

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

# Meta-alinhamento de ontologias utilizando a abordagem presa-predador

**Nicolas Ferranti**

JUIZ DE FORA  
NOVEMBRO, 2017

# Meta-alinhamento de ontologias utilizando a abordagem presa-predador

NICOLAS FERRANTI

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Bacharelado em Ciência da Computação

Orientador: Jairo Francisco de Souza  
Coorientador: Stênio Sã Rosário Furtado Soares

JUIZ DE FORA  
NOVEMBRO, 2017

# META-ALINHAMENTO DE ONTOLOGIAS UTILIZANDO A ABORDAGEM PRESA-PREDADOR

Nicolas Ferranti

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Jairo Francisco de Souza  
Dr. em Informática (PUC-RIO)

Stênio Sã Rosário Furtado Soares  
Dr. em Computação (UFF)

Victor Ströele de Andrade Menezes  
Dr. em Engenharia de Sistemas e Computação (UFRJ)

Heder Soares Bernardino  
Dr. em Modelagem Computacional (LNCC/MCTI)

JUIZ DE FORA  
30 DE NOVEMBRO, 2017

*Aos meus amigos e irmãos.*

*Aos pais, pelo apoio e sustento.*

## Resumo

Todo ano, diversos novos alinhadores de ontologias são propostos na literatura, cada um utilizando uma premissa diferente, o que implica em desempenhos distintos de acordo com as características das ontologias. Um meta-alinhador consiste de um algoritmo que combina diversas abordagens a fim de obter melhores resultados em diferentes cenários. Para atingir esse objetivo, é necessária a definição de um critério de uso das técnicas. A meta-heurística presa-predador surgiu recentemente por meio do caminho aberto pela introdução dos algoritmos evolucionários e tem como inspiração natural a interação entre animais. Neste trabalho, é apresentado um meta-alinhador de ontologias que combina várias técnicas verificando o uso da abordagem presa-predador no âmbito da parametrização das mesmas. A abordagem é avaliada pelo *benchmark* da OAEI e comparada com a literatura.

**Palavras-chave:** Ontologia, Web Semântica, Alinhamento de Ontologias, Presa-Predador

# Abstract

Every year, several new aligners of ontologies are proposed in the literature, each one using a different premise, which implies in different performances according to the characteristics of the ontologies. A meta-aligner consists of an algorithm that combines several approaches in order to obtain better results in different scenarios. To achieve this goal, it is necessary to define a criterion for the use of techniques. The meta-heuristic prey-predator has recently emerged by way of the introduction of evolutionary algorithms and has as natural inspiration the interaction between animals. In this work, a meta-aligner of ontologies is presented that combines several techniques verifying the use of the prey-predator approach in the scope of the parametrization of the same. The approach is evaluated by the OAEI benchmark and compared to the literature.

**Keywords:** Ontology, Semantic Web, Ontology Matching, Prey-Predator

## Agradecimentos

Aos meus pais Antônio, Silvana e minha família, por todo o exemplo recebido e o apoio incondicional que possibilitou realizar esse sonho.

Aos professores Jairo e Stênio pela orientação, amizade e principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos amigos do Laboratório de Aplicações e Inovação em Computação(LApIC), pelo apoio, companheirismo e conhecimento transferido ao longo desses anos.

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

*“Procure ser um homem de valor, em vez  
de ser um homem de sucesso.”*

*Albert Einstein*



# Conteúdo

|  |           |
|--|-----------|
| <b>Lista de Figuras</b>                                    | <b>7</b>  |
| <b>Lista de Tabelas</b>                                    | <b>8</b>  |
| <b>Lista de Abreviações</b>                                | <b>9</b>  |
| <b>1 Introdução</b>  | <b>10</b> |
| 1.1 Justificativas . . . . .                               | 11        |
| 1.2 Objetivos . . . . .                                    | 12        |
| 1.3 Metodologia . . . . .                                  | 12        |
| <b>2 Revisão da Literatura</b>                             | <b>14</b> |
| 2.1 Técnicas de Alinhamento . . . . .                      | 14        |
| 2.2 Combinando Técnicas . . . . .                          | 15        |
| 2.3 Meta-Alinhamento de Ontologias . . . . .               | 17        |
| 2.4 Trabalhos Relacionados . . . . .                       | 18        |
| <b>3 Abordagem proposta para calibragem de alinhadores</b> | <b>20</b> |
| 3.1 Cálculo da direção . . . . .                           | 22        |
| 3.2 Cálculo do tamanho do passo . . . . .                  | 22        |
| 3.3 Intensificação da solução . . . . .                    | 23        |
| 3.4 Integração do calibrador . . . . .                     | 26        |
| <b>4 Avaliação da proposta</b>                             | <b>28</b> |
| 4.1 Teste de convergência . . . . .                        | 28        |
| 4.2 Avaliação sobre o Benchmark . . . . .                  | 29        |
| <b>5 Considerações Finais</b>                              | <b>35</b> |
| 5.1 Limitações . . . . .                                   | 35        |
| 5.2 Trabalhos futuros . . . . .                            | 36        |
| <b>A Apêndice</b>  | <b>37</b> |
| <b>Bibliografia</b>  | <b>41</b> |

## Lista de Figuras

|     |   |    |
|-----|---|----|
| 2.1 | Classificação das técnicas de alinhamento (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015) . . . . . | 15 |
| 2.2 | Duas ontologias simples e um alinhamento candidato (SHVAIKO; EUZENAT, 2013) . . . . .                           | 16 |
| 3.1 | Fluxograma de execução do algoritmo . . . . .   | 27 |
| 4.1 | Gráfico triangular com cobertura e precisão totais . . . . .  | 33 |

## Lista de Tabelas

|     |  |    |
|-----|--|----|
| 4.1 | Soma das diferenças entre o valor encontrado e o esperado como resultado                 | 29 |
| 4.2 | Resumo descritivo de cada teste . . . . .  | 31 |
| 4.3 | Resultados dos testes empíricos nos três modelos . . . . .                               | 33 |
| 4.4 | Resultados médios considerando um conjunto de execuções no modelo PP<br>Padrão . . . . . | 33 |
| A.1 | Resultados obtidos pelo modelo PP Padrão . . . . .                                       | 38 |
| A.2 | Resultados PP Simples . . . . .  | 39 |
| A.3 | Resultados PP Simples + BLI . . . . .  | 40 |

## Lista de Abreviações

|        |   |
|--------|---|
| AG     | Algoritmo Genético                              |
| API    | <i>Application Programming Interface</i>        |
| BLI    | Busca Local Iterativa                           |
| CPU    | <i>Central Processing Unit</i>                  |
| DCC    | Departamento de Ciência da Computação           |
| HTML   | <i>HyperText Markup Language</i>                |
| HTTP   | <i>Hypertext Transfer Protocol</i>              |
| LOD    | <i>Linking Open Data</i>                        |
| OAEI   | <i>Ontology Alignment Evaluation Initiative</i> |
| OWL    | <i>Web Ontology Language</i>                    |
| PP     | Presa-Predador                                  |
| PPOA   | <i>Prey-Predator Ontology Aligner</i>           |
| RDF    | <i>Resource Descripton Framework</i>            |
| SPARQL | <i>SPARQL Protocol and RDF Query Language</i>   |
| SV     | <i>Survivor Value</i>                           |
| UFJF   | Universidade Federal de Juiz de Fora            |
| URI    | <i>Uniform Resource Identifier</i>              |

# 1 Introdução

Com o passar dos anos, a necessidade de troca de informações de maneira rápida e eficiente se tornou um requisito para a maioria dos sistemas. Algumas décadas atrás, no formato em que se encontrava, a web não estava preparada para o crescimento ocorrido no volume de dados. Para melhorar os meios de busca, era necessário desenvolver meios de padronizar os dados e estabelecer relações semânticas entre eles. Dados publicados nesses formatos padronizados atendem aos princípios da Web Semântica.

A Web Semântica não é uma Web separada, mas uma extensão da atual. Nela a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas (BERNERS-LEE; HENDLER; LASSILA, 2001). No contexto da Web Semântica, o termo Dados Ligados (*Linked Data*) é utilizado para descrever um conjunto de práticas para publicar, compartilhar e conectar dados estruturados na Web (BIZER; HEATH; BERNERS-LEE, 2009) sobre diversos temas.

