



Universidade Federal de Pernambuco  
Centro de Informática

Graduação em Ciência da Computação

## **Normalização de ontologias ALC para o uso com o LeanCop**

Adriano Silva Tavares de Melo

Trabalho de Graduação

Recife  
11 de julho de 2011



Universidade Federal de Pernambuco  
Centro de Informática

Adriano Silva Tavares de Melo

## **Normalização de ontologias ALC para o uso com o LeanCop**

*Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.*

Orientador: *Fred Freitas*

Recife  
11 de julho de 2011



*Eu dedico esse trabalho a minha mãe, namorada,  
familiares e amigos.*



# Agradecimentos

Gostaria de agradecer à minha mãe, que sempre acreditou em mim, que me deu condições de estudar, que abdicou de várias coisas ao decorrer de sua vida para dar prioridade aos filhos, que nos incentivou a crescer, que nos ensinou valores, que, em fim, nos amou incondicionalmente.

Agradeço a todos da minha família que me apoiaram a estudar e a trilhar caminhos que me deixaram longe, em especial ao meu irmão, madrinha e avó, que sempre me falam o quanto são orgulhosos da minha decisão de sair da minha cidade para evoluir profissionalmente. Obrigado!

Agradeço à minha namorada, que tornou essa jornada longe dos amigos e família mais agradável e feliz.

Agradeço a todos que me ajudaram profissionalmente e academicamente, eu não teria conseguido chegar até aqui sem essa ajuda. Obrigado!





*Uma vida de sacrifício é o cume supremo da arte; é cheia de alegria  
verdadeira.*

—MOHANDAS KARAMCHAND GANDHI



# Resumo

resumo....

**Palavras-chave:** método das conexões, lógica de descrição, lógica, inteligência artificial



# Abstract

**Keywords:** <DIGITE AS PALAVRAS-CHAVE AQUI>



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Web Semântica	1
1.1.1	Aplicações	2
1.1.1.1	Gerenciamento de Conhecimento	2
1.1.1.2	Comércio eletrônico Business to Consumer (B2C)	2
1.1.1.3	Comércio eletrônico Business to Business (B2B) e agentes pessoais	3
1.1.2	Tecnologias	3
1.1.2.1	Metadados Explícitos	3
1.1.2.2	Ontologias	4
1.1.2.3	Lógica	5
1.1.2.4	Agentes	5
1.2	Organização da Monografia	5
<b>2</b>	<b>Lógica de Descrição <i>ALC</i></b>	<b>7</b>
2.1	Sintaxe da Lógica de Descrição	7
2.2	OWL: Web Ontology Language	8
2.2.1	OWL Full	9
2.2.2	OWL DL	9
2.2.3	OWL Lite	10
<b>3</b>	<b>Normalização para o método das conexões</b>	<b>11</b>
3.1	Tradução de ontologias <i>ALC</i> para forma normal disjuntiva	11
3.1.1	Algoritmo Proposto	13
<b>4</b>	<b>leanCoP</b>	<b>15</b>
<b>5</b>	<b>Conclusão</b>	<b>17</b>
5.1	Definition 1 (First-order logic syntax).	17





# Lista de Figuras

1.1	Código HTML de uma página da americanas.com	4
1.2	Exemplo de código com semântica para um produto	4



# Lista de Tabelas

2.1	Notação da lógica de descrição	8
2.2	Sintaxe e semântica de alguns constructos de lógica de descrição com anotação de expressividade	9
3.1	Tabela de regras a serem aplicadas após o método purify	14



## CAPÍTULO 1

# Introdução

A World Wide Web é uma das tecnologias mais revolucionárias que o homem já inventou. Ela mudou em escala global a forma com que pessoas e empresas trocam informações, contribuindo para que o conhecimento se tornasse mais universal e que limites físicos e lingüísticos fossem cada vez mais minimizados.

A web como conhecemos hoje nasceu de uma proposta feita por Tim Berners-Lee à empresa CERN em 1989 [BL89]. O problema enfrentado pela empresa na época era a perda de informações internas por falta de documentação ou pela saída de algum funcionário. A solução proposta por Berners-Lee foi fazer uma rede de documentos interligados por hyperlinks em que cada setor da empresa poderia adicionar novos documentos.

A estrutura básica que Berners-Lee montou a 22 anos evoluiu a passos largos em relação à escalabilidade e padronização de protocolos e linguagens, tendo hoje cerca de 2 bilhões de usuários, mais de 30% da população do planeta.

Apesar do avanço das infra-estruturas e serviços para a Web, ainda há muito o que evoluir. Uma das propostas de mudanças é prover uma maior expressividade da linguagem que descreve os documentos na Web [Hef04]. Hoje, esses documentos não possuem um significado que possa ser extraído de forma concisa, apresentam ambigüidade, misturam os dados com elementos visuais e muitas vezes não podem ser indexados por engenhos de busca.

