

4

Estudo de Caso

4.1.

Introdução

Para demonstrar a estratégia implementada no Componente para Alinhamento Taxonômico de Ontologias (CATO), são detalhados três estudos de caso. Para cada um destes estudos de caso apresentados existe um par de ontologias complementares a ser alinhado pelo CATO. O primeiro par trata-se de ontologias específicas de domínios de aplicação. O segundo par trata-se de parte de uma ontologia genérica alinhada com uma ontologia específica de aplicação. O terceiro par trata-se de partes específicas de um mesmo assunto de ontologias genéricas.

Apesar do objetivo do CATO ser alinhar ontologias específicas de domínios de aplicação, os segundo e terceiro estudos de caso foram realizados com o intuito de avaliar o seu comportamento frente ao processamento de ontologias com um grande número de termos e estruturas hierárquicas bem diferentes.

As seis ontologias dos estudos de caso foram escolhidas depois de alguma pesquisa na Internet. Algumas localizações foram visitadas como: o *site* de ontologias publicadas do Google (Google, 2004) com aproximadamente dez links para novas fontes de ontologias, o diretório de pesquisa do SchemaWeb (SchemaWeb, 2004a) com aproximadamente cento e cinquenta ontologias, a biblioteca de ontologias *DAML - Darpa Agent Markup Language*, (DAML ontology library, 2004) com aproximadamente duzentos e oitenta ontologias, entre outros.

Apesar da quantidade de ontologias disponibilizadas nas localizações visitadas, poucas ontologias puderam ser analisadas por não serem de fontes conhecidas ou por apresentarem problemas em sua construção e, conseqüentemente, não ser possível sua utilização nos editores de ontologias ou por outros fatores como: *links* quebrados, repetições da mesma ontologia em mais

de uma localização, qualidade da ontologia, entre outros. A estratégia de seleção das ontologias é explicada no próximo tópico.

Neste capítulo sobre o estudo de caso, a palavra **classe** será utilizada como sinônimo da palavra **conceito**, que é utilizada ao longo deste trabalho, devido ao fato das nomenclaturas das *tags* em *OWL* utilizarem esta primeira palavra e, desta maneira, o entendimento do texto ser facilitado.

4.2.

Estratégia de Seleção das Ontologias

A escolha das ontologias utilizadas nos estudos de caso não envolveu interesse específico algum, fora o de pesquisa. Foram priorizadas ontologias disponibilizadas na linguagem de ontologias *OWL*, de instituições conhecidas e contextualizadas em domínios específicos de aplicações, as *Web Ontologies*. Escolhida uma ontologia de um domínio específico de aplicação, era preciso escolher uma outra ontologia, de domínio complementar ao da primeira, com classes equivalentes entre elas a serem alinhadas. As duas ontologias escolhidas também não deveriam ser criadas pelos mesmos autores. Com estas restrições, o número de possibilidades diminuía com cada requisito de escolha a ser satisfeito.

Por exemplo, as ontologias organizadas por palavras-chaves do diretório de ontologias do *DAML* (*DAML ontology library*, 2004) foram analisadas. Neste momento, o interesse de pesquisa era encontrar duas ontologias de aplicações complementares, disponibilizadas por grupos diferentes e, se possível, já escritas em *OWL*, pois, se estivessem escritas em *DAML*, a transformação para *OWL* seria necessária. A ontologia *CMU RI Employment Categories* (CMU, 2004e) da Universidade Carnegie Mellon foi encontrada referenciada pela palavra-chave *Academic Positions* e escolhida como a segunda ontologia do segundo estudo de caso. Feita esta escolha, a próxima ontologia a ser escolhida deveria ser de um domínio complementar ao domínio de categorias de empregos. A ontologia formada pelas classes do tipo pessoa cadastradas no thesaurus WordNet (Fellbaum, 2004) encontrada no diretório de pesquisa do SchemaWeb (SchemaWeb, 2004a) foi a ontologia complementar escolhida. A escolha destas duas ontologias baseou-se na vontade de analisar os resultados do alinhamento de parte de uma ontologia genérica, constituída por muitas classes, com uma ontologia específica.

Já no primeiro estudo de caso, o alinhamento foi realizado com duas ontologias específicas. A primeira ontologia escolhida, a *CMU RI Publications* (CMU, 2004d), também da Universidade Carnegie Mellon, é alinhada com a segunda ontologia escolhida, a *General University Ontology* (Mondeca, 2004c), de uma empresa comercial pertencente ao grupo de trabalho da *Web-Ontology* da W3C (W3CWebOntWorkingGroup, 2004).

No terceiro estudo de caso, partes específicas de um mesmo assunto foram extraídas de duas ontologias genéricas e estas partes foram comparadas na tentativa do alinhamento. As classes referentes à classe “Meio de Transporte” das ontologias genéricas SUMO (SUMO, 2004) e OpenCYC (OpenCYC, 2004) constituem as ontologias comparadas a serem alinhadas. Estas ontologias possuem estruturas hierárquicas bem diferentes.

Para cada um dos três estudos de caso escolhidos é utilizada a seguinte ordem de apresentação:

1. Resultado da Primeira Etapa;
2. Resultados das Etapas sem Ordenação Alfabética;
 - 2.1. Resultado da Segunda Etapa;
 - 2.2. Resultado da Terceira Etapa;
 - 2.3. Resultado do Alinhamento;
3. Resultados das Etapas com Ordenação Alfabética;
 - 3.1. Resultado da Segunda Etapa;
 - 3.2. Resultado da Terceira Etapa;
 - 3.3. Resultado do Alinhamento;
4. Avaliação dos Resultados;
5. Problemas Encontrados.

Vale ressaltar que os resultados das segunda e terceira etapas e do alinhamento propriamente dito são influenciados pelos dois módulos implementados do CATO, o sem ordenação alfabética, i.e., as classes e subclasses das ontologias comparadas são estruturadas com sua ordem original de criação, e o com ordenação alfabética, i.e., as classes e subclasses das ontologias comparadas são estruturadas ordenadas alfabeticamente. Devido a possíveis diferenças encontradas nos dois módulos implementados, os resultados são apresentados tanto no tópico dos resultados sem ordenação alfabética quanto no tópico dos resultados com ordenação alfabética.

4.3. Primeiro Estudo de Caso

Duas ontologias independentes, criadas por diferentes grupos, foram escolhidas como exemplo para o primeiro estudo de caso.

O primeiro grupo escolhido é o do projeto *Agent Transaction Language for Advertising Services – ATLAS* (ATLAS, 2004). O ATLAS é um projeto de pesquisa do grupo de agentes de software (CMU, 2004c) do Instituto de Robótica (CMU, 2004b) da Universidade Carnegie Mellon (CMU, 2004a), com a colaboração do Centro de Pesquisa da empresa Nokia (Nokia, 2004) e do Centro de Pesquisa Alemão para Inteligência Artificial (DFKI, 2004).

A ontologia de publicações do ATLAS, a *CMU RI Publications* (CMU, 2004d), é a primeira ontologia escolhida. A Figura 22 ilustra suas hierarquias em *DAML* (árvore da esquerda da figura)¹¹ e em *OWL* (árvore da direita da figura).

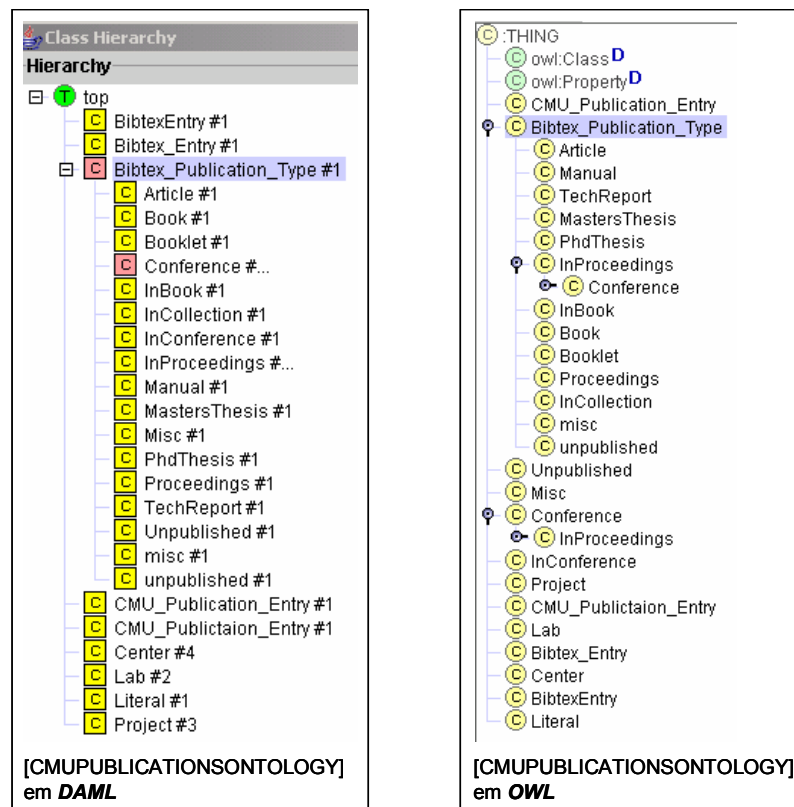


Figura 22 – Representações da ontologia de publicações escolhida

As ontologias do ATLAS são escritas na linguagem *DAML*. Por esta razão, a conversão de suas ontologias de *DAML* para *OWL* é necessária porque o CATO

¹¹ Os ícones das classes “Bibtex_Publication_Type” e “Conference” estão assinalados diferentemente na hierarquia em *DAML* da Figura 22 porque o editor de ontologias OilEd faz essa distinção como indicativo de classes iguais (SameClassAs, em *DAML*) cadastradas na ontologia.

recebe como entradas ontologias escritas em *OWL*. Tal conversão pode ser realizada por editores de ontologias que tenham esta funcionalidade, como o editor OilEd (OilEd, 2004) utilizado.

O endereço do código *OWL* dessa ontologia de publicações escolhida é disponibilizado no Anexo A deste documento. O código original em *DAML* é disponibilizado em (CMU, 2004d).

O segundo grupo escolhido para este primeiro estudo de caso é o da empresa francesa Mondeca SA (Mondeca, 2004a), uma das empresas do grupo de trabalho da *Web-Ontology* da W3C (W3CWebOntWorkingGroup, 2004).