Graças à fácil aceitação da Web Semântica e de Dados Ligados, tanto pela área acadêmica quanto pelo mercado, pessoas e organizações por todo o mundo começaram a desenvolver seus próprios conceitos para classificar os tipos de dados que trabalham, culminando no desenvolvimento de diversas ontologias que descrevem o domínio específico de cada dado, formando classes que permitem padronizar atributos e relações que dados de um mesmo domínio compartilham (SOUZA, 2012). Uma ontologia é um modelo de representação de dados de maneira estruturada que descreve conceitos dentro de um domínio (GRUBER, 1993).

Em parte, alguns problemas puderam ser solucionados, como a representação do conhecimento. Entretanto, novos desafios surgiram, pois cada ontologia é criada por organizações distintas com interesses distintos, logo, uma vez que ontologias são construídas para diferentes finalidades, por pessoas com especialidades e habilidades distintas e com diferentes visões de domínio, existem ontologias de mesmo domínio, ou domínios complementares, construídas com estruturas, nomes e características distintas (SOUZA, 2012). O alinhamento de ontologias é um processo que procura estabelecer relações entre

elementos das ontologias.

O desafio é, dado um cenário de integração entre bases, definir a relação entre conceitos das ontologias envolvidas, compatibilizando as estruturas de forma a representar a união dos conjuntos de dados em um novo modelo. Várias características do problema podem ser analisadas no desenvolvimento de uma abordagem para solucioná-lo, logo, muitos alinhadores podem ser encontrados na literatura, entretanto, seus desempenhos se limitam ao grupo de ontologias que se enquadram nas premissas definidas pelas técnicas que cada um emprega. As premissas estão relacionadas com as características que cada alinhador visa explorar das ontologias. O principal desafio deste trabalho é combinar as abordagens alinhadoras de forma a permitir que a solução se adapte a diversos tipos de ontologias de entrada.

## 1.1 Justificativas

Ontologias são construídas por pessoas com diversos níveis de especialização e visão de domínio, logo, conceitos que podem descrever o mesmo tipo de objeto podem se encontrar representados de forma distinta, tanto na sintaxe dos termos quanto na estrutura de relações, gerando um problema de heterogeneidade na semântica dos dados. Para solucionar problemas de heterogeneidade, é preciso uma forma de especificar, sem ambiguidade, os vocabulários subjacentes aos sistemas de informação (FARINELLI; ALMEIDA, 2014).

O problema de alinhamento de ontologias é complexo e suas características possibilitam que seja abordado por diversas técnicas computacionais. Lambrix e Tan (2006) propõem um alinhador para trabalhar com ontologias da área de biomedicina utilizando diversas métricas de similaridade sintática, como n-gram e distância de edição. Zhang et al. (2009) utilizam o WordNet como *background knowledge*, agregando valor semântico aos conceitos e propriedades em conjunto com uma análise de similaridade estrutural entre os grafos. Como não existe uma técnica que se sobressaia dentre as outras em todos os aspectos, é comum empregar abordagens de meta-alinhamento.

Um meta-alinhador combina diversas técnicas de alinhamento, a fim de explorar vários aspectos da heterogeneidade para evitar que o desempenho do alinhamento seja restrito a alguma característica das ontologias. Faria et al. (2016) apresenta uma ferramenta

que combina abordagens alinhadoras com pesos distintos definidos previamente para combinar estratégias, entretanto a abordagem não pode ser vista como meta-alinhamento uma vez que a combinação dos resultados é predeterminada.

Algoritmos evolucionários são empregados nos meta-alinhadores recentes, pois permitem que a solução adapte os parâmetros de cada função de similaridade em tempo de execução, podendo utilizar e testar um conjunto de técnicas de agrupamento final.

## 1.2 Objetivos

O objetivo deste trabalho é propor o uso da meta-heurística presa-predador aplicada no cenário de meta-alinhamento de ontologias. Espera-se que a técnica seja eficiente e eficaz ao parametrizar os métodos de alinhamento de forma a obter uma solução que se aproxime da ótima em tempo polinomial. A solução adotada por este trabalho, objetiva modelar o problema como um problema de otimização. Um problema de otimização é um problema de encontrar um elemento de um conjunto chamado conjunto de restrições que otimize uma determinada função objetivo (TILAHUN; ONG, 2015).

Este trabalho pretende avaliar o uso da meta-heurística presa-predador como uma nova técnica a ser aplicada no cenário de alinhamento de ontologias.

## 1.3 Metodologia

Para avaliar abordagens para alinhamento de ontologias, geralmente é utilizada a base de avaliação da OEAI<sup>1</sup> (*Ontology Alignment Evaluation Initiative*). A OEAI é uma iniciativa internacional coordenada para avaliar e chegar a um consenso sobre as diversas técnicas de alinhamento existentes, possuindo um conjunto de teste diversificado e apto a validar os pontos fortes e fracos de cada métrica. Este trabalho utilizará o *benchmark* da OEAI para avaliar o desempenho do meta-alinhador. O objetivo será atingido se o comportamento do sistema se adaptar a diversos tipos de ontologia, alcançando as relações esperadas pelo conjunto de teste.

Algoritmos evolutivos são frequentemente empregados para este problema, a

---

<sup>1</sup><http://oaei.ontologymatching.org/>

---

meta-heurística presa-predador será avaliada no que tange a adaptação ao problema. Também será avaliado o comportamento da meta-heurística em relação a outras, em específico, um algoritmo genético.



## 2 Revisão da Literatura

O problema de alinhamento de ontologias pode ser tratado de diversas formas. Este capítulo traz uma revisão dos métodos de alinhamento e das técnicas de meta-alinhamento presentes na literatura.

### 2.1 Técnicas de Alinhamento

Segundo Otero-Cerdeira, Rodríguez-Martínez e Gómez-Rodríguez (2015), técnicas de alinhamento de ontologias podem ser classificadas em níveis, sendo o primeiro dividido em:

- Alinhadores em nível dos elementos: técnicas que obtêm as correspondências considerando as entidades nas ontologias isoladamente, ignorando que são partes da estrutura da ontologia.
- Alinhadores a nível estrutural: técnicas que obtêm as correspondências analisando como as entidades se encaixam dentro da estrutura das ontologias.
- Baseadas em conteúdo: técnicas com foco na informação interna proveniente das ontologias que serão alinhadas.
- Baseadas em contexto: essas técnicas consideram para a correspondência as informações externas que podem surgir de relações entre ontologias ou outros recursos externos (contexto).

A Figura 2.1 apresenta de forma mais especializada os níveis e a classificação de técnicas de alinhamento de ontologias.

Com base no sistema de classificação, é possível agrupar as técnicas presentes na literatura por categoria, tornando mais fácil a comparação entre elas. Akbari, Fathian e Badie (2009) apresenta uma medida de similaridade baseada na distância de Levenshtein para a comparação de strings aplicada em alinhamento de ontologias. Técnicas de comparação de strings se baseiam no nome e na descrição das entidades da ontologia.