## 1.1 Web Semântica

*"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web, the content, links, and transactions between people and computers. A **Semantic Web** which should make this possible has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The **intelligent agents** people have touted for ages will finally materialize."*

Tradução literal: *"Eu tenho um sonho para a Web [em que os computadores] tornam-se capazes de analisar todos os dados na Web, o conteúdo, links, e as transações entre pessoas e computadores. A **Web Semântica** que deve tornar isso possível ainda está para surgir; mas quando isso acontecer, os mecanismos dia-a-dia da burocracia do comércio e nossas vidas diárias serão tratados por máquinas falando com máquinas. Os **agentes inteligentes** que as pessoas têm falado por anos vão finalmente se concretizar."*

A Web Semântica citada no texto de Berners-Lee acima é uma iniciativa de pesquisadores da área de inteligência artificial e lingüística computacional que estudam como adequar a Web de hoje a uma infra-estrutura que a tornará mais acessível às máquinas. Essa nova roupagem que os pesquisadores querem dar à Web permitirá que serviços mais sofisticados possam ser construídos, por exemplo:

### 1.1.1 Aplicações

#### 1.1.1.1 Gerenciamento de Conhecimento

Gerenciamento de conhecimento está relacionado à aquisição, acesso e manutenção de conhecimento dentro de uma empresa ou organização. Essa atividade se tornou e está se estabelecendo como uma necessidade básica em grandes empresas visto que o conhecimento que é gerado internamente agrega valor, pode se tornar um diferencial competitivo e também pode aumentar a produtividade de seus colaboradores. Com o uso de tecnologias criadas para a Web Semântica, soluções para G.C. podem melhorar em vários aspectos, entre eles:

- Organização do conhecimento existente a partir de seu significado;
- Geração de novas informações de forma automática;
- Checagem de inconsistências semânticas em documentos;
- Substituição de consultas baseadas em palavras-chave por perguntas em linguagem natural;

#### 1.1.1.2 Comércio eletrônico Business to Consumer (B2C)

O comércio eletrônico entre vendedores e consumidores é um dos modelos de negócio na Internet que melhor se estabeleceu, sites como amazon <sup>1</sup>, americanas <sup>2</sup> e mercado livre <sup>3</sup> possuem público fiel e que os visitam por vários objetivos. É muito comum para a geração que cresceu imersa na Web entrar em sites de compra como esses a procura do melhor preço antes de decidir fazer uma compra. Muitas vezes o produto não é adquirido em uma loja virtual, mas a pesquisa inicial de preços é que muitas vezes determina a escolha do produto. Observando esse comportamento, sites como o buscapé <sup>4</sup> fazem o trabalho de indicar qual é a loja que está com o melhor preço.

A Web Semântica pode ajudar nesse cenário provendo interfaces de consulta mais completas aos sites que fazem comparação de preços, porém, com muito mais detalhes técnicos sobre o produto. Supondo que cada produto tem, por exemplo, uma ontologia que o descreve em detalhes (provida pelo fabricante ou por sites de review de produtos), o consumidor poderá fazer comparações muito mais detalhadas, ajudando-o a encontrar o produto que vai suprir a sua necessidade.

---

<sup>1</sup> site: amazon.com

<sup>2</sup> site: americanas.com

<sup>3</sup> site: mercadolivre.com.br

<sup>4</sup> site: buscape.com.br

### 1.1.1.3 Comércio eletrônico Business to Business (B2B) e agentes pessoais

A maioria das pessoas que compram serviços ou produtos na Web só conhecem o comércio eletrônico do tipo B2C, mas existem tecnologias para comércio do tipo B2B, Business to Business, agentes computacionais de empresas que se comunicam para fechar acordos e otimizar o ciclo de negócios que muitas vezes já podem ser previstos e modelados.

Com a popularização da Web Semântica e a introdução de agentes pessoais e agentes que representam negócios, eles poderão se comunicar de forma mais natural e aplicações para otimizar tarefas do dia-a-dia que poderão ser produzidos. Por exemplo, um médico que possua um agente pessoal que negocie a sua agenda com os agentes pessoais de seus clientes pode ser utilizado para remarcar seus atendimentos em caso de uma viagem ou imprevisto, agindo como uma secretária virtual.

## 1.1.2 Tecnologias

Aplicações como as citadas acima já existem, mas o trabalho de engenharia para conseguir bons resultados é alto devido às tecnologias que são adotadas hoje. Vamos usar o case do site de comparação de preços BuscaPé na próxima seção da monografia.

### 1.1.2.1 Metadados Explícitos

A primeira tarefa de engenharia de um agente coletor de preço, como os do buscaPé, é fazê-lo visitar vários sites de compras todo dia à procura de modificações nas listas de produtos para saber quais estão disponíveis naquela loja. É feito então o *parsing* do HTML de cada site de compras à procura das informações de preço, descrição, avaliação e detalhes de cada produto. A limitação dessa abordagem é que sempre que um dos sites de compra mudar o layout (estrutura do HTML), um novo script de parsing deverá ser escrito. A grande demanda técnica de uma aplicação como essa é a escrita de agentes muito especializados para atingir bons resultados.

Na figura 1.1 está parte do código em HTML de uma página de produtos da americanas.com. As informações do produto estão cercados apenas de código para a renderização dessa página pelo browser. Ou seja, a única preocupação dos engenheiros da americanas.com foi a leitura por humanos da informação do produto. Uma aplicação que deseje usar as informações dos produtos da loja vai ter que fazer um agente especializado no parsing desse código.

Motores de busca que também se baseiam em parsing de páginas para extrair informações da Web dificilmente saberão, por exemplo, qual é o preço de um produto nesse site da americanas.com, já que há várias informações de preço na página e o parsing que é feito não é otimizado para sites específicos.

A consequência para o usuário final é que ele terá que usar um site específico como o buscapé ou terá que fazer buscas a um engenho de busca por palavras-chave para achar os sites de compra que possuam um produto e em uma segunda etapa, fazer a análise de preços manualmente.