No ano de 2003, a empresa foi uma das ganhadoras do prêmio *European Information Society Technologies* (IST-PRIZE, 2004). Este prêmio é o mais distinto para produtos e serviços inovadores no campo de tecnologias da sociedade de informação. O prêmio está aberto às companhias ou às organizações que apresentam um produto de tecnologia inovadora com um mercado potencial promissor.

Um dos produtos para gerência da semântica do conhecimento da empresa Mondeca, o ITM – *Intelligent Topic Manager* – (Mondeca, 2004b), ilustrado na Figura 23, faz uso de ontologias. Estas ontologias incluem definições de classes, tipos associados, relações, tipos de dados e restrições que garantem a coerência de seus termos.

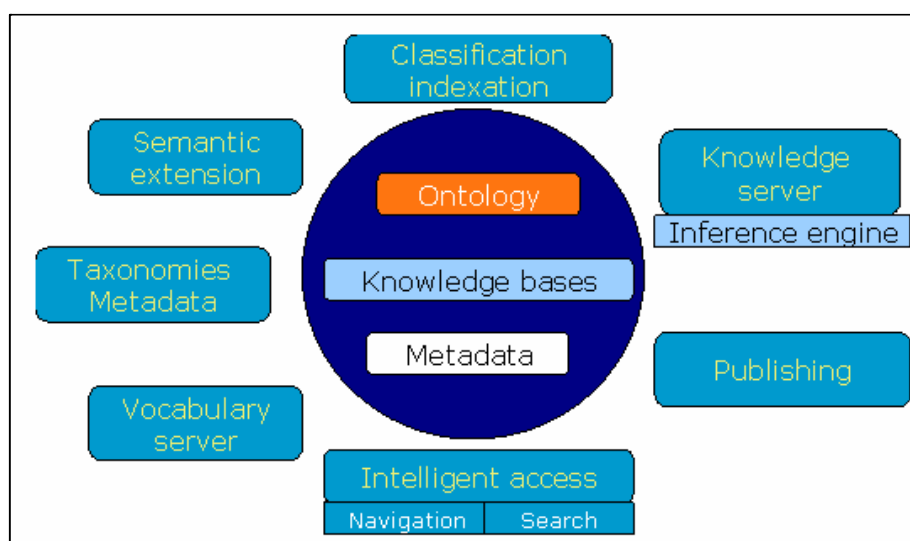


Figura 23 – Gerenciamento de Conhecimento no ITM

Uma das ontologias utilizada no ITM, a *General University Ontology* (Mondeca, 2004c), é a segunda ontologia escolhida para este estudo de caso. O endereço de seu código *OWL* é disponibilizado no Anexo A deste documento.

As ontologias comparadas neste primeiro estudo de caso podem ser classificadas como ontologias complementares específicas do domínio de publicação. A Figura 24 ilustra parte das hierarquias destas ontologias comparadas.

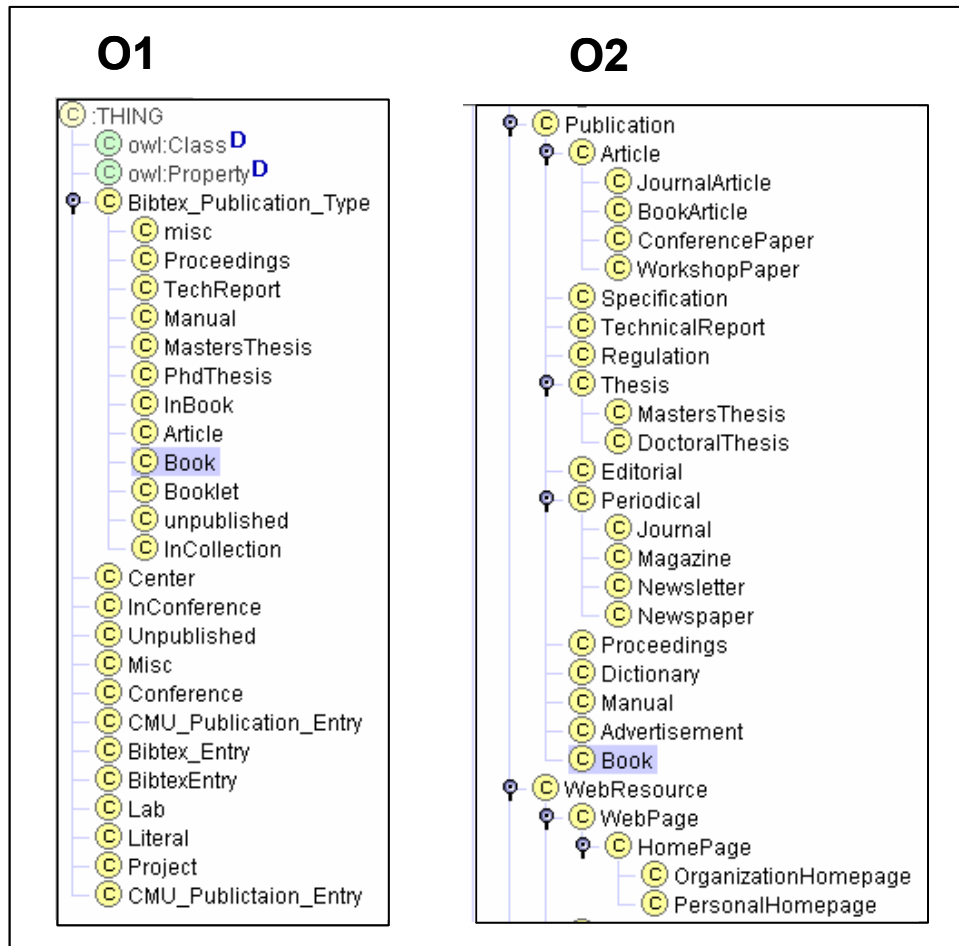


Figura 24 – Ontologias comparadas

Como as ontologias foram criadas por grupos distintos, é de se esperar que possuam diferenças tanto lexicais quanto estruturais. Por exemplo, a classe “PhdThesis” da primeira ontologia escolhida (árvore da esquerda na Figura 24) tem diferenças lexical e estrutural em relação à classe “DoctoralThesis” da segunda ontologia escolhida (árvore da direita na Figura 24). A diferença lexical é porque as classes são escritas com diferentes nomes e a diferença estrutural é porque possuem diferentes superclasses como, por exemplo, sua classe pai (um nível hierárquico acima). No entanto, a análise destas diferenças por um usuário classificaria, facilmente, as classes exemplificadas como equivalentes.

4.3.1. Resultado da Primeira Etapa

Inicialmente, como não existem cadastrados os sinônimos das classes das ontologias comparadas e as classes iguais lexicalmente possuem diferenças estruturais, portanto, não satisfazem a condição de poda da primeira etapa, a conclusão da primeira etapa de execução do CATO não resulta em alinhamento algum.

Para tentar algum alinhamento, os sinônimos ilustrados entre colchetes na Figura 25 foram cadastrados no banco de dados utilizado e a primeira etapa do CATO foi executada novamente. Nesta nova execução, as classes “TechReport” de **O1** e “TechnicalReport” de **O2** foram identificadas como sinônimos e, por isso, são comparadas. Como estas classes não satisfazem a condição de poda porque possuem classes pais diferentes, um nível hierárquico acima, não são alinhadas. O mesmo acontece com as classes “PhdThesis” de **O1** e “DoctoralThesis” de **O2**. Assim, a conclusão da primeira etapa de execução do CATO com o uso de sinônimos também não resulta em alinhamento algum.

```
Sinonimos de file:firstOnto.owl#TechReport: [TechnicalReport]
Sinonimos de file:firstOnto.owl#PhdThesis: [DoctoralThesis]
Sinonimos de file:secondOnto.owl#TechnicalReport: [TechReport]
Sinonimos de file:secondOnto.owl#DoctoralThesis: [PhdThesis]
```

Figura 25 – Sinônimos cadastrados identificados

Vale lembrar que as informações dos sinônimos identificados só são propagadas para as demais etapas de execução do CATO se satisfizerem às condições de alinhamento da primeira etapa. Como nesse estudo de caso elas não satisfazem tais condições, os sinônimos não são utilizados e, conseqüentemente, seu uso não resulta impacto algum no alinhamento final das ontologias.

4.3.2. Resultados das Etapas sem Ordenação Alfabética

Os sub-tópicos a seguir mostram os resultados das etapas executadas no CATO com as hierarquias das árvores comparadas sem ordenação alfabética, onde as classes e subclasses das ontologias são estruturadas com sua ordem original de criação.

4.3.2.1.

Resultado da Segunda Etapa (sem ordenação)

A Figura 26 ilustra as hierarquias das duas ontologias comparadas e seus grupos de equivalência identificados.

As classes “Bibtex_Publication_Type” de **O1** e “Publication” de **O2** são identificadas como *similares* porque suas subclasses “Proceedings” possuem igualdade lexical e similaridade estrutural. Assim, os grupos de equivalência, identificados pelos círculos superiores na Figura 26, são limitados por estas respectivas classes. Fechados os grupos de equivalência, todas as classes dentro destes, iguais lexicalmente e estruturalmente, são identificadas como *similares*. Isto acontece com as classes “Proceedings”, “Book” e “Manual”, de ambas as ontologias.

No entanto, algumas classes idênticas lexicalmente como, por exemplo, as classes “MastersThesis”, de ambas ontologias, pertencentes aos grupos de equivalência superiores na Figura 26, não são identificadas como *similares* porque possuem diferenças estruturais entre elas e próximas a elas, o que dificulta tanto a identificação de sua similaridade lexical quanto a formação de novos grupos de equivalência. Em **O1**, a classe “MastersThesis” é subclasse de “Bibtex_Publication_Type” e em **O2**, é subclasse de “Thesis” que, por sua vez, é subclasse de “Publication”. Como as classes “Bibtex_Publication_Type” de **O1** e “Publication” de **O2** foram identificadas como *similares*, então, existe a diferença de um nível hierárquico entre as classes “MastersThesis” em **O1** e em **O2** e, por isso, não há a identificação de similaridade destas classes.