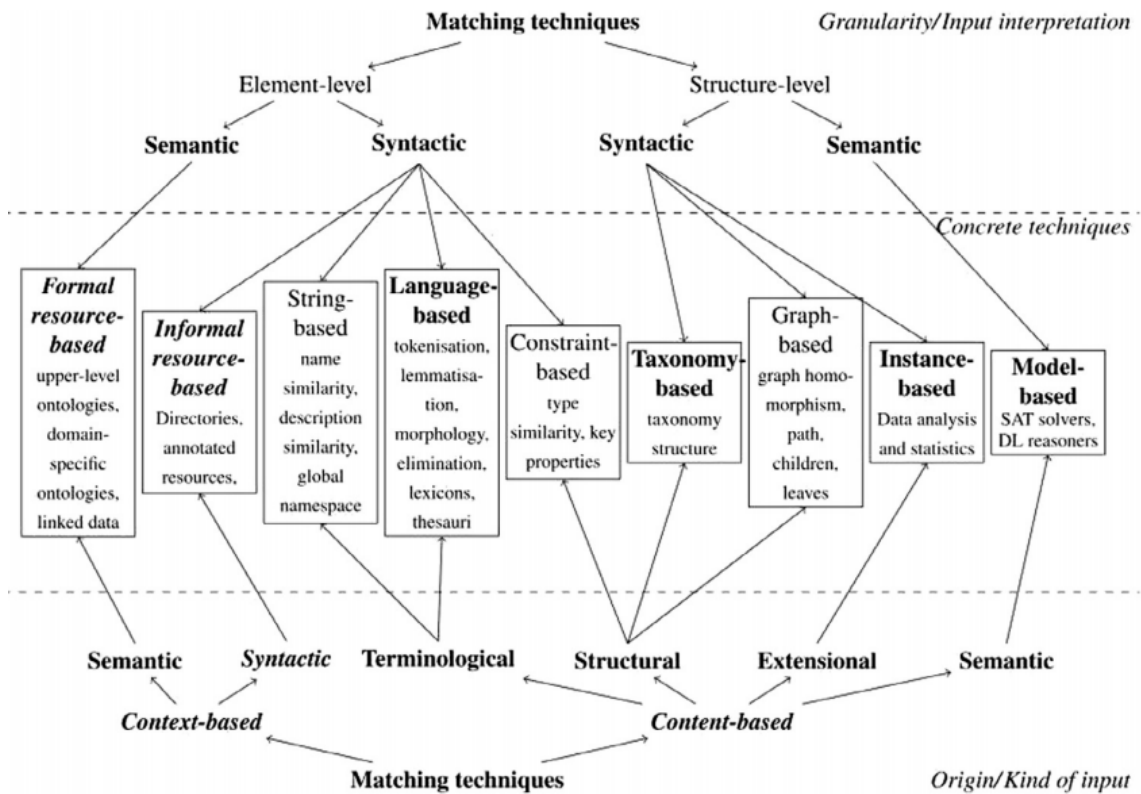


Figura 2.1: Classificação das técnicas de alinhamento (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015)

Existem várias métricas para cálculo de distância entre strings que podem ser usadas nesses métodos como Jaccard, n-gram, Levenshtein, TFIDF, euclidiana etc. Joslyn, Paulson e White (2009) aplica técnicas da teoria dos grafos ao problema. Essas técnicas consideram as ontologias a serem alinhadas como grafos rotulados, ou até mesmo árvores, caindo no problema de isomorfismo entre dois grafos. Não existe uma técnica que se sobressaia sobre as demais de forma genérica, um exemplo pode ser observado na Figura 2.2. A Figura 2.2 apresenta duas ontologias e um alinhamento candidato. Enquanto a comparação de strings identifica facilmente a equivalência entre os atributos *title*, a análise do grafo pode indicar que *Monograph* e *Essay* são menos gerais do que *Book*. Logo, é interessante estudar o uso dessas técnicas em conjunto para melhorar a acurácia de um alinhamento.

## 2.2 Combinando Técnicas

Diversos trabalhos exploram a possibilidade de empregar técnicas de alinhamento distintas em conjunto, esta seção apresenta, através de um trabalho, como isto pode ser feito.

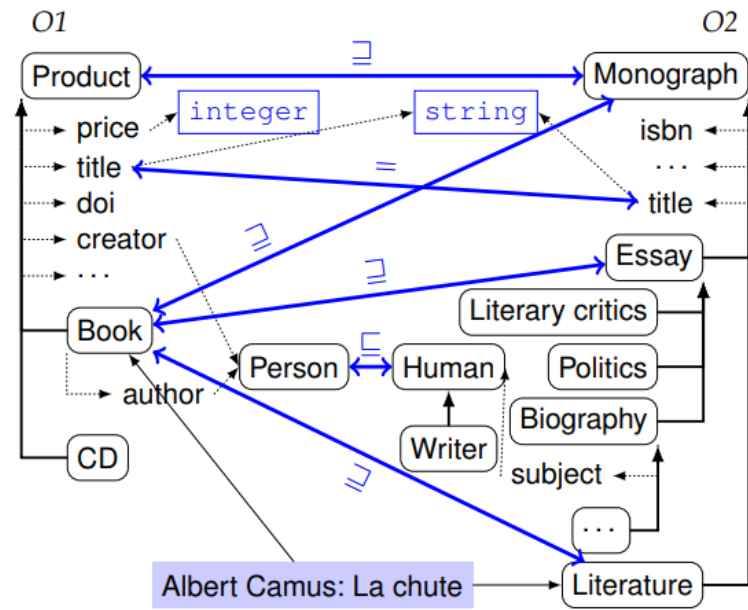


Figura 2.2: Duas ontologias simples e um alinhamento candidato (SHVAIKO; EUZENAT, 2013)

Com o uso de técnicas de alinhamento híbridas, o Lily, um sistema de alinhamento de ontologia, é capaz de resolver alguns problemas relacionados a ontologias heterogêneas (WANG; WANG, 2016). Avaliado pelo *benchmark* da OAEI em 2016, os resultados alcançados nas duas ontologias de teste foram superiores ou iguais a todos os outros alinhadores com relação ao *F-measure*, uma média harmônica entre os valores de precisão e cobertura que medem a taxa de acertos das correspondências. O Lily constrói um subgrafo semântico na tentativa de eliminar a interpretação heterogênea dos elementos das ontologias. Todo o cálculo de similaridade é realizado sobre esse subgrafo. A similaridade é computada por meio de técnicas de similaridade entre strings e similaridade estrutural. Ao final, as similaridades computadas são combinadas utilizando pesos experimentalmente definidos. Os pesos são fundamentais para definir o nível de confiança para uma dada abordagem e, nesse caso, são atribuídos estaticamente, prejudicando o desempenho da solução em ontologias cuja experimentação não foi aplicada. Uma alternativa viável é o desenvolvimento de meios para que o algoritmo possa se adaptar dinamicamente a uma dada entrada e atribuir automaticamente os pesos mais adequados para aquela ontologia. Abordagens que calibram o peso de técnicas de alinhamento em tempo de execução são denominadas meta-alinhadoras de ontologias.

## 2.3 Meta-Alinhamento de Ontologias

O termo meta-alinhamento de ontologias foi apresentado por Euzenat, Shvaiko et al. (2007) e descreve sistemas que parametrizam automaticamente um conjunto de funções de alinhamento de ontologias. Martinez-Gil e Aldana-Montes (2012) define um conjunto de características comuns no que tange aos meta-alinhadores de ontologias:

1. Não é necessário que o processo de meta-alinhamento seja realizado em tempo de execução. As funções de alinhamento podem ser computadas em *background* e aplicadas em tempo de execução uma vez que o processo executado por elas é determinista e as relações não mudam de uma execução para a outra.
2. O processo de meta-alinhamento deve ser automático, logo, deve ser possível que seja implementado por alguma ferramenta de alinhamento.
3. O processo deve se comportar como um especialista, caso a melhor função de alinhamento não seja conhecida, o processo deve ser capaz de experimentar pesos e combinações a fim de retornar a função mais próxima possível da melhor função de alinhamento.
4. Uma estratégia de meta-alinhamento é avaliada com a função de alinhamento retornada.

O meta-alinhamento lida com a integração de alinhadores heterogêneos, visando encontrar os melhores parâmetros que possam afetar os resultados do alinhamento. O problema de meta-alinhamento é modelado como um problema de otimização e as abordagens mais importantes empregam heurísticas em conjunto com algoritmos evolutivos, gulosos ou baseados em conjuntos de regras (SOUZA et al., 2014). Como a literatura não apresenta o uso da meta-heurística presa-predador para a calibração de funções de alinhamento, e algoritmos evolucionários são frequentemente empregados para esse fim, este trabalho experimentou o uso do presa-predador para a calibragem das funções de alinhamento.

## 2.4 Trabalhos Relacionados

Esta seção destaca alguns trabalhos relacionados com foco em sistemas de meta-alinhamento.