A abordagem da Web Semântica para resolver problemas como esse não é fazer agentes especializados (como os do buscapé), e sim, anotar metadados semânticos dos documentos disponíveis na Web. O exemplo dado anteriormente seria escrito na figura 1.2.

```

<div class="chosenProds infoP">
  <strong>Notebook Itautec W7440 c/ Intel® Core i3 370M 2.5GHz 4GB 500GB DVD-RW LED 14.5q
  uot; Windows 7 Premium – Itautec</strong>
</div>
<div class="infoProd">
  <p><strong>Design e tecnologia na ponta dos seus dedos</strong><br><br>
  ....
</div>
<dl>
  <dt>Marca</dt>
  <dd class="">Itautec</dd>
  <dt>Processador</dt>
  <dd class="">Core i3 370M 2.4GHz.</dd>
  <dt>Barramento</dt>
  <dd class="">2.5 GT/s</dd>
  <dt>Cache</dt>
  <dd class="">3 MB</dd>
  ...
</dl>

```

**Figura 1.1** Código HTML de uma página da americanas.com

```

<product
  <class>notebook</class>
  <vendor>Itautec</vendor>
  <model>W7440</model>
  <attributes>
    <attr name="processor">Intel Core i3 370M 2.4GHz</attr>
    <attr name="barramento">2.5 GT/s</attr>
    <attr name="cache">3 MB</attr>
  </attributes>
</product>

```

**Figura 1.2** Exemplo de código com semântica para um produto

### 1.1.2.2 Ontologias

O termo ontologia vem da filosofia, nesse contexto, é um ramo da filosofia que se dedica a estudar a natureza da existência, concentra-se em identificar e descrever o que existe no universo. Em computação, uma ontologia é um artefato para descrever um domínio. Consiste em uma lista finita de termos e relações entre eles. Os termos denotam conceitos importantes de um domínio [AH08].

Grande parte dos trabalhos referentes à Web Semântica estão ligados a ontologias, inclusive este. As linguagem de descrição de ontologias mais importantes para a Web são:

- XML: usado para dirigir a sintaxe de documentos estruturados. Não impõe restrições semânticas no conteúdo do documento;
- XML Schema: linguagem para impor restrições na estrutura dos documentos XML;
- RDF: modelo de dados para recursos (objetos) e relações entre eles. As restrições semânticas são fixas e podem ser representados a partir da sintaxe do XML;
- RDF Schema: descreve as propriedades e classes dos objetos RDF;
- OWL: linguagem rica para modelagem de classes, propriedades, relações entre classes (e.g. disjunção), restrições de cardinalidade, características de propriedades (e.g. simetria). Mais detalhes sobre a OWL serão dados no capítulo 2 dessa monografia.



### 1.1.2.3 Lógica

Lógica é a disciplina que estuda os princípios do raciocínio. Ela provê linguagens formais para expressar conhecimento, a semântica formal para a interpretação de sentenças sem precisar realizar operações sobre a base de conhecimento e a transformação de conhecimento implícito em conhecimento explícito, através de deduções a partir da base de conhecimento [AH08].

Lógica é mais geral que ontologias, ela pode ser usada por agentes inteligentes para tomada de decisões e escolha de ações. Por exemplo, um agente de B2C pode dar um desconto a um cliente baseado na seguinte regra:

$$\forall xy, cliente(x) \wedge produto(y) \wedge clienteFiel(x) \rightarrow desconto(x, y, 5\%)$$

Onde  $cliente(x)$  indica que  $x$  é um cliente/consumidor,  $produto(y)$  indica que  $y$  é um produto de uma loja,  $clienteFiel(x)$  indica que  $x$  é um cliente fiel da loja e  $desconto(x, y, 5\%)$  indica que o cliente  $x$  terá um desconto de 5% no produto  $y$ .

### 1.1.2.4 Agentes

Um agente é tudo o que pode ser considerado capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente por meio de atuadores [RN02]. Agentes lógicos são aqueles que executam ações através de uma base de conhecimento e possuem um requisito fundamental, quando ele formula uma pergunta para a base de conhecimento, a resposta deve seguir o que já foi informado anteriormente.

Agentes para a Web Semântica utilizam as três tecnologias que já foram descritas:

- Metadados serão usados para identificar e extrair informações da Web;
- Ontologias serão usadas para dar assistência às consultas realizadas à Web, interpretar informações recuperadas e para comunicação com outros agentes;
- Lógica será usada para processar informações recuperadas, chegar a conclusões e tomar decisões;

## 1.2 Organização da Monografia

Esta monografia está dividida em cinco capítulos. No Capítulo 1, é apresentada uma visão geral sobre a Web Semântica, exemplificando com aplicações e citando as tecnologias que estão sendo usadas. No Capítulo 2, são apresentados conceitos referentes a Lógica de Descrição e a sua ligação com a linguagem de descrição de ontologias OWL. No Capítulo 3 são descritos os algoritmos para normalização de ontologias para a Forma Normal Positiva. No Capítulo 4 o LeanCop é apresentado e é mostrado a validação do trabalho realizado. No Capítulo 6 apresenta as considerações finais sobre o trabalho, bem como propostas de trabalhos futuros.



# Lógica de Descrição *ALC*

Lógica de Descrição é uma família de linguagens de representação de conhecimento que pode ser usada para representar o conhecimento de domínio de uma aplicação de forma estruturada e formal [BCM<sup>+</sup>03].