Pela igualdade lexical das classes “Conference”, em ambas as ontologias, e sua similaridade estrutural, indicada pela nova classe em **O1** e pela nova subclasse em **O2**, os próximos grupos de equivalência, representados pelos círculos inferiores na Figura 26, são identificados e suas classes são classificadas como *similares*.

O resultado final desta etapa de execução são as seguintes classes identificadas como *similares*: “Proceedings”, “Book”, “Manual” e “Conference”, de ambas as ontologias, e “Bibtex_Publication_Type” de **O1** *similar* à classe “Publication” de **O2**.

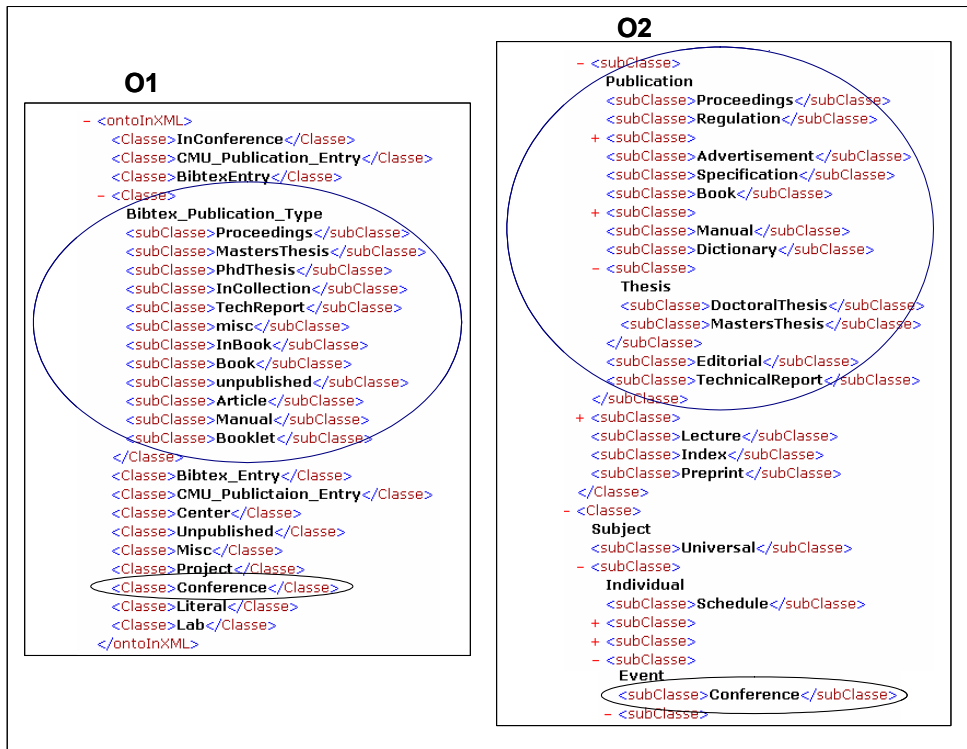


Figura 26 – Grupos de equivalência identificados no módulo sem ordenação

4.3.2.2.

Resultado da Terceira Etapa (sem ordenação)

A terceira etapa de execução do CATO calcula os percentuais de similaridade das classes identificadas como *similares* na etapa anterior. A Figura 27 ilustra os resultados de tais cálculos.

```

Similarities Level:
Bibtex_Publication_Type -> Publication    *** Similarity Level: 23.076923%
Conference -> Conference    *** Similarity Level: 100.0%
Proceedings -> Proceedings    *** Similarity Level: 100.0%
Book -> Book    *** Similarity Level: 100.0%
Manual -> Manual    *** Similarity Level: 100.0%
    
```

Figura 27 – Percentuais de similaridade calculados

4.3.2.3.

Resultado do Alinhamento (sem ordenação)

Ao concluir as três etapas de execução, o CATO identifica as classes equivalentes, i.e., aquelas que satisfazem todas as condições de cada uma de suas etapas, e realiza o alinhamento. Neste estudo de caso, com as árvores estruturadas na sua ordem original de criação, as classes “Conference”, ”Proceedings”, “Book” e “Manual” são as classes alinhadas pelo CATO. Isto porque são identificadas como *similares* pela sua segunda etapa de execução e classificadas como *bem*

similares, com um percentual de similaridade maior ou igual a 75%, pela sua terceira etapa de execução.

A Figura 28 ilustra parte da ontologia final do CATO com a classe “Book”, original da primeira ontologia, identificada pelo *namespace* “file:firstOnto.owl”, com a informação adicionada da classe equivalente identificada, a classe “Book”, original da segunda ontologia, identificada pelo *namespace* “file:secondOnto.owl”.

```

- <owl:Class rdf:about="file:firstOnto.owl#Book">
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Manual" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Proceedings" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#InConference" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Conference" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#MastersThesis" />
</owl:disjointWith>
<owl:disjointWith rdf:resource="file:firstOnto.owl#TechReport" />
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#InBook" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#PhdThesis" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Misc" />
</owl:disjointWith>
<rdfs:subClassOf rdf:resource="file:firstOnto.owl#Bibtex_Publication_Type" />
- <owl:equivalentClass>
  <owl:Class rdf:about="file:secondOnto.owl#Book" /> } ← Informação Adicionada
</owl:equivalentClass>
<owl:disjointWith rdf:resource="file:firstOnto.owl#Unpublished" />
<owl:disjointWith rdf:resource="file:firstOnto.owl#Booklet" />
</owl:Class>

```

Figura 28 – Informação adicionada, resultado do alinhamento com o CATO

4.3.3.

Resultados das Etapas com Ordenação Alfabética

Os sub-tópicos a seguir mostram os resultados das etapas executadas no CATO com as hierarquias das árvores comparadas estruturadas com ordenação alfabética, i.e., as classes e subclasses das ontologias são estruturadas ordenadas alfabeticamente.

4.3.3.1.

Resultado da Segunda Etapa (com ordenação alfabética)

A Figura 29 ilustra as hierarquias das duas ontologias comparadas e seus grupos de equivalência identificados.

As classes “Bibtex_Publication_Type” de O1 e “Publication” de O2 são identificadas como *similares* porque suas subclasses “Book” possuem igualdade

lexical e similaridade estrutural. Assim, os grupos de equivalência, identificados pelos círculos superiores na Figura 29, são limitados por estas respectivas classes. Fechados os grupos de equivalência, todas as classes dentro destes, iguais lexicalmente e estruturalmente, são identificadas como *similares*. Isto acontece com as classes “Book” e “Proceedings”, de ambas as ontologias.

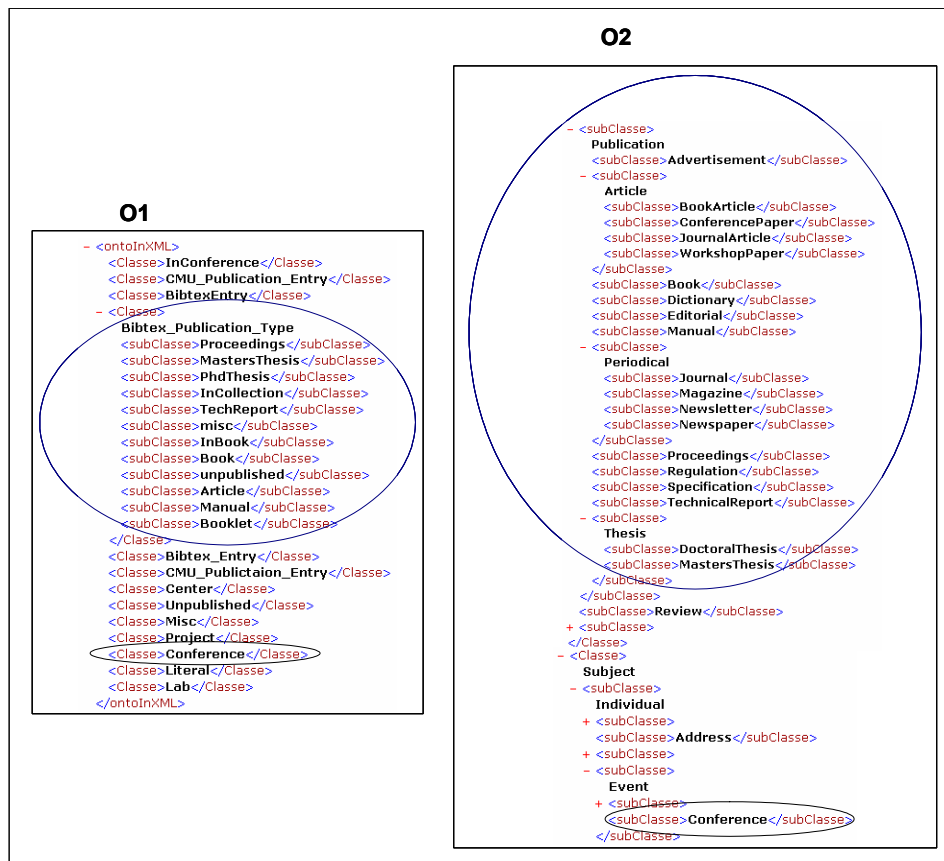


Figura 29 – Grupos de equivalência identificados no módulo com ordenação alfabética

Algumas classes idênticas lexicalmente como, por exemplo, as classes “MastersThesis” de ambas ontologias, pertencentes aos grupos de equivalência superiores na Figura 29, não são identificadas como *similares* porque possuem diferenças estruturais entre elas e próximas à elas, o que dificulta tanto a identificação de sua similaridade lexical quanto a formação de novos grupos de equivalência. Em **O1**, a classe “MastersThesis” é subclasse de “Bibtex_Publication_Type” e em **O2**, é subclasse de “Thesis” que, por sua vez, é subclasse de “Publication”. Como as classes “Bibtex_Publication_Type” de **O1** e “Publication” de **O2** foram identificadas como *similares*, então, existe a diferença de um nível hierárquico entre as classes “MastersThesis” em **O1** e em **O2** e, por isso, não há a identificação de similaridade destas classes.

As classes “Manual”, de ambas as ontologias, também não são identificadas como *similares* porque com a ordenação alfabética existem diferenças estruturais entre elas. Em **O1**, existe a classe “Booklet” logo abaixo da classe “Manual”, e em **O2**, existe a sub-árvore com a classe “Periodical” representando sua raiz e as subclasses “Journal”, “Magazine”, “Newsletter” e “Newspaper” representando suas folhas.