Souza (2012) apresenta o GNoSIS+, um meta-alinhador de ontologias que utiliza um algoritmo genético (AG) para parametrizar um conjunto preestabelecido de técnicas de alinhamento. O aprendizado do algoritmo é baseado em um grupo de alinhamentos de referência definidos na entrada por um engenheiro de ontologias. A premissa é que alguns relacionamentos podem ser facilmente apontados, então o AG calibra as funções do sistema baseado na referência a fim de prepará-lo para uma situação real de aplicação. É interessante destacar a representação do problema pelo GNoSIS+. Considerando  $\Xi = \{F_1, F_2, \dots, F_n\}$  um conjunto de funções de alinhamento, cada cromossomo possui  $n$  genes ( $|\Xi| = n$ ) e cada gene representa um valor real  $w \in [0, 1]$  que representa o peso a ser aplicado sobre cada função. O objetivo é minimizar a diferença entre o valor encontrado e o valor definido pelo engenheiro de ontologias para um relacionamento em específico. Xue e Tang (2017) também emprega um algoritmo evolucionário com a mesma representação de indivíduo, entretanto, a função objetivo passa a ser maximizar o valor da média harmônica do *f-measure*. O *f-measure* é uma medida que leva em conta as taxas de precisão e cobertura entre os mapeamentos obtidos pelo algoritmo com os que eram esperados. A função objetivo de cada trabalho, guia os respectivos algoritmos para caminhos diferentes. Para a abordagem de Xue e Tang (2017), é necessário avaliar cada item do resultado obtido a cada iteração com a base de referência, acarretando em um custo computacional maior do que apenas comparar o resultado obtido com o valor de confiança definido pelo engenheiro, como é feito por Souza (2012).

O MapPSO é uma solução que emprega a técnica de enxame de partículas para lidar com o problema de alinhamento (BOCK; HETTENHAUSEN, 2012). O enxame de partículas é uma técnica com inspiração natural baseada no comportamento social de indivíduos, como, por exemplo, a reunião de pássaros para encontrar um local com alimento suficiente (SHI; EBERHART, 1998). A abordagem busca apenas relações do tipo equivalência (1:1) e é utilizado um alinhador predefinido que implemente uma função de distância. A função de distância define um nível de similaridade para um dado par de conceitos. É importante notar que o MapPSO não calibra um conjunto de funções

alinhadoras pois utiliza apenas uma, entretanto pode ser considerado uma abordagem de meta-alinhamento por buscar um alinhamento ótimo fazendo uso de alinhadores predefinidos. A representação do indivíduo difere dos trabalhos anteriores. Nesta abordagem, cada solução é representada como um alinhamento candidato. Suponha que  $\vec{X}_p$  represente um alinhamento de duas ontologias constituído de  $k = 5$  correspondências ( $c$ ). A partícula é representada por  $\vec{X}_p = \{c_{(p,1)}, c_{(p,2)}, c_{(p,3)}, c_{(p,4)}, c_{(p,5)}\}$  onde cada  $c_{(p,i)}$  indica um valor confiança para o relacionamento  $(p, i)$ .

A função objetivo do MapPSO busca encontrar a maior quantidade de alinhamentos possíveis, podendo prejudicar o desempenho em ontologias onde por natureza a taxa de correspondências é baixa.

A literatura mostra que abordagens evolutivas apresentam bons resultados quando aplicadas no alinhamento de ontologias, o que permite fundamentar que o uso da meta-heurística presa-predador tem potencial para construir resultados efetivos para o problema.

A definição da representação do indivíduo impacta na forma como o esforço da abordagem pode ser reproduzido. Uma vez que a representação seja baseada no conjunto de pesos, os parâmetros encontrados podem ser armazenados e recuperados sem muito esforço, enquanto que a representação baseada no conjunto de alinhamentos candidatos requer que todo o processamento seja executado novamente. Logo, os pesos das abordagens de conjuntos de pesos tem contribuição melhor para a construção de um meta-alinhador mais genérico.

### 3 Abordagem proposta para calibragem de alinhadores

Para tratar o problema é aplicada a meta-heurística presa-predador baseada na interação entre animais. A primeira versão da meta-heurística foi introduzida inicialmente por Laumanns, Rudolph e Schwefel (1998). Na abordagem de Laumanns, Rudolph e Schwefel (1998) as presas não se movem naturalmente, estão sujeitas ao movimento dos predadores para que então possam responder de forma a se adaptar no espaço de solução, melhorando a qualidade da solução. Tilahun e Ong (2015) apresentam uma abordagem também baseada na interação presa-predador entre animais, entretanto, o comportamento e a forma como os indivíduos interagem entre si se difere dos outros trabalhos. O trabalho de Tilahun e Ong (2015) é a base para o desenvolvimento deste trabalho.

Considere  $S$  um conjunto de correspondências conhecidas de equivalência. O conjunto  $S$  é formado por tuplas  $(e_{1i}, e_{2i}, =, s_i)$ , onde  $e_{1i}$  e  $e_{2i}$  são entidades de ontologias distintas,  $=$  denota a relação do tipo equivalência e  $s_i$  é a similaridade conhecida, informada pelo engenheiro de ontologias, entre  $e_{1i}$  e  $e_{2i}$ . Seja  $f$  uma função de similaridade composta da soma ponderada de outras funções, ao aplicar a função  $f$  em  $e_{1i}$  e  $e_{2i}$ , espera-se encontrar o valor  $s_i$ , ou seja,  $f(e_{1i}, e_{2i}) = s_i$ . Como exemplo, considere o conjunto  $S' = (e_{11}, e_{21}, =, 1), (e_{12}, e_{22}, =, 1), (e_{13}, e_{23}, =, 1)$ , com todas as correspondências possuindo similaridade igual a 1. Considerando uma função  $\bar{f}'(e_{11}, e_{21}) = g_1(e_{11}, e_{21})p_1 + g_2(e_{11}, e_{21})p_2 + g_3(e_{11}, e_{21})p_3$ , onde  $g_i$  representa o valor de similaridade definido pela função  $i$  que são constantes do problema e  $p_i$  representa o peso atribuído à função  $i$ . Logo, para cada alinhamento conhecido fornecido na entrada, é possível construir, como visto em Souza (2012), um sistema linear:

$$\begin{aligned}
\bar{f}'(e_{11}, e_{21}) &= s_1 \cdot g_1(e_{11}, e_{21})p_1 + g_2(e_{11}, e_{21})p_2 + g_3(e_{11}, e_{21})p_3 = 1 \\
\bar{f}'(e_{12}, e_{22}) &= s_2 \cdot g_1(e_{12}, e_{22})p_1 + g_2(e_{12}, e_{22})p_2 + g_3(e_{12}, e_{22})p_3 = 1 \\
\bar{f}'(e_{13}, e_{23}) &= s_3 \cdot g_1(e_{13}, e_{23})p_1 + g_2(e_{13}, e_{23})p_2 + g_3(e_{13}, e_{23})p_3 = 1
\end{aligned} \tag{3.1}$$

O objetivo é encontrar os melhores valores de  $p_i$  de forma a minimizar a soma das diferenças entre o valor encontrado e o valor esperado. Como a meta-heurística presa-predador é populacional, a representação do indivíduo é baseada no conjunto de pesos  $p_i$ . Para modelar o problema, cada indivíduo recebe um conjunto de valores reais  $p_i$ , inicialmente dentro do intervalo  $[0, 1]$  cuja alteração no valor impacta diretamente na confiança atribuída às funções associadas.

Na abordagem desenvolvida, um conjunto de soluções  $p_i$  viáveis é construído de forma aleatória. Para cada solução  $x_i$ , é atribuído um valor de sobrevivência,  $SV(x_i)$ , que é diretamente ligado à função objetivo do problema. Seja  $F(x)$  a função objetivo, descrita como a soma das diferenças, ou seja, onde quanto menor o valor, melhor é a avaliação da solução, definimos:

$$SV(x_i) = 1/F(x_i) \tag{3.2}$$

Isso pode ser considerado como quão bem localizada está uma presa para fugir de um predador ou a força de uma presa para ultrapassar o predador. Após o valor de sobrevivência (SV) de cada membro da solução ser calculado, o membro com o menor SV será designado como um predador e o resto como presas. Uma vez que as presas e o predador são definidos, as presas precisam fugir do predador e tentar seguir as melhores presas em termos de valores de sobrevivência ou encontrar um bom esconderijo ao mesmo tempo. O que leva à definição de como se dará esse movimento. Ao tratar do movimento, é preciso definir duas questões: a direção e o tamanho do passo.