A motivação para estudar lógica de descrição neste trabalho vem da Web Semântica. Para que as máquinas possam fazer inferências sobre os documentos da Web, é preciso que a linguagem de descrição dos documentos vá além da semântica básica definida pelo RDF Schema e consiga definir e descrever classes e propriedades sobre os objetos encontrados na Web.

## 2.1 Sintaxe da Lógica de Descrição

Nessa seção será mostrada a sintaxe básica da lógica de descrição. A tabela 2.1 mostra o alfabeto de símbolos usado pela linguagem.

Os elementos mais básicos são os conceitos atômicos e propriedades atômicas. Descrições de conceitos podem ser construídas indutivamente a partir dos construtores com conceitos e propriedades.

$C, D \rightarrow$	$A \mid$	(conceito atômico)
	$\top \mid$	(conceito universal)
	$\perp \mid$	(conceito vazio)
	$\neg A \mid$	(negação de conceito atômico)
	$C \sqcap D \mid$	(interseção de conceitos)
	$\forall R.C \mid$	(restrição de valor)
	$\exists R.\top$	(restrição existencial)

Uma interpretação  $\mathcal{I}$  consiste em um conjunto não vazio  $\Delta^{\mathcal{I}}$  (domínio da interpretação) e uma função de interpretação, que para conceito atômico  $A$  é o conjunto  $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$  e para cada propriedade atômica  $R$  é a relação binária  $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . As funções de interpretação se estendem a descrição de conceitos a partir das definições indutivas [BCM<sup>+</sup>03] como as que estão abaixo:

$$\begin{aligned}
 \top^{\mathcal{I}} &= \Delta^{\mathcal{I}} \\
 \perp^{\mathcal{I}} &= \emptyset \\
 \neg A^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus A^{\mathcal{I}} \\
 (C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
 (\forall R.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \forall b. (a, b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\} \\
 (\exists R.\top)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \exists b. (a, b) \in R^{\mathcal{I}}\}
 \end{aligned}$$

alfabeto	
a, b	indivíduos
A, B	conceitos atômicos
C, D	descrição de conceitos
R, S	papeis (propriedades)
f, g	símbolos de funções
conectivos	
$\sqcap$	interseção
$\sqcup$	união
$\neg$	negação
relações	
$\sqsubseteq$	inclusão
$\equiv$	equivalência

**Tabela 2.1** Notação da lógica de descrição

Esse trabalho é restrito à família ALC, que compreende os conceitos e propriedades atômicas, negação de conceitos, interseção, união, restrições de valor e existencial, *top* (verdade) e *bottom* (absurdo). A tabela 2.2 mostra além de ALC, outras famílias de DL existentes [BCM<sup>+</sup>03].

Uma ontologia ou base de conhecimento em ALC é composta pela tripla  $(N_C, N_R, N_O)$ , onde  $N_C$  é o conjunto de conceitos,  $N_R$  é o conjunto de predicados, e  $N_O$  é o conjunto de indivíduos, que são as instâncias de  $N_C$  e  $N_R$ . A base de conhecimento ou ontologia também pode ser descrita como o par  $(\tau, \alpha)$ , onde  $\tau$  é a terminologia do domínio (TBox), equivalente a  $N_C \cup N_R$  e  $\alpha$  é a instanciação da base, que corresponde a  $N_O$ , também conhecida como *assertional box* (ABox).

Os axiomas são compostos por elementos de  $N_O$  e um conjunto finito de GCIs (*general concept inclusions*). Podem assumir a forma  $C \sqsubseteq D$  ou  $C \equiv D$  (uma equivalência ( $\equiv$ ) é o mesmo que  $(C \sqsubseteq D) \wedge (D \sqsubseteq C)$ ), onde  $C, D$  são conceitos e  $\sqsubseteq$  é uma inclusão.

## 2.2 OWL: Web Ontology Language

A *Web Ontology Language*, OWL, foi escolhida pela w3c<sup>1</sup>, grupo que regula os padrões na Web, como a linguagem de descrição de ontologias para a Web Semântica [Hef04]. Alguns dos requisitos que ela atendeu foram [AH08]: i) sintaxe bem definida; Como o objetivo da Web Semântica é tornar os documentos da Web mais fáceis de serem processados por máquinas, este é um requisito básico. ii) semântica formal; Descrever a base de conhecimento de forma logicamente precisa é fundamental para fazer inferências como dedução de conceitos, checagem de consistência na base de conhecimento e instanciação de indivíduos a uma classe. iii) suporte a raciocínio; Uma vez que a linguagem possui uma semântica formal, atividades de