Pela igualdade lexical das classes “Conference”, em ambas as ontologias, e sua similaridade estrutural, indicada pela nova classe em **O1** e pela nova subclasse em **O2**, os próximos grupos de equivalência, representados pelos círculos inferiores na Figura 29, são identificados e suas classes são classificadas como *similares*.

O resultado final desta etapa de execução são as seguintes classes identificadas como *similares*: “Proceedings”, “Book” e “Conference”, de ambas as ontologias, e “Bibtex_Publication_Type” de **O1** *similar* à classe “Publication” de **O2**.

4.3.3.2.

Resultado da Terceira Etapa (com ordenação alfabética)

A terceira etapa de execução do CATO calcula os percentuais de similaridade das classes identificadas como *similares* na etapa anterior. A Figura 30 ilustra os resultados de tais cálculos.

```
Similarities Level:
Bibtex_Publication_Type -> Publication    *** Similarity Level: 15.384616%
Conference -> Conference    *** Similarity Level: 100.0%
Book -> Book    *** Similarity Level: 100.0%
Proceedings -> Proceedings    *** Similarity Level: 100.0%
```

Figura 30 – Percentuais de similaridade calculados

4.3.3.3.

Resultado do Alinhamento (com ordenação alfabética)

Ao concluir as três etapas de execução, o CATO identifica as classes equivalentes, i.e., aquelas que satisfazem todas as condições de cada uma de suas etapas, e realiza o alinhamento. Neste estudo de caso, com as árvores ordenadas alfabeticamente, as classes “Conference”, “Book” e “Proceedings” são as classes alinhadas pelo CATO. Isto porque são identificadas como *similares* pela sua segunda etapa de execução e classificadas como *bem similares*, com um

percentual de similaridade maior ou igual a 75%, pela sua terceira etapa de execução.

A Figura 31 ilustra parte da ontologia final do CATO com a classe “Book”, original da primeira ontologia, identificada pelo *namespace* “file:firstOnto.owl”, com a informação adicionada da classe equivalente identificada, a classe “Book”, original da segunda ontologia, identificada pelo *namespace* “file:secondOnto.owl”.

```

- <owl:Class rdf:about="file:firstOnto.owl#Book">
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Manual" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Proceedings" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#InConference" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Conference" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#MastersThesis" />
</owl:disjointWith>
<owl:disjointWith rdf:resource="file:firstOnto.owl#TechReport" />
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#InBook" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#PhdThesis" />
</owl:disjointWith>
- <owl:disjointWith>
  <owl:Class rdf:about="file:firstOnto.owl#Misc" />
</owl:disjointWith>
<rdfs:subClassOf rdf:resource="file:firstOnto.owl#Bibtex_Publication_Type" />
- <owl:equivalentClass>
  <owl:Class rdf:about="file:secondOnto.owl#Book" />
</owl:equivalentClass>
<owl:disjointWith rdf:resource="file:firstOnto.owl#Unpublished" />
<owl:disjointWith rdf:resource="file:firstOnto.owl#Booklet" />
</owl:Class>

```

← **Informação Adicionada**

Figura 31 – Informação adicionada, resultado do alinhamento com o CATO

4.3.4.

Avaliação dos Resultados

Das treze classes da primeira ontologia (**O1**) e das vinte e três classes da segunda ontologia (**O2**) referentes ao domínio de publicação, oito classes no total (“Conference”, “Manual”, “Book” e “Proceedings”, de ambas as ontologias) são alinhadas automaticamente pelo CATO, no módulo sem ordenação e seis classes no total (“Conference”, “Book” e “Proceedings”, de ambas as ontologias) no módulo com ordenação alfabética. Oito novas classes poderiam ser facilmente alinhadas se fosse permitida a intervenção do usuário (as classes “Article” e “MastersThesis”, de ambas ontologias e “PhdThesis” de **O1** com “DoctoralThesis” de **O2** e “TechReport” de **O1** com “TechnicalReport” de **O2**).

Visto que as ontologias comparadas foram desenvolvidas sem um propósito comum entre elas e por grupos totalmente independentes, o resultado encontrado pelo CATO é satisfatório.

Neste estudo de caso, os resultados finais dos módulos sem e com ordenação alfabética, para o processo de alinhamento, foram diferentes. As seis classes equivalentes, “Conference”, “Book” e ”Proceedings”, de ambas as ontologias, são alinhadas nos dois módulos. Porém, as classes equivalentes “Manual”, de ambas as ontologias, não são alinhadas no módulo com ordenação alfabética porque possuem diferenças estruturais próximas a elas neste módulo. Além disso, os percentuais de similaridade entre as classes raízes do primeiro grupo de equivalência identificado (“Bibtex_Publication_Type” de **O1** e “Publication” de **O2**) são diferentes em ambos os módulos, comprovando que seu cálculo também é influenciado pelos percentuais de similaridade de suas subclasses.

4.3.5. Problemas Encontrados

A qualidade do alinhamento poderia ser sensivelmente melhorada se os sinônimos das classes já estivessem cadastrados no banco de dados ou se houvesse o uso de um banco de sinônimos melhor. No entanto, como foi visto, se existem diferenças nos nomes das classes pais (um nível hierárquico acima) ou avôs (dois níveis hierárquicos acima) de uma classe e seu sinônimo, a informação identificada do sinônimo não é utilizada no processo de alinhamento. Este fato é um problema que merece melhor análise para minimizar a falta de completude do alinhamento realizado pelo CATO.

Neste exemplo de estudo de caso, o uso de sinônimos possibilitaria o CATO alinhar automaticamente as oito classes facilmente alinhadas por um usuário (as classes “Article” e “MastersThesis”, de ambas ontologias, e a classe “PhdThesis” de **O1** com a classe “DoctoralThesis” de **O2** e a classe “TechReport” de **O1** com a classe “TechnicalReport” de **O2**) se suas classes pais e avôs fossem descritas com os mesmos nomes ou se existisse outra condição de poda a ser satisfeita (por exemplo, além da igualdade dos nomes destas classes também a igualdade dos nomes com seus sinônimos para as classes pais e avôs, como é feito para as classes que dispararam a análise da condição de poda).

Em relação à execução do CATO, nenhum problema foi encontrado com este estudo de caso escolhido. As ontologias de entrada foram alinhadas automaticamente, sem erro algum.

4.4.

Segundo Estudo de Caso

Este estudo de caso compara parte de uma ontologia genérica com uma ontologia específica onde, além do nível de abstração das classes, também as estruturas hierárquicas das ontologias comparadas são bem diferentes. A Figura 32 ilustra estas diferentes hierarquias. Os endereços das ontologias escritas em *OWL* estão disponíveis no Anexo B deste documento.

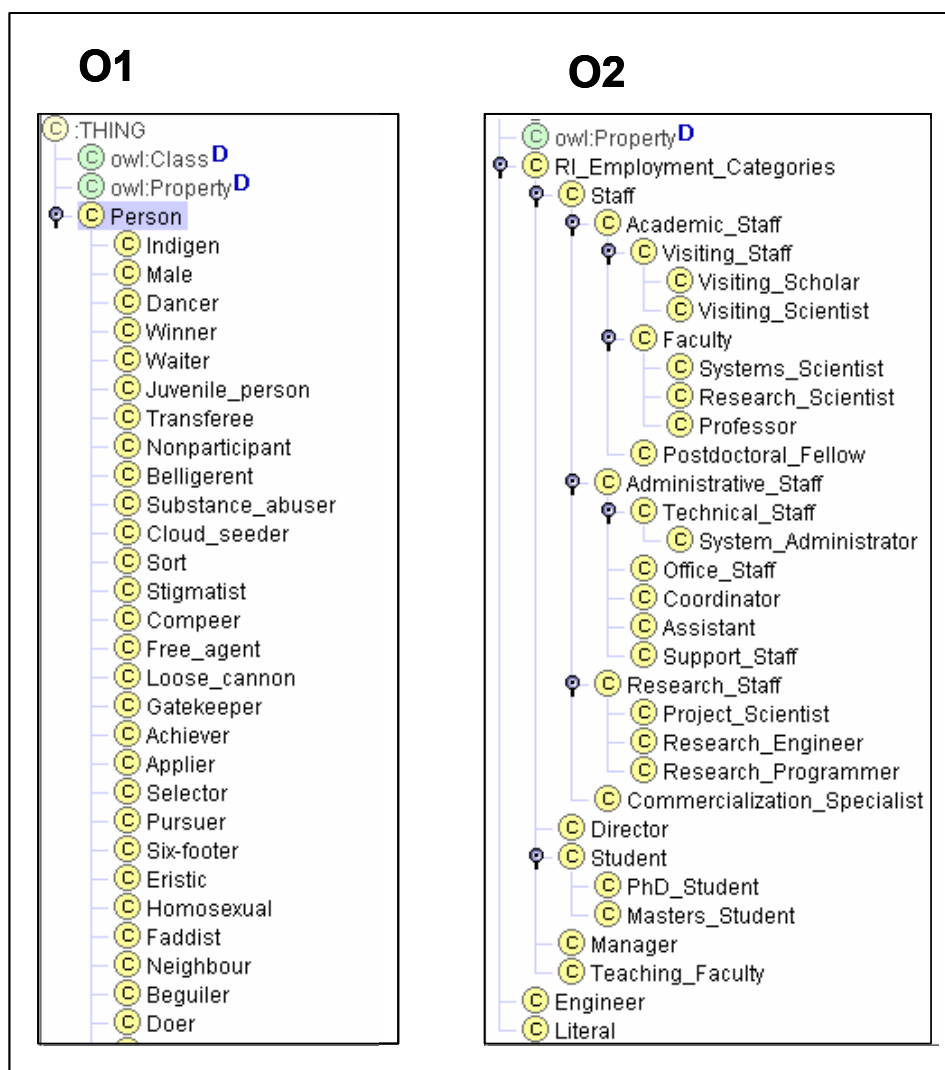


Figura 32 – Ontologias comparadas

A primeira ontologia escolhida é formada pelas classes do tipo pessoa cadastradas no thesaurus WordNet (Fellbaum, 2004). Esta ontologia é

disponibilizada no diretório de pesquisa do SchemaWeb em (SchemaWeb, 2004b).