### 3.1 Cálculo da direção

Considerando que as presas precisam fugir ou tentar se esconder, é definida uma probabilidade de acompanhamento que decide se a presa deve seguir as mais aptas ou procurar se esconder na vizinhança. Caso uma presa  $x_i$  escolha seguir as demais, tomando  $\{x_1, x_2, \dots, x_p\}$  como o conjunto de presas com valor de sobrevivência maior que  $x_i$ , o cálculo da nova direção, definido por Tilahun e Ong (2015), é dado por:

$$y_i = \sum_{j=1}^p e^{SV(x_j)^\tau - r_{ij}} (x_j - x_i) \quad (3.3)$$

Onde,  $r_{ij}$  representa a distância entre as duas presas e  $\tau$  um valor escolhido para ponderar o peso do valor de sobrevivência. Se a probabilidade de seguir não for alcançada, uma direção aleatória  $y_r$  é construída e então avaliada na presa  $x_i$  calculando a distância para o predador.

$$d_1 = \|x_{predador} - (x_i + y_i)\| \quad (3.4)$$

$$d_2 = \|x_{predador} - (x_i - y_i)\| \quad (3.5)$$

Se  $d_1 < d_2$  então tomar a direção  $-y_r$  faz com que a presa  $x_i$  fique mais distante do predador, caso contrário é utilizada  $y_r$ . Após os cálculos de direção, fugindo ou seguindo, é necessário calcular quanto o indivíduo vai caminhar na direção encontrada.

### 3.2 Cálculo do tamanho do passo

O tamanho do passo define o quão longe a presa vai caminhar na direção escolhida. É importante ressaltar que a natureza do problema é contínua tornando inviável a exploração de todo o espaço de busca, logo a definição do passo é importante para que não se perca uma boa solução no meio do caminho. Baseados na premissa de que uma presa longe do predador não correrá tão rápido quanto uma perto, Tilahun e Ong (2015) definem o passo como:

$$\lambda_i = \frac{\lambda_{MAX}\varepsilon_1}{e^{\beta|SV(x_i)-SV(x_{predador})|}^\omega} \quad (3.6)$$

Onde,  $\lambda_{MAX}$  representa o maior tamanho do passo,  $\varepsilon_1$  um número escolhido aleatoriamente de forma uniforme no intervalo  $[0, 1]$  e as constantes  $\beta$  e  $\omega$  são definidas previamente antes da execução do algoritmo. Neste trabalho foi acrescentado um valor de granularidade ( $G$ ) na equação do passo que contribui para a discretização do problema, generalizando a equação de movimento das presas temos, para a direção  $y_i$  adequada à escolha de seguir ou fugir:

$$x_i \leftarrow x_i + G\lambda_i\left(\frac{y_i}{\|y_i\|}\right) \quad (3.7)$$

O predador sempre se movimentará na direção da presa com pior SV, com um certo nível de aleatoriedade, como descreve a Equação 3.8:

$$x_{predador} \leftarrow x_{predador} + \lambda_{MAX}(\varepsilon_5)\left(\frac{y_r}{\|y_r\|}\right) + \lambda_{MIN}(\varepsilon_6)\left(\frac{x'_i - x_{predador}}{\|x'_i - x_{predador}\|}\right) \quad (3.8)$$

Onde  $\lambda_{MIN}$  e  $\lambda_{MAX}$  são constantes definidas previamente representando o passo mínimo e máximo respectivamente,  $\varepsilon_5$  e  $\varepsilon_6$  são valores reais aleatórios no intervalo  $[0,1]$ ,  $y_r$  uma direção gerada aleatoriamente e  $x'_i$  representa a posição da pior presa.

### 3.3 Intensificação da solução

Caso a presa em avaliação seja a de melhor SV em toda a população, não ocorre caminhar, segundo Tilahun e Ong (2015) é aconselhável que nesta presa seja executado um processo de intensificação da solução a cada iteração. No trabalho desenvolvido, é utilizada a busca local apresentada por Souza (2012). Um algoritmo de busca local define, para cada solução, uma vizinhança composta por um conjunto de soluções com características muito próximas (AARTS; LENSTRA, 2003). A busca local utilizada percorre todos os pesos de uma solução criando duas novas soluções para cada peso visitado. O processo de criação se dá pela soma e subtração do valor de granularidade ( $G$ ) em cada peso do indivíduo, ou seja, se  $x_i$  é um indivíduo com conjunto de pesos  $(g_1, g_2, \dots, g_n)$ , ao iterar sobre o primeiro peso são criadas duas novas soluções  $(g_1 + G, g_2, \dots, g_n)$  e  $(g_1 - G, g_2, \dots, g_n)$ .

O processo é executado para cada peso e, ao final, o melhor aprimorante é escolhido para substituir o antigo indivíduo se sua aptidão for superior.

Com a definição das funções de movimento e de intensificação, os passos do algoritmo podem ser especificados como:

1. Definir os parâmetros e gerar um conjunto de soluções viáveis
2. Calcular o valor de sobrevivência para cada presa e definir a melhor presa, o predador e as presas restantes
3. Fazer com que as presas e o predador se movimentem segundo as funções descritas anteriormente
4. Se o critério de parada for atendido, terminar a execução, senão, voltar ao passo 2

No intuito de diversificar a população criada pelo algoritmo, foi definido um número  $\kappa$  que representa a quantidade de vezes que o processo deve se repetir, executando todos os passos desde a criação da população até a busca local ao final, caso seja

determinado. O pseudocódigo a seguir apresenta a estrutura do algoritmo.

---

**Algoritmo 1: ALGORITMO PRESA-PREDADOR**

---

**Entrada:** Parâmetros do sistema

**Saída:** Vencedor (Melhor solução)

```

1 início
2   Cria população inicial P de tamanho  $m$  atribuindo SV aos indivíduos;
3   Vencedor  $\leftarrow$  Melhor Solução de P; // Melhor solução do algoritmo
4   PresaI  $\leftarrow$  Melhor Solução de P; // Melhor solução da população atual
5   para  $j$  de 0 até  $\kappa$  faça
6     para  $i$  de 0 até Número de iterações faça
7       Criar nova população vazia P';
8       para  $p$  de 1 até  $m-1$  faça
9         Probabilidade de seguir  $\leftarrow$  Valor Randômico;
10        se Probabilidade de seguir  $\leq$  Prob. Parâmetro então
11          | P' = P'  $\cup$  Movimenta Seguindo(p)
12        senão
13          | P' = P'  $\cup$  Movimenta Fugindo(p)
14        fim se
15      fim para
16      P' = P'  $\cup$  Movimenta Predador(Pior solução de P);
17      PresaI  $\leftarrow$  Busca Local (Melhor Solução de P', Granularidade);
18      P' = P'  $\cup$  PresaI;
19      Atribui SV para cada indivíduo de P';
20      P  $\leftarrow$  P';
21    fim para
22    se  $SV(PresaI) > SV(Vencedor)$  então
23      | Vencedor  $\leftarrow$  PresaI;
24    fim se
25    Cria população inicial P de tamanho  $m$  atribuindo SV aos
      indivíduos;
26  fim para
27 fim
28 retorna Vencedor

```

---

## 3.4 Integração do calibrador

O trabalho desenvolvido faz uso da ferramenta de alinhamento apresentada em Souza (2012). A ferramenta possui arquitetura distribuída e permite que o módulo calibrador de parâmetros seja substituído por outras abordagens desde que o padrão das mensagens seja mantido, respeitando o fluxo de execução. O fluxo de execução se inicia com os arquivos de entrada. Considerando que a abordagem é supervisionada, cada teste precisa informar o conjunto de pré-alinhamentos de treinamento, juntamente com o par de ontologias, o alinhamento de referência completo para avaliação final e por último, o conjunto de funções que serão utilizadas para treinar o algoritmo. As funções de alinhamento são pré-cadastradas no sistema e fazem uso de métricas distintas para avaliar aspectos diversificados das entidades da ontologia. Algumas das métricas das funções presentes no conjunto teste que avaliou o sistema são apresentadas a seguir.

- Avaliar a similaridade das entidades com base na similaridade entre os comentários que descrevem as entidades
- Definir o grau de similaridade baseado na semelhança entre os termos da entidade
- Calcular a similaridade com base nos identificadores em comum que compõem um subgrafo das relações de cada entidade
- Avaliar a semelhança das entidades com base nas instâncias de mesmo identificador
- Calcular a similaridade das entidades com base na semelhança entre os termos que identificam propriedades de tipo de dado e de objeto

Visto que os arquivos de entrada foram informados corretamente, dentro dos padrões estabelecidos, a entrada é processada e modelada computacionalmente, construindo o sistema linear que será otimizado e as estruturas que avaliarão os pesos encontrados. Em seguida, o calibrador utilizando o presa-predador é acionado e, após o término de sua execução, o meta-alinhador computa os alinhamentos candidatos utilizando os pesos encontrados e retorna as equivalências mais relevantes. Para escolher quais alinhamentos são mais relevantes, foi adotado um método que computa as funções e os pesos para cada par de entidades candidatas, ordena pelos maiores graus de similaridade encontrados e

seleciona sempre o maior par como relacionamento escolhido, removendo os escolhidos do restante da lista. A Figura 3.1 apresenta um diagrama de fluxo que reflete a sequência dos dados dentro do sistema.