<sup>1</sup>sítio oficial: <http://w3.org>

Nome	Sintaxe	Semântica	Expressividade
Verdade	$\top$	$\Delta^I$	$AL$
Absurdo	$\perp$	$\emptyset$	$AL$
Interseção	$C \sqcap D$	$C^I \cap D^I$	$AL$
União	$C \sqcup D$	$C^I \cup D^I$	$U$
Negação	$\neg C$	$\Delta^I \setminus A^I$	$C$
Restrição de valor	$\forall R.C$	$\{a \in \Delta^I \mid \forall b. (a, b) \in R^I \rightarrow b \in C^I\}$	$AL$
Restrição existencial	$\exists R.C$	$\{a \in \Delta^I \mid \exists b. (a, b) \in R^I \wedge b \in C^I\}$	$\varepsilon$
Restrição numérica não qualificada	$\geq nR$ $\leq nR$ $= nR$	$\{a \in \Delta^I \mid  \{b \in \Delta^I \mid (a, b) \in R^I\}  \geq n\}$ $\{a \in \Delta^I \mid  \{b \in \Delta^I \mid (a, b) \in R^I\}  \leq n\}$ $\{a \in \Delta^I \mid  \{b \in \Delta^I \mid (a, b) \in R^I\}  = n\}$	$N$
Restrição numérica qualificada	$\geq nR.C$ $\leq nR.C$ $= nR.C$	$\{a \in \Delta^I \mid  \{b \in \Delta^I \mid (a, b) \in R^I \wedge b \in C^I\}  \geq n\}$ $\{a \in \Delta^I \mid  \{b \in \Delta^I \mid (a, b) \in R^I \wedge b \in C^I\}  \leq n\}$ $\{a \in \Delta^I \mid  \{b \in \Delta^I \mid (a, b) \in R^I \wedge b \in C^I\}  = n\}$	$Q$

**Tabela 2.2** Sintaxe e semântica de alguns constructos de lógica de descrição com anotação de expressividade

raciocínio podem ser realizadas. iv) expressividade; alguns domínios precisam de construtos mais elaborados para que possam ser descritos. Quanto maior a expressividade da linguagem, naturalmente fica mais fácil de descrever um domínio, apesar de aumentar a complexidade e tempo de processamento.

Entre os requisitos citados no parágrafo anterior estão expressividade e suporte a raciocínio. Apesar de ambos poderem estar na linguagem, são antagônicos, quanto maior for a expressividade da linguagem, mais complexas e demoradas serão as atividades de raciocínio sobre a linguagem. Para criar fronteiras nesse conflito entre expressividade e complexidade de raciocínio, a w3c criou três versões de OWL: OWL Full, OWL DL e OWL Lite.

### 2.2.1 OWL Full

A Web Ontology Language em sua versão mais expressiva, usando todas as primitivas da linguagem, é chamada de OWL Full. Essa combinação inclui, por exemplo, aplicar uma restrição de cardinalidade na classe que contém todas as outras classes, limitando a quantidade de classes que a ontologia pode ter.

OWL Full é completamente compatível com RDF, tanto sintaticamente, quanto em sua semântica. A desvantagem de OWL Full é que ela é tão poderosa que é indecidível em relação às atividades de raciocínio.

### 2.2.2 OWL DL

OWL DL (DL é a sigla para *Description Logic*, Lógica de Descrição em português) é a família de OWL que corresponde à lógica de Descrição. A sua grande vantagem é que ela é decidível, dando a possibilidade de realização de atividades de raciocínio de forma mais eficiente. A

desvantagem de OWL DL é que ela perde a compatibilidade com RDF, qualquer documento em OWL DL pode ser descrito como um documento em RDF, mas o contrário não é verdade.

### **2.2.3 OWL Lite**

OWL Lite é uma família de OWL que é mais limitada do que OWL Full e OWL DL. Ela não dá suporte a, por exemplo, disjunção entre classes, união e complemento. A grande vantagem dessa linguagem é uma maior facilidade para o desenvolvimento de ferramentas, e a sua desvantagem é a perda de expressividade.

## Normalização para o método das conexões

O método das conexões proposto por W. Bibel [Bib82] é um método para prova automática de teoremas descritos em lógica de primeira ordem ou em lógica proposicional [BH93]. Um dos trabalhos recentes de Freitas et al [FSS10] foi a extensão desse método para lógica de descrição *ALC*.

O artigo intitulado *A Connection Method for Reasoning with the Description Logic ALC* [FSS10] propõe algoritmos tanto para o método, quanto para a normalização que precisa ser feita na base de conhecimento para que seja possível a representação necessária para o método das conexões usando apenas uma matriz.

O objetivo deste trabalho é implementar algum algoritmo de normalização para o método das conexões como os citados no texto de Freitas et al. Dois algoritmos foram propostos com esse objetivo [FSS10]; o primeiro utiliza-se de uma tabela com nove regras que devem ser aplicadas à base de conhecimento a fim de obter a forma normal positiva. O segundo, intitulado "*A more complex and efficient normalization*" não cria novos símbolos durante a sua execução, fazendo-o mais eficiente que o primeiro em relação ao uso de memória.

No cronograma deste trabalho estava prevista a implementação desse segundo algoritmo, porém, ao decorrer do desenvolvimento, o orientando propôs um terceiro algoritmo que é ainda mais eficiente em relação ao uso de memória, mas com o efeito colateral de precisar adicionar um símbolo a mais na base de conhecimento no pior caso. O restante desse capítulo se dedicará a dar definições para o entendimento dos últimos dois algoritmos comentados acima e também descreverá as suas implementações.

### 3.1 Tradução de ontologias *ALC* para forma normal disjuntiva

Para que o leitor consiga entender melhor os algoritmos de tradução, alguns conceitos precisam ser fixados.

Métodos diretos com o método das conexões são formulados para provar que uma fórmula ou um conjunto de fórmulas é um teorema, se e somente se cada interpretação gerada é uma tautologia. Tautologias normalmente tomam a forma  $L \vee \neg L$ , nesse caso, a fórmula precisa estar na Forma Normal Disjuntiva (DNF).