A segunda ontologia escolhida, a *CMU RI Employment Categories* (CMU, 2004e), é sobre as categorias de empregos do Instituto de Robótica da Universidade Carnegie Mellon, também do grupo de pesquisa ATLAS (ATLAS, 2004), de onde a primeira ontologia do primeiro estudo de caso foi escolhida.

Esta experiência foi interessante porque permitiu a avaliação do CATO frente ao alinhamento de ontologias formadas por um grande número de termos. Como a primeira ontologia (**O1**) é formada pelas classes do WordNet e este trata-se de um thesaurus, é esperado que existam várias classes descritas e estas sejam mais genéricas que as criadas na segunda ontologia (**O2**), que é constituída por classes específicas do domínio de tipos de empregos.

4.4.1.

Resultado da Primeira Etapa

Inicialmente, como não existem cadastrados os sinônimos das classes das ontologias comparadas e as classes iguais lexicalmente possuem diferenças estruturais, portanto, não satisfazem a condição de poda da primeira etapa, a conclusão da primeira etapa de execução do CATO não resulta em alinhamento algum.

Para tentar algum alinhamento, os sinônimos ilustrados entre colchetes na Figura 33 foram cadastrados no banco de dados utilizado e a primeira etapa do CATO foi executada novamente. Nesta nova execução, as classes “Technologist” (subclasse de “Person”) de **O1** e “Research_Programmer” (subclasse de “Research_Staff”) de **O2** foram identificadas como sinônimos e, por isso, são comparadas. Como estas classes não satisfazem a condição de poda porque possuem classes pais diferentes, um nível hierárquico acima, não são alinhadas. O mesmo acontece com as classes “Scholar” e “Learner” (subclasses de “Person”) de **O1** identificadas como sinônimos da classe “Student” (subclasse de “RI_Employment_Categories”) de **O2**. Assim, a conclusão da primeira etapa de execução do CATO com o uso de sinônimos também não resulta em alinhamento algum.


```
Sinonimos de file:firstOnto.owl#Technologist: [Research_Programmer]
Sinonimos de file:firstOnto.owl#Learner: [Student]
Sinonimos de file:firstOnto.owl#Scholar: [Student]
Sinonimos de file:secondOnto.owl#Student: [Learner, Scholar]
Sinonimos de file:secondOnto.owl#Research_Programmer: [Technologist]
```

Figura 33 – Sinônimos cadastrados identificados

Como as informações dos sinônimos identificadas não satisfazem as condições de alinhamento da primeira etapa de execução do CATO, então, não são propagadas para as demais etapas de execução e, conseqüentemente, o uso dos sinônimos não resulta impacto algum no alinhamento final das ontologias, neste estudo de caso.

4.4.2. Resultados das Etapas sem Ordenação Alfabética

Os sub-tópicos a seguir mostram os resultados das etapas executadas no CATO com as hierarquias das árvores comparadas sem ordenação alfabética, onde as classes e subclasses das ontologias são estruturadas com sua ordem original de criação.

4.4.2.1. Resultado da Segunda Etapa (sem ordenação)

A Figura 34 ilustra as hierarquias das duas ontologias comparadas e seus grupos de equivalência identificados.

O1

```
<subClasse>Dancer</subClasse>
<subClasse>Pisser</subClasse>
<subClasse>Hugger</subClasse>
<subClasse>Mutilator</subClasse>
<subClasse>Manipulator</subClasse>
<subClasse>Possessor</subClasse>
<subClasse>Visionary</subClasse>
<subClasse>Acquaintance</subClasse>
<subClasse>Complexifier</subClasse>
<subClasse>Battler</subClasse>
<subClasse>Allayer</subClasse>
<subClasse>Snuffer</subClasse>
<subClasse>Dupe</subClasse>
<subClasse>Pisces</subClasse>
<subClasse>Dissident</subClasse>
<subClasse>Protester</subClasse>
<subClasse>Engineer</subClasse>
<subClasse>Contriver</subClasse>
<subClasse>Left-hander</subClasse>
<subClasse>Insured</subClasse>
<subClasse>Miracle_worker</subClasse>
<subClasse>Emotional_person</subClasse>
<subClasse>Transsexual</subClasse>
<subClasse>Trier</subClasse>
<subClasse>Disputant</subClasse>
<subClasse>Toreador</subClasse>
<subClasse>Female_person</subClasse>
<subClasse>Groaner</subClasse>
<subClasse>Indweller</subClasse>
<subClasse>Applied_scientist</subClasse>
<subClasse>Lefty</subClasse>
<subClasse>Machine</subClasse>
<subClasse>Inhabitant</subClasse>
<subClasse>Straight_person</subClasse>
```

O2

```
- <ontoInXML>
<Classe>Teaching_Faculty</Classe>
- <Classe>
  RI_Employment_Categories
  <subClasse>Manager</subClasse>
  <subClasse>Director</subClasse>
- <subClasse>
  Staff
  - <subClasse>
    Research_Staff
    <subClasse>Project_Scientist</subClasse>
    <subClasse>Research_Engineer</subClasse>
    <subClasse>Research_Programmer</subClasse>
  </subClasse>
- <subClasse>
  Administrative_Staff
  <subClasse>Coordinator</subClasse>
  <subClasse>Office_Staff</subClasse>
  <subClasse>Support_Staff</subClasse>
  + <subClasse>
    <subClasse>Assistant</subClasse>
  </subClasse>
  <subClasse>Commercialization_Specialist</subClasse>
  + <subClasse>
    </subClasse>
  </subClasse>
- <subClasse>
  Student
  <subClasse>PhD_Student</subClasse>
  <subClasse>Masters_Student</subClasse>
  </subClasse>
</Classe>
<Classe>Literal</Classe>
<Classe>Engineer</Classe>
</ontoInXML>
```

Figura 34 – Grupos de equivalência identificados no módulo sem ordenação

Em **O1**, só existe uma classe raiz, a classe “Person”, e todas as demais classes da ontologia são subclasses desta classe raiz. Em **O2**, existem classes tanto sem subclasses como, por exemplo a classe “Engineer”, quanto com subclasses como, por exemplo, a classe “RI_Employment_Categories”. Como em **O2** não existe uma classe raiz única cadastrada, como a classe “Person” de **O1**, então, o CATO vai formando os grupos de equivalência sempre com uma das subclasses de **O1** com as classes sem subclasses de **O2**. Assim, as únicas classes equivalentes, as classes “Engineer”, de ambas as ontologias, são identificadas como *similares* pelos grupos de equivalências formados por suas classes.

4.4.2.2.

Resultado da Terceira Etapa (sem ordenação)

A terceira etapa de execução do CATO calcula os percentuais de similaridade das classes identificadas como *similares* na etapa anterior. A Figura 35 ilustra os resultados de tais cálculos.

```
Similarities Level:  
Person -> Person    *** Similarity Level: 0.17361112*  
Engineer -> Engineer *** Similarity Level: 100.0%
```

Figura 35 – Percentuais de similaridade calculados

4.4.2.3.

Resultado do Alinhamento (sem ordenação)

Ao concluir as três etapas de execução, o CATO identifica as classes equivalentes, i.e., aquelas que satisfazem todas as condições de cada uma de suas etapas, e realiza o alinhamento. Neste estudo de caso, com as árvores estruturadas na sua ordem original de criação, as classes “Engineer”, de ambas as ontologias, são as classes alinhadas pelo CATO. Isto porque, são identificados como *similares* pela sua segunda etapa de execução e classificados como *bem similares*, com um percentual de similaridade maior ou igual a 75%, pela sua terceira etapa de execução.

A Figura 36 ilustra parte da ontologia final do CATO com a classe “Engineer”, original da segunda ontologia, identificada pelo *namespace* “file:secondOnto.owl”, com a informação adicionada da classe equivalente identificada, a classe “Engineer”, original da primeira ontologia, identificada pelo *namespace* “file:firstOnto.owl”.

```

<rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
<j.0:comment>one who outlives another; "he left his farm to his survivors"</j.0:comment>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Libra" j.0:label="Libra">
  <j.0:description>(astrology) a person who is born while the sun in in Libra</j.0:description>
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>(astrology) a person who is born while the sun in in Libra</j.0:comment>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Freewheeler" j.0:label="Freewheeler">
  <j.0:description>someone acting freely or even irresponsibly</j.0:description>
  <j.0:comment>someone acting freely or even irresponsibly</j.0:comment>
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Controller" j.0:label="Controller">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>a person who directs and restrains</j.0:comment>
  <j.0:description>a person who directs and restrains</j.0:description>
</owl:Class>
- <owl:Class rdf:about="file:secondOnto.owl#Engineer">
- <owl:equivalentClass>
  <owl:Class rdf:about="file:firstOnto.owl#Engineer" />
</owl:equivalentClass>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Divider" j.0:label="Divider">
  <j.0:description>a person who separates something into parts or groups</j.0:description>
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>a person who separates something into parts or groups</j.0:comment>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Applied_scientist" j.0:label="Applied_scientist">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>a person who uses scientific knowledge to solve practical problems</j.0:comment>
  <j.0:description>a person who uses scientific knowledge to solve practical problems</j.0:description>
</owl:Class>

```

← **Informação Adicionada**

Figura 36 – Informação adicionada, resultado do alinhamento com o CATO

4.4.3.

Resultados das Etapas com Ordenação Alfabética

Os sub-tópicos a seguir mostram os resultados das etapas executadas no CATO com as hierarquias das árvores comparadas estruturadas com ordenação alfabética, i.e., as classes e subclasses das ontologias são estruturadas ordenadas alfabeticamente.

4.4.3.1.

Resultado da Segunda Etapa (com ordenação alfabética)

A Figura 37 ilustra as hierarquias das duas ontologias comparadas e seus grupos de equivalência identificados.

Em **O1**, só existe uma classe raiz, a classe “Person”, e todas as demais classes da ontologia são subclasses desta classe raiz. Em **O2**, existem classes tanto sem subclasses como, por exemplo, a classe “Engineer”, quanto com subclasses como, por exemplo, a classe “RI_Employment_Categories”. Como em **O2** não existe uma classe raiz única cadastrada, como a classe “Person” de **O1**, então, o CATO vai formando os grupos de equivalência sempre com uma das subclasses de **O1** com as classes sem subclasses de **O2**. Assim, as únicas classes equivalentes, as classes “Engineer”, de ambas as ontologias, são identificadas como *similares* pelos grupos de equivalências formados por suas classes.

