://nossocenario.com/>
ex:felicidade ex:expressaPor ex:n Conj
ex:n Conj
    ex:expressaEmocao ex:acao1
    ex:grauDeCrenca 0.84^^xsd:decimal
```

Em seguida, escolhemos uma ontologia estruturada em OWL sobre emoções, já existente na Web<sup>3</sup>, para compor, junto com os dados da oficina a nossa rede de ontologias heterogêneas, e aplicamos no nosso metamodelo, como apresentado na figura 6 abaixo.

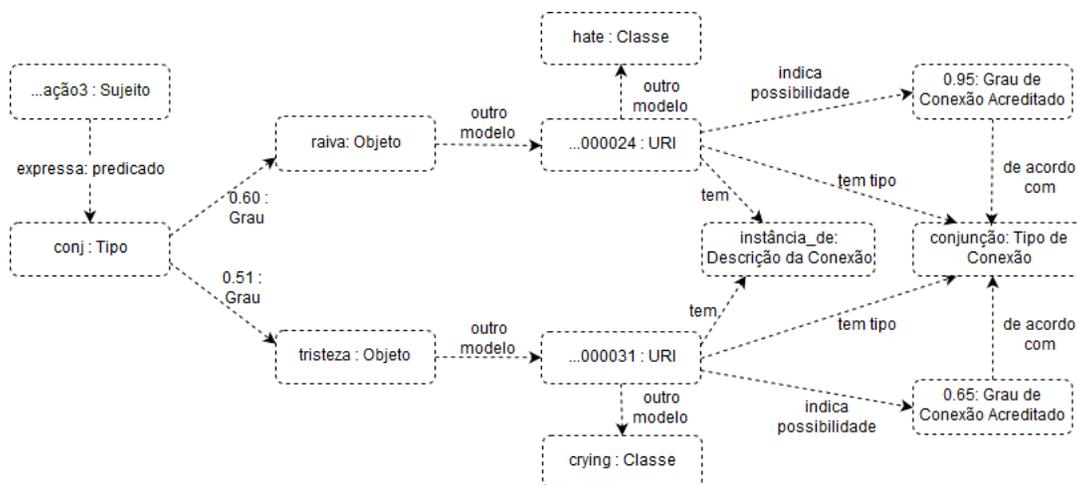


Figura 6: Exemplo do uso do nosso metamodelo no cenário de teste.

<sup>3</sup><http://purl.obolibrary.org/obo/MFOEM>

Nesse cenário, um modelo *fuzzy* RDF define o grau com qual a "ação3" pode expressar as emoções "tristeza" e "raiva" (configuração interna do modelo *fuzzy* RDF). O nosso metamodelo permite relacionar que o objeto "raiva" é equivalente a Classe "anger" em OWL no modelo de ontologia estruturada da *Web* (expressa pelo URI que termina em "000009"), com 95% de crença. Da mesma maneira, relacionamos o objeto "tristeza" no modelo de ontologias "leves" com a Classe "crying" em OWL no modelo de ontologia estruturada da *Web*<sup>4</sup> (expressa pelo URI finalizado com "000031"), e há uma crença nisso de 65%.

Para chegar aos valores de crença entre as ontologias de diferentes modelos, diversas formas podem ser adotadas, como análise semântica entre os termos provenientes das ontologias heterogêneas, comparação com banco de dados sobre o tema, etc. No nosso exemplo sobre a ontologia de emoções, poderíamos ainda, imitando o desenvolvimento da matriz de probabilidades, coletar *feedbacks* que auxiliem na evolução dos valores de crença das relações entre os dados das diferentes ontologias.

## 5 Discussão

A natureza formal das ontologias as torna mais habilitadas para a interpretabilidade de conceitos e raciocínio automatizado na máquina, pois fornece uma semântica bem definida para termos de um domínio. No entanto, o uso de diferentes formalismos para representar um domínio do conhecimento impede a interoperabilidade adequada entre diferentes sistemas. Isso não é atenuado quando se considera o uso de formalismos, como ontologias *Web* e ontologias "leves".