**Definição 1** (Forma Normal Disjuntiva, cláusula). *Uma fórmula em DNF é uma conjunção de disjunções. Ou seja, tomam a forma:*

$$\bigcup_{i=1}^n C_i, \text{ ou, } C_1 \vee \dots \vee C_n.$$

onde cada  $C_i$  é uma cláusula. Uma cláusula é uma conjunção de literais. Ou seja, tomam a forma:

$$\bigcap_{j=1}^m L_{i,j}, \text{ ou, } L_{i,1} \wedge \dots \wedge L_{i,m}, \text{ também representado por } \{L_{i,1}, \dots, L_{i,m}\}$$

onde cada  $L_{i,j}$  é um literal, resultando na fórmula:

$$\bigcap_{i=1}^n \bigcup_{j=1}^m L_{i,j}, \text{ ou, } (L_{1,1} \wedge \dots \wedge L_{1,m}) \vee (L_{n,1} \wedge \dots \wedge L_{n,m})$$

podendo ser chamada também de forma causal disjuntiva, representada por:

$$\{\{(L_{1,1}, \dots, L_{1,m}), \dots, (L_{n,1}, \dots, L_{n,m})\}\}$$

A definição acima é a definição herdada da lógica de primeira ordem, para ser válida também para a lógica de descrição o conceito de conjunções e disjunções deve ser estendido.

**Definição 2** (Conjunção ALC). Uma conjunção ALC é um literal  $L$ , uma conjunção  $(E_0 \wedge, \dots, \wedge E_n)$ , ou uma restrição existencial  $\exists x.E$ , onde  $E$  é uma expressão qualquer em lógica de descrição.

**Definição 3** (Disjunção ALC). Uma disjunção ALC é um literal  $L$ , uma disjunção  $(E_0 \vee, \dots, \vee E_n)$ , ou uma restrição de valor  $\forall x.E$ , onde  $E$  é uma expressão qualquer em lógica de descrição.

**Definição 4** (Conjunção ALC pura, Conjunção ALC não pura). Uma conjunção ALC pura é uma conjunção ALC que na forma normal negada não contém restrições de valor  $(\forall x.E)$  e também não contém disjunções  $(E \vee, \dots, \vee E)$ . O conjunto de conjunções ALC puras é representado por  $\hat{C}$ . Uma conjunção ALC não pura é uma conjunção ALC que não é pura.

**Definição 5** (Disjunção ALC pura, Disjunção ALC não pura). Uma disjunção ALC pura é uma disjunção ALC que na forma normal negada não contém restrições existenciais  $(\exists x.E)$  e também não contém conjunções  $(E \wedge, \dots, \wedge E)$ . O conjunto de disjunções ALC puras é representado por  $\check{D}$ . Uma disjunção ALC não pura é uma disjunção ALC que não é pura.

**Definição 6** (Impureza em uma expressão não pura). Impureza em expressões ALC não puras são conjunções em disjunções não puras ou disjunções em conjunções não puras. O conjunto de impurezas é chamado de conjunto de impurezas ALC e é representado por  $I$ .

**Definição 7** (Forma Normal Positiva). Um axioma ALC está na Forma Normal Positiva sse ele está em uma das seguintes formas:

- i)  $A \sqsubseteq \hat{C}$
- ii)  $\check{D} \sqsubseteq A$
- iii)  $\hat{C} \sqsubseteq \check{D}$

onde  $A$  é um conceito atômico,  $\hat{C}$  é uma conjunção ALC pura,  $\check{D}$  é uma disjunção ALC pura.



O método das conexões utiliza matrizes para realizar provas de teoremas. No início deste trabalho, ainda não era possível de fazer as provas com matrizes aninhadas, ou seja, havia sempre a necessidade de normalizar a base de conhecimento na forma normal positiva (definição 7). No entanto, Jens Otten em um trabalho intitulado *A Non-clausal Connection Calculus* [Ott11] mostrou como aplicar o método das conexões sem o passo da normalização. Apesar da recente evolução do método, para o objetivo deste trabalho, ainda é necessário fazer a normalização, já que para lógica de descrição, o método ainda não foi modificado para usar matrizes aninhadas.

A próxima seção desta monografia descreve o algoritmo proposto pelo autor para a normalização para a forma normal positiva.

### 3.1.1 Algoritmo Proposto

Esta seção da monografia descreve a implementação realizada neste trabalho.

O método das conexões é um método direto, ou seja, uma consulta à base de conhecimento toma verifica se uma fórmula é uma tautologia e toma a forma  $KB \rightarrow \alpha$ , onde  $\alpha$  é um axioma e  $KB$  (*Knowledge base*) é da forma  $\bigcap_{i=1}^n A_i$ , onde  $A_i$  também é um axioma. Expandindo a fórmula  $\neg KB \vee \alpha$ , temos:

$\neg \bigcap_{i=1}^n A_i \vee \alpha$  [ou,  $\neg(A_1 \wedge \dots \wedge A_n) \vee \alpha$ ], que pode ser transformada para:

$\bigcup_{i=1}^n \neg A_i \vee \alpha$  [ou,  $\neg A_1 \vee \dots \vee \neg A_n \vee \alpha$ ]

Cada  $\neg A_i$  ou  $\alpha$  precisam estar na forma normal positiva, ou seja, precisam estar em uma das formas: i)  $A \sqsubseteq \hat{C}$ , ii)  $\check{D} \sqsubseteq A$ , ou, iii)  $\hat{C} \sqsubseteq \check{D}$ , onde  $A$  é um conceito,  $\hat{C}$  é uma conjunção pura e  $\check{D}$  é uma disjunção pura.