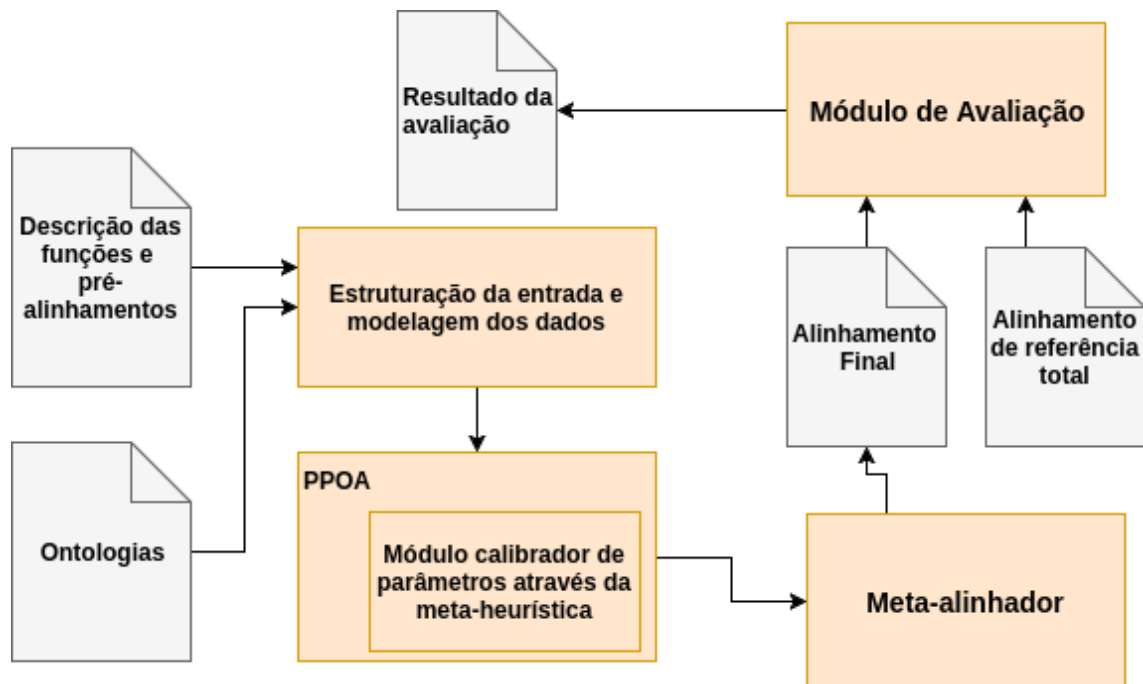


Figura 3.1: Fluxograma de execução do algoritmo

## 4 Avaliação da proposta

Algoritmos heurísticos são difíceis de avaliar, principalmente os que possuem aspectos randômicos (HOOS; STUTZLE, 2007). Logo, é comum utilizar métodos empíricos na avaliação. Embora haja uma definição matemática para o algoritmo, essa informação não é suficiente para concluir sobre os variados aspectos do seu comportamento. Como a abordagem evolutiva empregada possui diversos parâmetros que regulam seu funcionamento, a seção abaixo apresenta o comportamento deste trabalho em um teste de convergência.

### 4.1 Teste de convergência

Para avaliar o comportamento do algoritmo, é necessário definir alguns parâmetros de execução. Foi utilizado um valor de 0,005 para a granularidade, o valor foi reproduzido nos testes. Tilahun e Ong (2015) utilizam valores de  $\beta = 1$  e  $\omega = 1$ , que foram replicados nos testes desse modelo. Os demais parâmetros foram definidos empiricamente como: chance de seguir em 50%,  $\tau = 0,09$ ,  $\lambda_{MAX} = 20$  e  $\lambda_{MIN} = 1$ . Ainda, é definido um limite inferior para a diferença alcançada com a finalidade de evitar erros numéricos ao longo da execução, o limite é da ordem de 1E-11. Os resultados apresentados na Tabela 4.1 expõem o comportamento do algoritmo através da melhor solução encontrada, variando parâmetros de tamanho da população e número de iterações aplicados na Equação 4.1. Como a semente é randômica, pode ocorrer variação entre os resultados encontrados de uma execução para outra.

$$\left\{ \begin{array}{l}
 f_{a_1} = 0,4x + 0,1y + 0,2z \\
 f_{c_1} = f_{a_1}a + 0,1b \\
 f_{c_1}c + 0,2d + 0,4e = 1 \\
 \\
 f_{a_2} = 0,3x + 0,2y + 0,4z \\
 f_{c_2} = f_{a_2}a + 0,7b \\
 f_{c_2}c + 0,1d + 0,1e = 1 \\
 \\
 f_{a_3} = 0,5x + 0,1y + 0,1z \\
 f_{c_3} = f_{a_3}a + 0,4b \\
 f_{c_3}c + 0,7d + 0,2e = 1
 \end{array} \right. \quad (4.1)$$

Tabela 4.1: Soma das diferenças entre o valor encontrado e o esperado como resultado

| Iterações | População |         |           |            |
|-----------|-----------|---------|-----------|------------|
|           | 50        | 100     | 500       | 1000       |
| 1         | 1.86302   | 1.84622 | 1.81194   | 1.79356935 |
| 100       | 1.36082   | 1.31162 | 1.24825   | 1.2102177  |
| 200       | 0.77582   | 0.7175  | 0.49725   | 0.4201752  |
| 300       | 0.05282   | 0.0425  | 0.0000224 | 0.0000007  |
| 400       | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |
| 500       | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |
| 600       | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |
| 700       | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |
| 800       | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |
| 900       | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |
| 1000      | 0.00033   | 0.00018 | 0.0000224 | 0.0000007  |

É possível observar que o algoritmo alcança um valor mínimo entre as gerações 200 e 400, apontando estagnar em um ponto de mínimo local enquanto outras melhores soluções poderiam ter sido encontradas, embora as soluções sejam muito próximas.

## 4.2 Avaliação sobre o Benchmark

Todo ano a OAEI disponibiliza um conjunto de instâncias para testes de alinhadores de ontologias. Para fins comparativos com outras abordagens da literatura, o algoritmo



desenvolvido foi avaliado sobre o *Benchmark* da OAEI referente ao ano de 2007. Neste *benchmark*, é utilizada uma ontologia como referência dentro do domínio de referências bibliográficas. A ontologia de referência é descrita sobre a linguagem OWL-DL e serializada em RDF/XML, possui 33 classes nomeadas, 24 propriedades de objeto, 40 propriedades de dados, 56 indivíduos nomeados e 20 indivíduos anônimos, além de conter ligações para recursos externos a fim de representar informações não bibliográficas. Os recursos externos expressam conceitos de pessoa, organização e evento. Os testes tomam como base uma ontologia de referência e consistem em analisar o alinhamento dessa ontologia com uma segunda ontologia. A segunda ontologia é distinta em cada teste e é construída a partir de modificações na ontologia de referência, removendo ou alterando parte das informações disponibilizadas no intuito de analisar o comportamento do algoritmo. A Tabela 4.2 apresenta um resumo que relaciona o identificador do teste com as informações que foram removidas ou modificadas. É importante destacar, que a faixa de testes 3xx representa o cenário mais próximo do real, pois objetiva alinhar a ontologia de referência com ontologias reais sobre referências bibliográficas.