O conhecimento em *fuzzy* surgiu das necessidades de vários campos de aplicação [14]. A *Web* requer abordagens capazes de apoiar a representação das informações geradas pela experiência humana. O avanço da iniciativa de dados vinculados requer que os modelos de representação sejam interpretáveis por máquina e interoperáveis. Para isso, nossa pesquisa objetivou compreender como as informações geradas pela experiência do usuário podem ser representadas por ontologias "leves" [10] (*i.e.*, sem impor regras rígidas, uma linguagem específica ou limitações de um vocabulário específico), como ele pode ser transformado em modelos interpretáveis por máquina e como ele pode ser traduzido para um modelo interoperável da *Web*. Nossa investigação desenvolveu uma técnica para obter um conjunto de dados representado em RDF *fuzzy*, assim como um metamodelo formal que definiu blocos de construção elementares que são instanciados para elementos correspondentes nos formalismos alvo, permitindo a construção de uma rede de ontologias heterogêneas.

A avaliação conduzida da pesquisa envolveu um ambiente educacional com dois grupos de crianças participando de atividades educativas apoiadas por um robô. Este é um cenário rico envolvendo conceitos subjetivos como emoções e ações expressando

---

<sup>4</sup><http://purl.obolibrary.org/obo/MFOEM>

emoções. Após 12 iterações, a matriz apresentou uma convergência consistente, atribuindo maiores probabilidades às ações que representam as expressões emocionais [6]. A velocidade de convergência foi abaixo da expectativa dos pesquisadores. Outro aspecto refere-se à não existência de um padrão bem estabelecido de reconhecimento em relação às emoções entre os grupos estudados de crianças. Estudos com outras funções de adaptação podem ser necessários para melhorar o modelo.

O conjunto de dados RDF gerado pode ser vinculado a outras ontologias, incluindo as bem formalizadas. Em nosso cenário de avaliação, por exemplo, vinculamos o RDF gerado à ontologia de emoções que está disponível na linguagem OWL na *Web* e fornece uma extensa representação de fenômenos afetivos.

Descobrimos que é possível acomodar entidades conceituais no metamodelo para criar conexões de dados de um modelo para outro. Exploramos o conceito de grau de crença, inspirado nos modelos *fuzzy* RDF, para determinar o grau de confiança entre elementos de diferentes modelos. O cenário de aplicação apresentado mostrou os benefícios e limitações de nossos achados. Os exemplos práticos ilustraram como um conceito em uma ontologia da *Web* foi conectado a um sujeito em ontologias RDF *fuzzy* com um determinado grau de confiança com base em nosso modelo.

Nosso modelo se concentra no suporte à interoperabilidade de dados. Trabalhos futuros envolverão um refinamento do metamodelo, além de implementar e testar outras formalizações. Pretendemos estudar como nossa proposta pode ser empregada em cenários em que as ontologias evoluem ao longo do tempo, uma vez que novos conceitos são criados, modificados e removidos ao longo da execução do sistema.

## 6 Conclusão

As ontologias são artefatos centrais para representar conhecimento em sistemas de informação. No entanto, diferentes modelos conceituais dificilmente são conciliados no mesmo espaço conceitual. Nesta pesquisa, propusemos um metamodelo para possibilitar que diferentes modelos de ontologia coexistam em uma rede de ontologias. Propusemos uma implementação de ontologias "leves" baseada em uma matriz de probabilidades, métodos para sua criação e atualização, além de sua transformação em modelos *fuzzy* RDF. Nossa abordagem foi avaliada em um cenário educacional, no qual várias rodadas geraram o *feedback* necessário para ajustar o comportamento simplificado de um robô, que utilizou nossa matriz de probabilidades como uma implementação de uma ontologia leve. O robô executou ações representando expressões de emoção de acordo com o significado dado pelas crianças durante a oficina conduzida. Os resultados revelaram que o metamodelo é adequado para representar tipos de conexão distintos considerando um grau de conexão de crença. Aspectos como estratégias de inferência, aprofundamento em questões teóricas e o uso do metamodelo em situações do mundo real devem ser investigados, sendo aspectos principais

em trabalhos futuros.