O primeiro passo do algoritmo é a separação de axiomas de equivalência de expressões e inclusões. Os axiomas de equivalência serão substituídos por dois axiomas de inclusão, e.g.,  $(A \equiv B \rightarrow A \sqsubseteq B \wedge B \sqsubseteq A)$ .

---

**Algorithm 1** Início do processo de normalização para toda a base de conhecimento

---

```

for all  $A \equiv B \in S_{EQ}$  do {  $A, B$  são expressões,  $S_{EQ}$  é o conjunto dos axiomas de equivalência }
    Normalize-Axiom ( $A \sqsubseteq B$ )
    Normalize-Axiom ( $B \sqsubseteq A$ )
end for
for all  $A \sqsubseteq B \in S_I$  do {  $A, B$  são expressões,  $S_I$  é o conjunto dos axiomas de inclusão }
    Normalize-Axiom ( $A \sqsubseteq B$ )
end for

```

---

A chamada ao método *Normalize-Axiom()* remove as impuresas do lado esquerdo e do lado direito de cada axioma.

O método *purify(A)* indentifica se  $A$  (expressão em DL) é uma conjunção, disjunção ou conceito. Caso  $A$  seja uma conjunção não pura, as impuresas são removidas, caso contrário,  $A$  permanece com a estrutura original. De forma análoga, caso  $A$  seja uma disjunção não pura, as impuresas serão removidas e  $A$  passará a ser pura. E o último caso, caso  $A$  seja um conceito, ele não será modificado.

**Algorithm 2** Normalize-Axiom ( $A \sqsubseteq B$ )

---

```

 $O = \{O - (A \sqsubseteq B)\}$ 
 $pure\_left = \text{purify}(A)$ 
 $pure\_right = \text{purify}(B)$ 
if ( $pure\_left \sqsubseteq pure\_right$ )  $\in S_{PNF}$  then  $\{S_{PNF}$  é o conjunto das fórmulas na forma normal positiva $\}$ 
     $O = \{O \sqcup (pure\_left \sqsubseteq pure\_right)\}$ 
else
     $O = \{O \sqcup (pure\_left \sqsubseteq N)\}$ 
     $O = \{O \sqcup (N \sqsubseteq pure\_right)\}$ 
end if

```

---

expressão	regra
$\hat{C}_1 \sqsubseteq \hat{C}_2$	$(\hat{C}_1 \sqsubseteq A') \wedge (A' \sqsubseteq \hat{C}_2)$
$\check{D}_1 \sqsubseteq \check{D}_2$	$(\check{D}_1 \sqsubseteq A') \wedge (A' \sqsubseteq \check{D}_2)$
$\check{D} \sqsubseteq \hat{C}$	$(\check{D} \sqsubseteq A') \wedge (A' \sqsubseteq \hat{C})$

**Tabela 3.1** Tabela de regras a serem aplicadas após o método purify

O *if* do método *Normalize-Axiom* () verifica se o axioma  $A \sqsubseteq B$  está na forma normal positiva, caso não esteja, as expressões são separadas em dois axiomas para respeitar a tabela 3.1.

## CAPÍTULO 4

# leanCoP

O leanCoP é um provador automático de teoremas para lógica de primeira ordem escrito em Prolog [OB03]. Em testes realizados com a biblioteca TPTP [Sut09], o leanCoP mostrou boa performance, mesmo comparado a provadores de teoremas que são estado da arte. Dentre os 2200 problemas inclusos na biblioteca TPTP, o leanCoP foi capaz de resolver 667 (30.3%), 532 problemas a mais que o leanTAP [BP94] e 932 a menos que o Otter [Mcc95].

O objetivo do uso do leanCoP neste trabalho é validar o algoritmo de normalização descrito em 3.1.1, caso o leanCoP consiga ser usado com a base de conhecimento descrita em *ALC*, significa que os axiomas da base estão na forma normal positiva.

É importante notar que a ontologia OWL, documento que contém os axiomas da base de conhecimento, é descrita em lógica de descrição. Algumas ferramentas se mostraram necessárias para que essa base em OWL fosse usada com o leanCoP. Primeiro, a ontologia é convertida para o formato TPTP a partir da biblioteca owlapi 1.0 <sup>1</sup>, o arquivo gerado passa então por uma transformação para que use a linguagem de lógica de primeira ordem do TPTP. Após essa transformação, os axiomas da base podem ser utilizadas pelo leanCoP.

Os códigos abaixo mostram exemplos da linguagem usada pelo leanCoP:

```
fof(axiom_13, axiom, (
    ! [X] :
      ( iAB(X)
    <=> ( iA(X)
        | iB(X) ) ) ) ).
```

O exemplo acima é um axioma expresso em lógica de primeira ordem que é equivalente a:

$$\forall x(AB(x) \iff A(x) \vee B(x))$$

A owlapi, biblioteca que transforma de OWL para TPTP, define em lógica de primeira ordem alguns conceitos básicos que são necessários para que os teoremas separem o domínio das instâncias de classes do domínio dos tipos pre-definidos em OWL. o exemplo de código abaixo mostra três dos onze axiomas que são gerados pela biblioteca com esse objetivo:

```
fof(axiom_0, axiom, (
    ! [X] :
      ( abstractDomain(X)
```

---

<sup>1</sup>download: <http://sourceforge.net/projects/owlapi/files/OWL%20API%20%28OWL%201.0%29/>

```

| dataDomain(X) ) ) ).

fof(axiom_1, axiom, (
  ? [X] : abstractDomain(X) ) ) ).

fof(axiom_2, axiom, (
  ? [X] : dataDomain(X) ) ) ).