Para a execução desse teste, o modelo desenvolvido foi flexibilizado para que a avaliação pudesse contemplar aspectos distintos. A primeira flexibilização se deu pela remoção da busca local do algoritmo, assim apenas as operações de movimento foram executadas, fazendo com que a melhor presa estivesse sujeita a ser ultrapassada (PP Simples). A segunda modificação fez com que a melhor solução encontrada pela primeira flexibilização fosse submetida à busca local apresentada na seção (3.3), avaliando a vizinhança e seguindo para o melhor aprimorante de forma iterativa, enquanto um vizinho melhor pudesse ser encontrado (PP Simples + BLI). Já o terceiro modelo, reproduz o comportamento do algoritmo sem modificações, com o movimento da população de acordo com o que foi apresentado e uma busca local simples na melhor presa, fazendo apenas um movimento na direção do melhor vizinho uma única vez (PP Padrão). Para avaliar a acurácia do sistema foram utilizadas as medidas de Precisão, Cobertura e *f-measure*.

*Definição 1.* Seja A um alinhamento de referência e B um alinhamento obtido pelo modelo,

Tabela 4.2: Resumo descritivo de cada teste

| #Teste | Resumo  |
|--------|---|
| 101    | Ontologias idênticas  |
| 103    | Generalização de linguagem  |
| 104    | Restrição de linguagem  |
| 201    | Nome de entidades ausentes  |
| 202    | Entidades nomeadas aleatoriamente e ausência de comentários       |
| 203    | Comentários ausentes  |
| 204    | Convenção de nomenclatura distinta                                |
| 205    | Sinônimos substituindo o nome de entidades                        |
| 206    | Idioma distinto   |
| 208    | Convenção de nomenclatura distinta e comentários ausentes         |
| 209    | Sinônimos substituindo o nome de entidades e comentários ausentes |
| 210    | Idioma distinto e comentários ausentes                            |
| 221    | Especializações suprimidas  |
| 222    | Hierarquia reduzida   |
| 223    | Hierarquia expandida  |
| 224    | Indivíduos ausentes   |
| 225    | Restrições ausentes   |
| 228    | Propriedades e relações entre objetos ausentes                    |
| 230    | Propriedades mais especializadas                                  |
| 232    | Hierarquia suprimida e indivíduos ausentes                        |
| 233    | Hierarquia suprimida e propriedades ausentes                      |
| 236    | Instâncias e propriedades ausentes                                |
| 237    | Hierarquia reduzida e indivíduos ausentes                         |
| 238    | Hierarquia expandida e indivíduos ausentes                        |
| 239    | Hierarquia reduzida e propriedades ausentes                       |
| 240    | Hierarquia expandida e propriedades ausentes                      |
| 241    | Hierarquia suprimida, indivíduos e propriedades ausentes          |
| 246    | Hierarquia reduzida, indivíduos e propriedades ausentes           |
| 247    | Hierarquia expandida, indivíduos e propriedades ausentes          |
| 301    | Ontologia real sobre bibliografia: BibTeX/MIT                     |
| 302    | Ontologia real sobre bibliografia: BibTeX/UMBC                    |
| 304    | Ontologia real sobre bibliografia: BibTeX/INRIA                   |

a precisão mede a porcentagem da resposta que está correta:

$$Precisão = \frac{|A \cap B|}{|B|} \quad (4.2)$$

*Definição 2.* Seja A um alinhamento de referência e B um alinhamento obtido pelo modelo, a cobertura mede a habilidade de fornecer alinhamentos corretos dentro da referência:

$$Cobertura = \frac{|A \cap B|}{|A|} \quad (4.3)$$

*Definição 3.* Dado uma precisão P e uma cobertura C, o *f-measure* é a média harmônica entre P e C:

$$f\text{-measure} = \frac{2PC}{P + C} \quad (4.4)$$

Todos os testes apresentados neste trabalho foram executados em uma máquina Ubuntu 14.04 LTS, arquitetura 64-bit, com 16GB de memória RAM e processador Intel i7-4790 com oito núcleos de 3.6 GHz. A Tabela 4.3 apresenta a média de precisão, cobertura e *f-measure* para os três modelos dentro das faixas de teste da OAEI. O melhor resultado observado foi o do terceiro modelo, que contempla o movimento dos indivíduos no espaço de solução e a busca local executada na melhor presa que retorna o melhor vizinho imediato a essa presa, caso o mesmo exista. Como o terceiro modelo apresentou resultado médio melhor que os outros, foi realizada uma análise mais aprofundada neste modelo. Uma vez que o algoritmo possui fatores randômicos, a execução foi repetida a fim de avaliar a estabilidade do modelo. O resultado geral é apresentado na Tabela A.1 e os resultados médios para um conjunto de 4 execuções são apontados na Tabela 4.4.

Com o resultado médio da Tabela 4.4, o presa-predador, denominado PPOA, foi comparado ao GNoSiS+, o AG de Souza (2012), na Figura 4.1. A Figura 4.1 apresenta valores crescentes de cobertura do meio para a esquerda e, da mesma forma, valores crescentes de precisão do meio para a direita. Assim, quanto mais alto e mais centrado o ponto estiver, maior será o valor de harmonização entre as duas medidas e, por consequência, melhor será a qualidade do modelo. Enquanto o GNoSiS+ obteve média total 96%, a abordagem desenvolvida neste trabalho esteve na margem de 90%.

Tabela 4.3: Resultados dos testes empíricos nos três modelos

|                | Faixa de instâncias | Valores Médios dentro da faixa |              |              |
|----------------|---------------------|--------------------------------|--------------|--------------|
|                |                     | Precisão                       | Cobertura    | $f$ -measure |
| PP Simples     | 101-104             | 1                              | 1            | 1            |
|                | 201-247             | 0,92                           | 0,92         | 0,92         |
|                | 301-304             | 0,7633                         | 0,7367       | 0,749762963  |
|                | Média Total:        | 0,9134375                      | 0,9153125    | 0,9143740388 |
| PP Simples+BLI | 101-104             | 1                              | 1            | 1            |
|                | 201-247             | 0,9396                         | 0,946        | 0,9427891387 |
|                | 301-304             | 0,8233333333                   | 0,79         | 0,806322314  |
|                | Média Total:        | 0,93625                        | 0,938125     | 0,9371865622 |
| PP Padrão      | 101-104             | 1                              | 1            | 1            |
|                | 201-247             | 0,9464                         | 0,9528       | 0,9495892165 |
|                | 301-304             | 0,82                           | 0,7866666667 | 0,8029875519 |
|                | Média Total:        | 0,94125                        | 0,943125     | 0,9421865672 |

Tabela 4.4: Resultados médios considerando um conjunto de execuções no modelo PP Padrão

|              | Média        | Desvio Padrão |
|--------------|--------------|---------------|
| Precisão     | 0,90678125   | 0,0562465884  |
| Cobertura    | 0,9084479167 | 0,0560783869  |
| $f$ -measure | 0,9070363222 | 0,0558457086  |

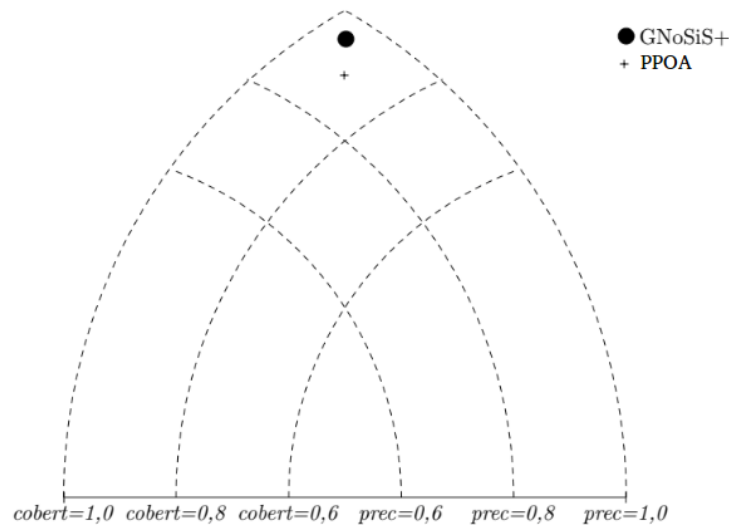


Figura 4.1: Gráfico triangular com cobertura e precisão totais

Ao comparar os resultados do PP Simples com o PP Simples+BLI, é possível ver que a busca local iterativa tem impacto positivo na qualidade das soluções, como apontam as medidas das faixas 2xx e 3xx, aprimorando a melhor solução encontrada pelo PP Simples. Em termos de tempo, as tabelas A.2 e A.3, que exibem os melhores resultados encontrados pelo modelo, informam o tempo associado à execução que obteve o resultado

exposto. Ao comparar o tempo total de execução dos dois testes completos, o resultado inesperado aponta um custo maior na execução do PP Simples do que na execução do PP Simples+BLI, sugerindo um estudo mais aprofundado no que se refere a questões como os impactos da criação de população de maneira randômica e a concorrência pelo tempo de CPU na máquina de teste, algo que não foi avaliado.