## Agradecimentos

Este trabalho teve apoio da Fundação de amparo à pesquisa do estado de São Paulo (FAPESP) (#2017/02325-5 e #2018/06416-8)<sup>5</sup>.

## Referências

- [1] Hugo Alves and André Santanchè. Folksonomized ontology and the 3e steps technique to support ontology evolvment. *Journal of Web Semantics*, 18(1):19 – 30, 2013.
- [2] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.
- [3] Make Block. Make block inc. <https://www.makeblock.com/steam-kits/mbot-2>. Accessed: 2019-01-08.
- [4] Rodrigo Bonacin, Ivo Calado, and Julio Cesar dos Reis. A metamodel for supporting interoperability in heterogeneous ontology networks. In Kecheng Liu, Keiichi Nakata, Weizi Li, and Cecilia Baranauskas, editors, *Digitalisation, Innovation, and Transformation*, pages 187–196. Springer, 2018.
- [5] K. K. Breitam. Ontology in computer science. *Semantic Web: Concepts, Technologies and Applications.*, 2007.
- [6] Ricardo Caceffo, Eliana Moreira, Rodrigo Bonacin, Julio Cesar dos Reis, Marleny Carbajal, João Vilhete D’Abreu, Camilla Brennad, Luma Lombello, José Armando Valente, and Maria Cecília Baranauskas. Collaborative meaning construction in socioenactive systems: Study with the mbot. In *21st International Conference on Human-Computer Interaction*, 2019.
- [7] Céline Van Damme, Martin Hepp, and Katharina Siorpaes. Folksonology: An integrated approach for turning folksonomies into ontologies. 2007.
- [8] N. Guarino. Formal ontology in information systems. *IOS Press, Amsterdam.*, 1998.

---

<sup>5</sup>As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do(s) autor(es) e não necessariamente refletem a visão da FAPESP.

- [9] E. Jacob. Ontologies and the semantic web. *Bulletin of the American Society for Information Science and Technology Magazine*, 2003.
- [10] Mauri Kaipainen, Peeter Normak, Katrin Niglas, Jaagup Kippar, and Mart Laanpere. Soft ontologies, spatial representations and multi-perspective explorability. *Expert Systems*, 25(5):474–483, 2008.
- [11] Mauri Kaipainen, Niklas Ravaja, Pia Tikka, Rasmus Vuori, Roberto Pugliese, Marco Rapino, and Tapio Takala. Enactive systems and enactive media: Embodied human-machine coupling beyond interfaces. *Leonardo*, 44:433–438, 2011.
- [12] Golbeck J. Banerjee J. Hendler J. Kalyanpur, A. Owl: Capturing semantic information using a standardized web ontology language. *Multilingual Computing Technology Magazine (15)(7)*, 2004.
- [13] Yanhui Lv, Zongmin Ma, and li Yan. Fuzzy rdf: A data model to represent fuzzy metadata. pages 1439 – 1445, 07 2008.
- [14] Mauro Mazzieri. A fuzzy rdf semantics to represent trust metadata. In *1st Workshop on Semantic Web Applications and Perspectives (SWAP2004)*, pages 83–89, 01 2004.
- [15] Elke Michlmayr. A case study on emergent semantics in communities. In *In Workshop on Semantic Network Analysis, International Semantic Web Conference (ISWC2005)*, 2005.
- [16] Van Du Nguyen, Hai Bang Truong, Mercedes G. Merayo, and Ngoc Thanh Nguyen. Intelligent collectives: Impact of diversity on susceptibility to consensus and collective performance. In *ICCCI (1)*, volume 11055 of *Lecture Notes in Computer Science*, pages 13–22. Springer, 2018.
- [17] Rogério Aparecido Sá Ramalho. *Desenvolvimento e utilização de ontologias em bibliotecas digitais: uma proposta de aplicação*. PhD thesis, Sao Paulo State University, Brazil, 2010.
- [18] Q Umer and D Mundy. Semantically intelligent semi-automated ontology integration. In *World Congress on Engineering*, 2012.