```

O primeiro axioma é igual a  $\forall x(\text{abstractDomain}(x) \vee \text{dataDomain}(x))$ , ou seja, todos os elementos do domínio ou são *abstractDomain* ou *dataDomain*. O segundo axioma é igual a  $\exists x(\text{abstractDomain}(x))$ , ou seja, o domínio *abstractDomain* não é vazio. E o terceiro axioma do exemplo é igual a  $\exists x(\text{dataDomain}(x))$ , ou seja, o domínio *dataDomain* também não é vazio. Abaixo estão todos os axiomas que são básicos para todas as bases:

$$\begin{aligned}
& \forall x(\text{abstractDomain}(x) \vee \text{dataDomain}(x)) \\
& \exists x(\text{abstractDomain}(x)) \\
& \exists x(\text{dataDomain}(x)) \\
& \forall x \neg(\text{abstractDomain}(x) \wedge \text{dataDomain}(x)) \\
& \forall x(\text{owlThing}(x) \implies \text{abstractDomain}(x)) \\
& \forall x(\text{owlNothing}(x) \implies \text{abstractDomain}(x)) \\
& \forall x(\text{abstractDomain}(x) \implies \text{owlThing}(x)) \\
& \forall x \neg(\text{owlNothing}(x)) \\
& \forall x(\text{xsd\_string}(x) \implies \text{dataDomain}(x)) \\
& \forall x(\text{xsd\_integer}(x) \implies \text{dataDomain}(x)) \\
& \forall x(\text{dataDomain}(x) \implies \neg(\text{xsd\_string}(x) \wedge \text{xsd\_integer}(x)))
\end{aligned}$$

Para fazer alguma atividade de raciocínio a partir dessa base, um número exponencial de consultas deveria ser gerado. Subsunsão, por exemplo, consistirá em perguntar se cada conceito pode ser sub-classe de outro conceito.

Devido a essa solução ser muito ineficiente, devido à quantidade transformações de arquivos que cada consulta deveria gerar, essa atividade de consulta acaba fugindo do escopo desse projeto, que é a implementação do algoritmo de normalização para o Método das Conexões com ontologias *ALC*.

## Conclusão

In order to describe the connection method as a formal inference system (in section 4), and the positive matricial normal form used in it, we will briefly describe the notation we use for first order logic, before examining the method. We are presuming readers to be acquainted to first order logic.

### 5.1 Definition 1 (First-order logic syntax).

The alphabets given by table 1 compose the FOL syntax notation. ? Table 1. First order logic syntax notation.



## Referências Bibliográficas

- [AH08] Grigoris Antoniou and Frank van Harmelen. *A Semantic Web Primer, 2nd Edition (Cooperative Information Systems)*. The MIT Press, 2 edition, 2008.
- [BCM<sup>+</sup>03] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, Cambridge, 2003.
- [Bez95] Jeff Bezos. amazon.com, 1995.
- [BH93] W. Bibel and S. Hölldobler. *Deduction: automated logic*. Academic Press, 1993.
- [Bib82] W. Bibel. *Automated theorem proving*. F. Vieweg, 1982.
- [BL89] Tim Berners-Lee. Information management: A proposal, 1989.
- [BM11] Kai Brunnler and George Metcalfe, editors. *Automated Reasoning with Analytic Tableaux and Related Methods - 20th International Conference, TABLEAUX 2011, Bern, Switzerland, July 4-8, 2011. Proceedings*, volume 6793 of *Lecture Notes in Computer Science*. Springer, 2011.
- [BP94] B. Beckert and J. Posegga. Leantap : Lean tableau-based theorem proving. 1994.
- [FSS10] Fred Freitas, Anne Schlicht, and Heiner Stuckenschmidt. A connection method for reasoning with the description logic alc. Technical report, UFPE and University of Mannheim, 2010.
- [Gal99] Marcos Galperin. mercadolive.com.br, 1999.
- [Hef04] Jeff Heflin. Owl web ontology language use cases and requirements. 2004.
- [McC95] W. Mccune. The otter theorem prover. 1995.
- [OB03] Jens Otten and Wolfgang Bibel. leancop: lean connection-based theorem proving. *J. Symb. Comput.*, 36(1-2):139–161, 2003.
- [Ott11] Jens Otten. A non-clausal connection calculus. In Brunnler and Metcalfe [BM11], pages 226–241.
- [RBML99] Romero Rodrigues, Rodrigo Borges, Ronaldo Morita, and Mario Letelier. buscape.com.br, 06 1999.

- [RN02] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach (2nd Edition)*. Prentice Hall series in artificial intelligence. Prentice Hall, 2 edition, December 2002.
- [Sic] Carlos Alberto Sicupira. americanas.com.
- [Sut09] G. Sutcliffe. The TPTP Problem Library and Associated Infrastructure: The FOF and CNF Parts, v3.5.0. *Journal of Automated Reasoning*, 43(4):337–362, 2009.