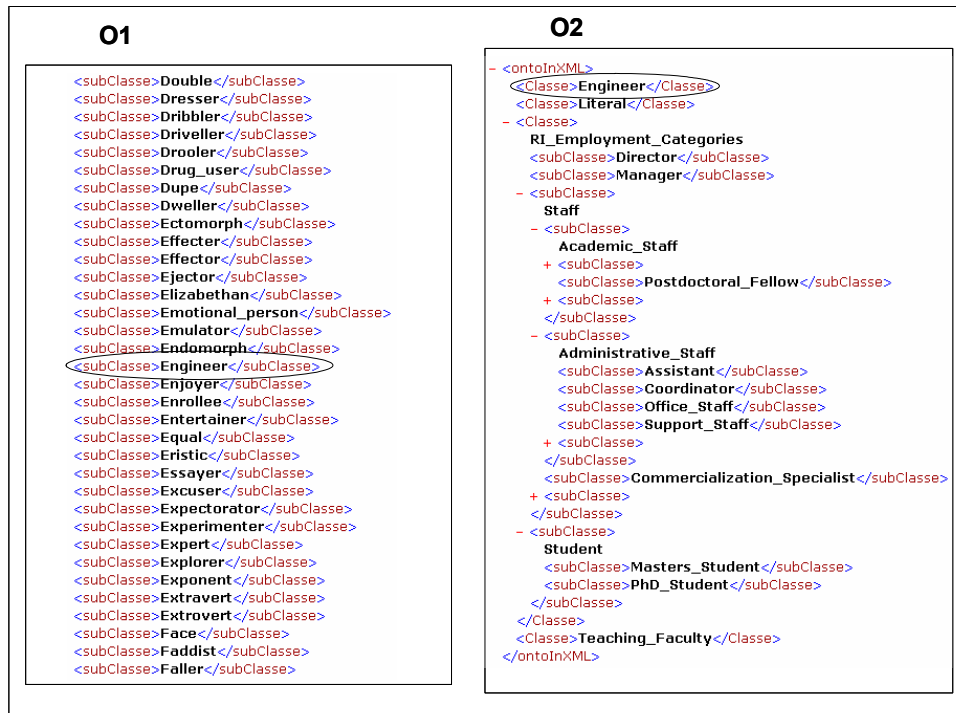


Figura 37 – Grupos de equivalência identificados no módulo com ordenação alfabética

4.4.3.2.

Resultado da Terceira Etapa (com ordenação alfabética)

A terceira etapa de execução do CATO calcula os percentuais de similaridade das classes identificadas como *similares* na etapa anterior. A Figura 38 ilustra os resultados de tais cálculos.

```

Similarities Level:
Person -> Person    *** Similarity Level: 0.17361112%
Engineer -> Engineer *** Similarity Level: 100.0%

```

Figura 38 – Percentuais de similaridade calculados

4.4.3.3.

Resultado do Alinhamento (com ordenação alfabética)

Ao concluir as três etapas de execução, o CATO identifica as classes equivalentes, i.e., aquelas que satisfazem todas as condições de cada uma de suas etapas, e realiza o alinhamento. Neste estudo de caso, com as árvores estruturadas ordenadas alfabeticamente, as classes “Engineer”, de ambas as ontologias, são as classes alinhadas pelo CATO. Isto porque são identificadas como *similares* pela sua segunda etapa de execução e classificadas como *bem similares*, com um percentual de similaridade maior ou igual a 75%, pela sua terceira etapa de execução.

A Figura 39 ilustra parte da ontologia final do CATO com a classe “Engineer”, original da segunda ontologia, identificada pelo *namespace* “file:secondOnto.owl”, com a informação adicionada da classe equivalente identificada, a classe “Engineer”, original da primeira ontologia, identificada pelo *namespace* “file:firstOnto.owl”.

```

<rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
<j.0:comment>one who outlives another; "he left his farm to his survivors"</j.0:comment>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Libra" j.0:label="Libra">
  <j.0:description>(astrology) a person who is born while the sun in in Libra</j.0:description>
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>(astrology) a person who is born while the sun in in Libra</j.0:comment>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Freewheeler" j.0:label="Freewheeler">
  <j.0:description>someone acting freely or even irresponsibly</j.0:description>
  <j.0:comment>someone acting freely or even irresponsibly</j.0:comment>
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Controller" j.0:label="Controller">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>a person who directs and restrains</j.0:comment>
  <j.0:description>a person who directs and restrains</j.0:description>
</owl:Class>
- <owl:Class rdf:about="file:secondOnto.owl#Engineer">
  - <owl:equivalentClass>
    <owl:Class rdf:about="file:firstOnto.owl#Engineer" />
  </owl:equivalentClass>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Divider" j.0:label="Divider">
  <j.0:description>a person who separates something into parts or groups</j.0:description>
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>a person who separates something into parts or groups</j.0:comment>
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Applied_scientist" j.0:label="Applied_scientist">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Person" />
  <j.0:comment>a person who uses scientific knowledge to solve practical problems</j.0:comment>
  <j.0:description>a person who uses scientific knowledge to solve practical problems</j.0:description>
</owl:Class>

```

← **Informação Adicionada**

Figura 39 – Informação adicionada, resultado do alinhamento com o CATO

4.4.4.

Avaliação dos Resultados

Neste exemplo de estudo de caso, das mais de quinhentas e setenta classes da primeira ontologia (**O1**) e das aproximadamente trinta classes da segunda ontologia (**O2**), as classes equivalentes “Engineer”, de ambas as ontologias, foram alinhadas pelo CATO, nos seus módulos sem e com ordenação alfabética.

Visto que as ontologias comparadas foram desenvolvidas sem um propósito comum entre elas e por grupos totalmente independentes, além de serem de tipos diferentes (**O1** é parte de uma ontologia genérica e **O2** é uma ontologia específica), o resultado encontrado pelo CATO é satisfatório. Isto porque o CATO traz a automação necessária para o alinhamento destas ontologias, pois, o alinhamento manual seria bastante trabalhoso, devido à grande quantidade de classes existentes a serem comparadas. Vale lembrar que as mais de quinhentas e setenta classes de **O1**, originalmente, não estão organizadas alfabeticamente como

ilustradas na Figura 37, de modo a facilitar a comparação. Esta organização foi realizada pelo CATO.

Neste estudo de caso, os resultados finais dos módulos sem e com ordenação alfabética, para o processo de alinhamento, não foram diferentes. As duas classes equivalentes "Engineer", de ambas as ontologias, são alinhadas nestes dois módulos. Além disso, a classe raiz "Person" de **O1** teve os mesmos valores de percentuais de similaridade calculados nos dois módulos, confirmando que a ordenação das classes neste estudo de caso não teve influência no alinhamento final do CATO.

4.4.5.

Problemas Encontrados

Para ontologias modeladas com estruturas diferentes como, por exemplo, uma ontologia com poucas classes gerais e muitas classes específicas, comparada com uma outra ontologia, com muitas classes gerais e poucas classes específicas, a identificação dos grupos de equivalência é difícil porque existe pouca similaridade estrutural entre as classes. Conseqüentemente, o número de classes alinhadas é bastante reduzido. Este número resume-se às classes alinhadas na primeira etapa de execução do CATO ou, talvez, às identificadas como iguais lexicalmente nos grupos de equivalência da segunda etapa.

Em relação à execução do CATO, nenhum problema foi encontrado com este estudo de caso escolhido. As ontologias de entrada foram alinhadas automaticamente, sem erro algum. No entanto, houve uma maior demora em comparação ao tempo de execução dos demais estudos de caso devido ao maior número de classes a serem comparadas neste estudo de caso.

4.5.

Terceiro Estudo de Caso

Este estudo de caso trata de ontologias criadas com as classes referentes à classe "Meio de Transporte". Para a comparação, as partes específicas ao assunto foram extraídas das ontologias genéricas SUMO (SUMO, 2004) e OpenCYC (OpenCYC, 2004). Os endereços das ontologias resumidas escritas em *OWL* estão disponíveis no Anexo C deste documento.

Esta experiência de alinhamento foi interessante porque, como as ontologias comparadas são genéricas, suas classes estão organizadas segundo uma hierarquia

de classes bem gerais, algumas delas ilustradas na Figura 40. Por exemplo, nas duas ontologias existem as classes “Ambulance” cadastradas. No entanto, na primeira ontologia (O1), esta classe é subclasse das seguintes classes: “EmergencyRoadVehicle”, que é subclasse de “RoadVehicle”, que é subclasse de “LandVehicle”, que, finalmente, é subclasse da raiz “Vehicle”. Na segunda ontologia (O2), a classe “Ambulance” é subclasse das seguintes classes: “EmergencyVehicle”, que é subclasse de “Vehicle”, que é subclasse de “TransportationDevice”, que é subclasse de “SelfPoweredDevice”, que, por fim, é subclasse de “MechanicalDevice”.

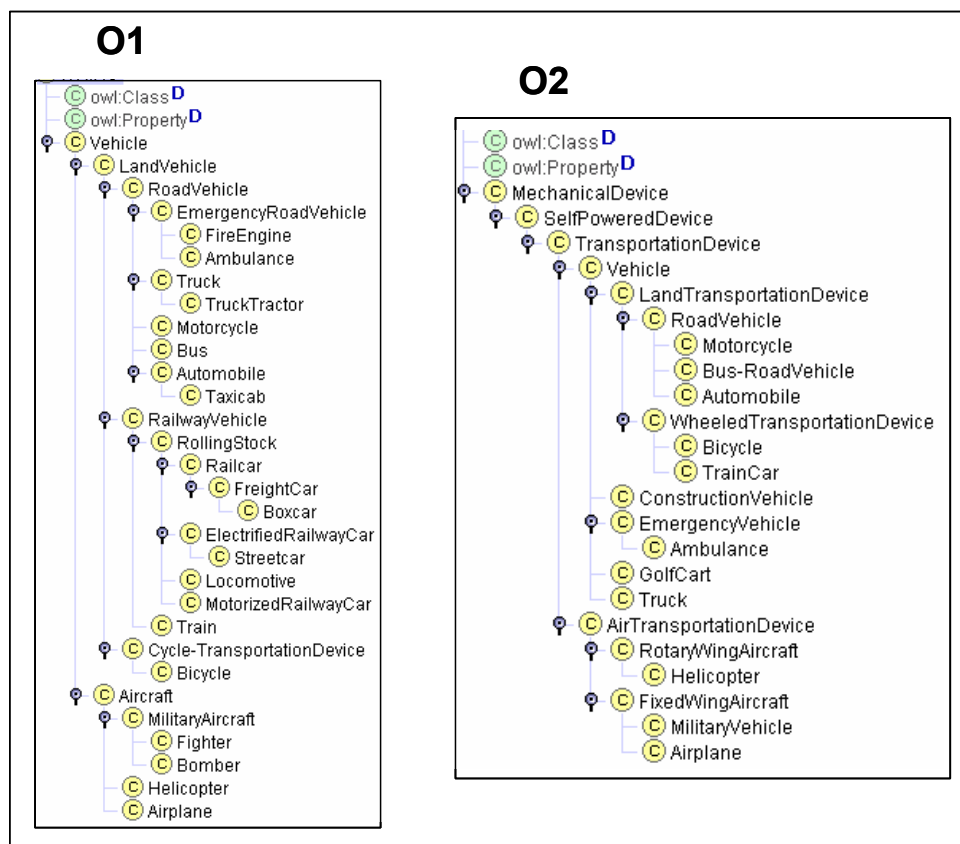


Figura 40 – Ontologias comparadas

O alinhamento realizado pelo CATO neste estudo de caso é bastante limitado porque, apesar da igualdade lexical de algumas de suas classes, estas não possuem similaridades estruturais. Assim, além das condições de poda da primeira etapa de execução do CATO não serem satisfeitas, resultando no não alinhamento das classes iguais lexicalmente, a identificação dos grupos de equivalência também é dificultada, o que faz com que a comparação estrutural, segunda etapa de execução, e o cálculo de medidas de similaridades, terceira etapa de execução, resultem em poucas classes alinhadas.

4.5.1. Resultado da Primeira Etapa

Inicialmente, como não existem cadastrados os sinônimos das classes das ontologias comparadas e as classes iguais lexicalmente possuem diferenças estruturais, portanto, não satisfazem a condição de poda da primeira etapa, a conclusão da primeira etapa de execução do CATO não resulta em alinhamento algum.

Para tentar algum alinhamento, os sinônimos ilustrados entre colchetes na Figura 41 foram cadastrados no banco de dados utilizado e a primeira etapa do CATO foi executada novamente. Nesta nova execução, as classes “MilitaryAircraft” de **O1** e “MilitaryVehicle” de **O2** foram identificadas como sendo sinônimos e, por isso, são comparadas. Como estas classes não satisfazem a condição de poda porque possuem classes pais diferentes, um nível hierárquico acima, não são alinhadas. O mesmo acontece com as classes “LandVehicle” de **O1** e “LandTransportationDevide” de **O2**, “Aircraft” de **O1** e “FixedWingAircraft” de **O2** e, por fim, as classes comparadas “Bus” de **O1** e “Bus-RoadVehicle” de **O2**. Assim, a conclusão da primeira etapa de execução do CATO com o uso de sinônimos também não resulta em alinhamento algum.

```
Sinonimos de file:firstOnto.owl#MilitaryAircraft: [MilitaryVehicle]
Sinonimos de file:firstOnto.owl#LandVehicle: [LandTransportationDevide]
Sinonimos de file:firstOnto.owl#Aircraft: [FixedWingAircraft]
Sinonimos de file:firstOnto.owl#Bus: [Bus-RoadVehicle]
Sinonimos de file:secondOnto.owl#LandVehicle: [LandTransportationDevide]
Sinonimos de file:secondOnto.owl#MilitaryVehicle: [MilitaryAircraft]
Sinonimos de file:secondOnto.owl#FixedWingAircraft: [Aircraft]
Sinonimos de file:secondOnto.owl#Bus-RoadVehicle: [Bus]
```

Figura 41 – Sinônimos cadastrados identificados

Neste estudo de caso, para exemplificar o uso dos sinônimos identificados, a classe “LandTransportationDevide” de **O2** foi renomeada para “LandVehicle” (facilmente identificada por um usuário como equivalente à classe “LandVehicle” de **O1**) e a primeira etapa do CATO executada mais uma vez. Agora, as condições de poda das seguintes classes são satisfeitas e, conseqüentemente, seus alinhamentos são realizados: a classe “Bus” de **O1** é alinhada com a classe “Bus-RoadVehicle” de **O2**, e as classes “RoadVehicle”, “Motorcycle” e “Automobile”, de ambas ontologias, são alinhadas.


```

URI da Classe:file:firstOnto.owl#RoadVehicle
URI da Classe Pai: file:firstOnto.owl#LandVehicle
URI da Classe Avo: file:firstOnto.owl#Vehicle
URI da Classe Equivalente: file:secondOnto.owl#RoadVehicle

- <owl:Class rdf:about="file:firstOnto.owl#RoadVehicle">
- <rdfs:subClassOf>
  <owl:Class rdf:about="file:firstOnto.owl#LandVehicle" />
</rdfs:subClassOf>
<owl:equivalentClass rdf:resource="file:secondOnto.owl#RoadVehicle" />
</owl:Class>

```

Figura 42 – Informação adicionada, resultado do alinhamento com o uso de sinônimos na primeira etapa do CATO

A Figura 42 ilustra as informações de equivalência identificadas pelo CATO da classe “RoadVehicle” de **O1** com a classe “RoadVehicle” de **O2**, tanto na resposta das informações existentes nesta classe após o término de execução da primeira etapa quanto em seu código resultado em *OWL*.

Com a renomeação realizada, as informações dos sinônimos identificados satisfazem as condições de alinhamento da primeira etapa de execução do CATO e, por isso, estas seriam propagadas para as demais etapas de execução. Conseqüentemente, o uso dos sinônimos impactaria no alinhamento final das ontologias.

No entanto, a apresentação deste estudo de caso continuará com as ontologias originais escolhidas sem qualquer modificação. Assim, a renomeação da classe “LandTransportationDevice” de **O2**, utilizada apenas para exemplificar o uso dos sinônimos, é desfeita e os alinhamentos resultantes desta renomeação (a classe “Bus” de **O1** alinhada com a classe “Bus-RoadVehicle” de **O2**, e as classes alinhadas “RoadVehicle”, “Motorcycle” e “Automobile”, de ambas ontologias) serão desconsiderados porque não mais satisfazem a condição de poda.

4.5.2.

Resultados das Etapas sem Ordenação Alfabética

Os sub-tópicos a seguir mostram os resultados das etapas executadas no CATO com as hierarquias das árvores comparadas sem ordenação alfabética, onde as classes e subclasses das ontologias são estruturadas com sua ordem original de criação.

4.5.2.1.

Resultado da Segunda Etapa (sem ordenação)

A Figura 43 ilustra as estruturas hierárquicas das ontologias comparadas. Devido ao fato das poucas classes iguais lexicalmente possuírem diferenças estruturais entre elas, a execução do CATO, neste módulo, não identificou

qualquer grupo de equivalência e, conseqüentemente, nenhuma classe foi identificada como *similar* nesta etapa de execução.

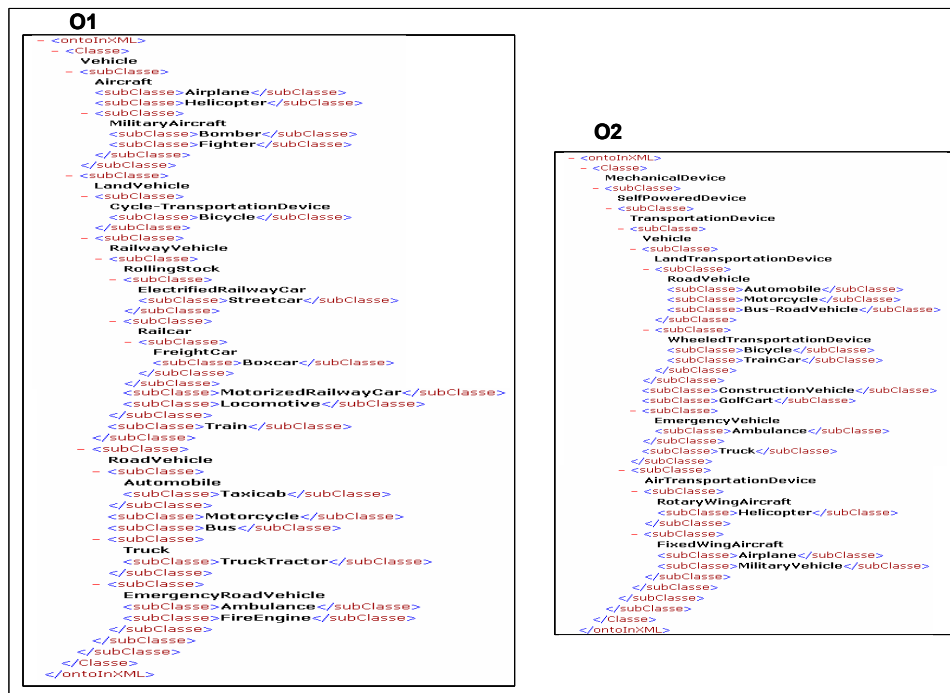


Figura 43 – Estruturas hierárquicas das ontologias comparadas

4.5.2.2.

Resultado da Terceira Etapa (sem ordenação)

A terceira etapa de execução do CATO calcula os percentuais de similaridade das classes identificadas como *similares* na etapa anterior.

Como nenhuma classe foi identificada como *similar* na etapa anterior deste módulo, então, nenhum cálculo de percentual de similaridade é realizado.

4.5.2.3.

Resultado do Alinhamento (sem ordenação)

Ao concluir as três etapas de execução, o CATO identifica as classes equivalentes, i.e., aquelas que satisfazem todas as condições de cada uma de suas etapas, e realiza o alinhamento. Neste estudo de caso, com as árvores estruturadas na sua ordem original de criação, nenhuma classe foi alinhada. Tal resultado é aceitável porque as ontologias comparadas são extraídas de ontologias genéricas e, por isso, possuem além de classes bem abstratas, identificadas por diferentes nomes, também modelagens conceituais bastante distintas.

4.5.3.

Resultados das Etapas com Ordenação Alfabética

Os sub-tópicos a seguir mostram os resultados das etapas executadas no CATO com as hierarquias das árvores comparadas estruturadas com ordenação alfabética, i.e., as classes e subclasses das ontologias são estruturadas ordenadas alfabeticamente.

4.5.3.1.

Resultado da Segunda Etapa (com ordenação alfabética)

A Figura 44 ilustra as hierarquias das duas ontologias comparadas e seus grupos de equivalência identificados.

Neste exemplo, como as duas ontologias comparadas possuem uma relação fraca entre elas, então, os muitos grupos de equivalência identificados resumem-se, praticamente, aos formados pelas classes equivalentes com alguma similaridade entre elas. Essa identificação é influenciada tanto pela igualdade lexical quanto pela organização estrutural dessas classes equivalentes. Por exemplo, as classes “RoadVehicle”, de ambas ontologias, são identificadas como *similares* e seus grupos de equivalência definidos porque são iguais lexicalmente e cada uma delas possui exatamente três subclasses.

Já as classes “LandVehicle” de **O1** e “Vehicle” de **O2** são identificadas como *similares* porque possuem uma grande quantidade de subclasses organizadas com similaridades estruturais. Com estas identificações e similaridades estruturais das classes superiores a “LandVehicle” de **O1** e “Vehicle” de **O2**, outros grupos de equivalência são fechados. Nestes grupos, as classes raízes "Aircraft" de **O1** e "AirTransportationDevice" de **O2** são identificadas como *similares*.

As seguintes outras classes são identificadas como *similares*, devido à igualdade lexical e similaridade estrutural: “Vehicle”, “Airplane” e “Truck”, de ambas as ontologias comparadas.

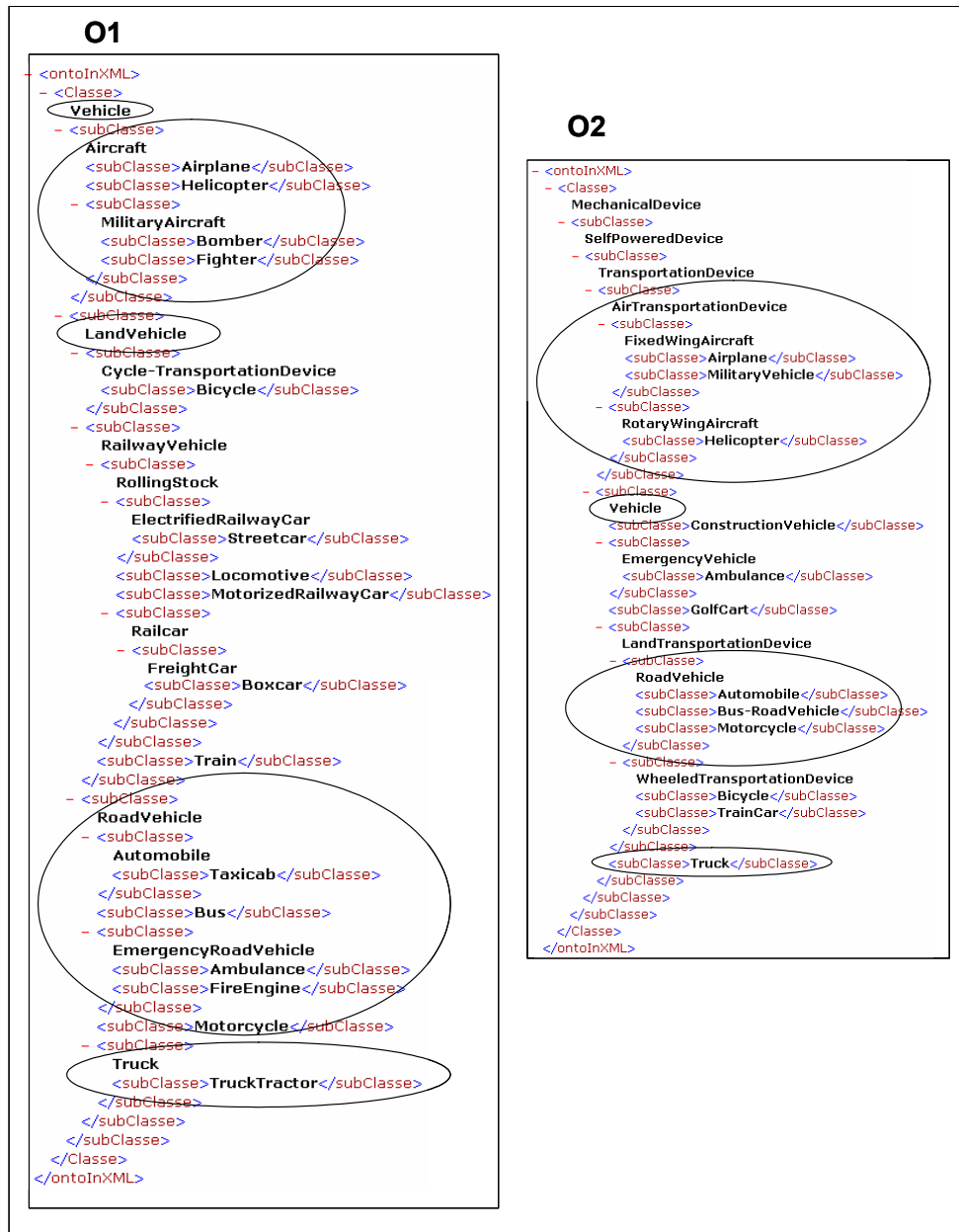


Figura 44 – Grupos de equivalência identificados no módulo com ordenação alfabética

4.5.3.2.

Resultado da Terceira Etapa (com ordenação alfabética)

A terceira etapa de execução do CATO calcula os percentuais de similaridade das classes identificadas como *similares* na etapa anterior. A Figura 45 ilustra os resultados de tais cálculos.

```

Similarities Level:
Vehicle -> Vehicle   *** Similarity Level: 9.027779%
Aircraft -> AirTransportationDevice *** Similarity Level: 25.0%
Airplane -> Airplane *** Similarity Level: 100.0%
LandVehicle -> Vehicle *** Similarity Level: 2.0833335%
RoadVehicle -> RoadVehicle *** Similarity Level: 8.333334%
Truck -> Truck     *** Similarity Level: 50.0%
    
```

Figura 45 – Percentuais de similaridade calculados

4.5.3.3.

Resultado do Alinhamento (com ordenação alfabética)

Ao concluir as três etapas de execução, o CATO identifica as classes equivalentes, i.e., aquelas que satisfazem todas as condições de cada uma de suas etapas, e realiza o alinhamento. Neste estudo de caso, com as árvores estruturadas ordenadas alfabeticamente, as classes “Airplane”, de ambas as ontologias comparadas, são as classes alinhadas pelo CATO. Isto porque são identificadas como *similares* pela sua segunda etapa de execução e classificadas como *bem similares*, com um percentual de similaridade maior ou igual a 75%, pela sua terceira etapa de execução.

A Figura 46 ilustra parte da ontologia final do CATO com a classe “Airplane”, original da primeira ontologia identificada pelo *namespace* “file:firstOnto.owl”, com a informação adicionada da classe equivalente identificada, a classe “Airplane”, original da segunda ontologia identificada pelo *namespace* “file:secondOnto.owl”.

```

- <owl:Class rdf:about="file:firstOnto.owl#Locomotive">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#RollingStock" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Cycle-TransportationDevice">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#LandVehicle" />
</owl:Class>
- <owl:Class rdf:about="file:secondOnto.owl#RoadVehicle">
  <rdfs:subClassOf rdf:resource="file:secondOnto.owl#LandTransportationDevice" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Bicycle">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Cycle-TransportationDevice" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Airplane">
  <owl:equivalentClass rdf:resource="file:secondOnto.owl#Airplane" />
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#Aircraft" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#FireEngine">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#EmergencyRoadVehicle" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#Bomber">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#MilitaryAircraft" />
</owl:Class>
- <owl:Class rdf:about="file:secondOnto.owl#Vehicle">
  <rdfs:subClassOf rdf:resource="file:secondOnto.owl#TransportationDevice" />
</owl:Class>
- <owl:Class rdf:about="file:secondOnto.owl#FixedWingAircraft">
  <rdfs:subClassOf rdf:resource="file:secondOnto.owl#AirTransportationDevice" />
</owl:Class>
- <owl:Class rdf:about="file:secondOnto.owl#TrainCar">
  <rdfs:subClassOf rdf:resource="file:secondOnto.owl#WheeledTransportationDevice" />
</owl:Class>
- <owl:Class rdf:about="file:firstOnto.owl#RailwayVehicle">
  <rdfs:subClassOf rdf:resource="file:firstOnto.owl#LandVehicle" />
</owl:Class>

```

← **Informação Adicionada**

Figura 46 – Informação adicionada, resultado do alinhamento com o CATO

4.5.4.

Avaliação dos Resultados

Apesar de ser utilizado para um propósito diferente do que foi implementado, i.e., comparação de ontologias genéricas em vez de comparação de ontologias específicas, o CATO trouxe resultado com seu alinhamento neste

exemplo de estudo de caso. Claro que este resultado não é tão bom como o conseguido na comparação de ontologias complementares, onde estas ontologias são de domínios específicos com classes relacionadas entre elas, mas por outro lado, é bem melhor do que o resultado conseguido com o alinhamento semi-automático ou manual, levando em consideração o tempo de resposta destes. Isto porque, as ontologias comparadas são genéricas e possuem um grande número de classes a serem comparadas sem uma organização hierárquica intuitiva. Assim, a comparação semi-automática ou manual demoraria um tempo razoável para ser concluída, ao passo que o CATO retorna a resposta em poucos minutos.

4.5.5.

Problemas Encontrados

Em relação à execução do CATO, nenhum problema foi encontrado com este estudo de caso escolhido. As ontologias de entrada foram alinhadas automaticamente, sem erro algum.