Embora os resultados obtidos até então não superem a referência da literatura, o meta-alinhador utilizando o presa-predador como calibrador de pesos apresentou resultados promissores nos parâmetros empregados. Ambos, GNoSiS+ e PPOA, podem ser empregados no mesmo cenário, atribuindo alinhamentos candidatos a ontologias com poucas referências e retornando uma primeira versão de alinhamento para os engenheiros. Outro cenário, onde espera-se que o desempenho seja melhor, é aquele em que se deseja refazer um alinhamento já existente dado que houve alguma alteração em uma das ontologias. A expectativa de melhores resultados se baseia no fato de que o conjunto de alinhamentos de referência de entrada é maior, provendo uma capacidade descritiva maior para o algoritmo. No que tange à escalabilidade, é possível observar que no conjunto de testes, os tempos associados são ínfimos quando comparados à necessidade de obtenção de solução relacionada ao problema, pois as restrições que caracterizam o cenário não se alteram ao longo do tempo, sugerindo que a aplicação poderia ser executada em tempo real sem grandes dificuldades para ontologias com dimensões semelhantes às testadas. Ontologias com maior volume de informação, de larga escala, requerem um tempo maior tanto no calibramento quanto no meta-alinhamento em si. Como este trabalho executa as duas tarefas em tempo de execução, a performance do sistema tende a cair, podendo inviabilizar sua aplicação nesse cenário. A OAEI possui conjuntos de testes que lidam com o alinhamento de ontologias em larga escala.

## 5 Considerações Finais

Este trabalho modelou o conjunto de instâncias de treinamento da OAEI em um sistema linear, buscando minimizar as folgas e excessos em cada linha do sistema para obter a melhor parametrização das funções de alinhamento, o trabalho utilizou a meta-heurística presa-predador para a calibragem dos pesos de cada função. Mesmo com um desvio padrão destoante para algumas instâncias, a natureza do problema em conjunto com a forma como foi modelado permite que, quando uma boa solução for encontrada, os pesos associados a essa solução possam ser persistidos e utilizados para reproduzir o experimento em ontologias que atendam aos critérios estabelecidos pelo teste. Uma vez que os relacionamentos entre as entidades das ontologias estão estabelecidos, não há necessidade de reprocessar o algoritmo para encontrar uma nova solução como ocorre em problemas de roteamento de veículos, onde novas soluções devem ser geradas com frequência pois as restrições variam ao longo do tempo. Logo, encontrar uma solução perto da ótima dentro de um conjunto de execuções em tempo polinomial, é suficiente pois os parâmetros encontrados podem ser aplicados em qualquer momento.

O método de alinhamento utilizado, que seleciona sempre o par de entidades com maior similaridade, pode estar contribuindo com um desvio padrão maior do que o esperado, um método que explore as similaridades encontradas e busque o conjunto de alinhamentos que maximize o somatório das similaridades envolvidas pode contribuir com uma melhor estabilidade do modelo, uma vez que a similaridade correta pode não ser a de maior valor encontrado mas está situada entre as melhores.

### 5.1 Limitações

O desempenho do algoritmo está intimamente ligado com as configurações definidas para execução, embora a minimização da diferença seja uma função objetivo que tenda a bons resultados, dois indivíduos com SV iguais, por exemplo, podem possuir conjuntos de pesos distintos. Isso implica em valores de acurácia diferentes em execuções distintas. O

*benchmark* da OAEI é criado de forma sistemática com base em dados sintéticos, o que dá margem para que possam ocorrer variações de desempenho e acurácia para ontologias do mundo real. Ainda assim, o *benchmark* é difundido na literatura, representando a principal referência para testes de alinhadores de ontologias. Por fim, o desempenho do calibrador de pesos desenvolvido é dependente de quais entidades são fornecidas na entrada, é interessante que as entidades sejam heterogêneas o suficiente para que possam descrever bem a ontologia, fazendo com que funções de natureza distinta sejam empregadas em conjunto sem problemas, isso significa que quanto mais homogênea a entrada for, mais preso à função que descreve aquele comportamento o sistema estará. Entende-se por homogeneidade quando todos os pares de entidades possuem as mesmas características, como, por exemplo, nomes semelhantes, abrindo vantagem para funções baseadas em *strings*.

## 5.2 Trabalhos futuros

A avaliação da abordagem se deu sobre um *benchmark* relativamente antigo apenas para fins comparativos com outras abordagens da época, futuramente, os resultados serão computados novamente sobre o *benchmark* mais recente da OAEI. O desempenho do algoritmo desenvolvido está atrelado às configurações definidas pelo usuário, logo é interessante submeter o sistema a algum algoritmo que possa calibrar os parâmetros de forma adequada. Existem outros métodos de composição dos resultados baseados na maior soma entre as similaridades encontradas para o conjunto de pares de entidades. Souza (2012) utiliza essa combinação mais elaborada que sugere melhores índices de precisão e cobertura. Ainda, seria interessante adaptar a meta-heurística para trabalhar com o alinhamento sem supervisão, ou seja, sem utilizar alinhamentos referência como base.

## A Apêndice

Neste capítulo se encontram as tabelas completas dos melhores resultados obtidos por cada modelo apresentado, cada tabela possui o identificador do teste em específico, o valor de precisão, cobertura, *f-measure* e o tempo coletado durante a execução do teste.



Tabela A.1: Resultados obtidos pelo modelo PP Padrão

| #Teste | Precisão | Cobertura | <i>f-measure</i> | Tempo (s) |
|--------|----------|-----------|------------------|-----------|
| 101    | 1,00     | 1,00      | 1,00             | 7.272     |
| 103    | 1,00     | 1,00      | 1,00             | 2.664     |
| 104    | 1,00     | 1,00      | 1,00             | 13.105    |
| 201    | 1,00     | 1,00      | 1,00             | 28.341    |
| 202    | 0,40     | 0,40      | 0,40             | 61.777    |
| 203    | 1,00     | 1,00      | 1,00             | 14.915    |
| 204    | 1,00     | 1,00      | 1,00             | 30.791    |
| 205    | 0,90     | 0,89      | 0,89             | 38.110    |
| 206    | 1,00     | 0,98      | 0,99             | 69.552    |
| 208    | 1,00     | 1,00      | 1,00             | 69.667    |
| 209    | 0,66     | 0,65      | 0,65             | 66.918    |
| 210    | 0,91     | 0,90      | 0,90             | 70.833    |
| 221    | 1,00     | 1,00      | 1,00             | 57.095    |
| 222    | 1,00     | 1,00      | 1,00             | 21.022    |
| 223    | 1,00     | 1,00      | 1,00             | 41.049    |
| 224    | 1,00     | 1,00      | 1,00             | 29.502    |
| 225    | 1,00     | 1,00      | 1,00             | 10.630    |
| 228    | 1,00     | 1,00      | 1,00             | 5.606     |
| 230    | 0,93     | 1,00      | 0,96             | 10.822    |
| 232    | 1,00     | 1,00      | 1,00             | 10.857    |
| 233    | 1,00     | 1,00      | 1,00             | 6.972     |
| 236    | 1,00     | 1,00      | 1,00             | 5.952     |
| 237    | 1,00     | 1,00      | 1,00             | 10.918    |
| 238    | 1,00     | 1,00      | 1,00             | 32.181    |
| 239    | 0,96     | 1,00      | 0,98             | 9.494     |
| 240    | 0,97     | 1,00      | 0,98             | 7.914     |
| 241    | 1,00     | 1,00      | 1,00             | 5.874     |
| 246    | 0,96     | 1,00      | 0,98             | 23.171    |
| 247    | 0,97     | 1,00      | 0,98             | 7.250     |
| 301    | 0,87     | 0,78      | 0,82             | 13.227    |
| 302    | 0,72     | 0,64      | 0,68             | 18.234    |
| 304    | 0,87     | 0,94      | 0,90             | 53.146    |
| Média: | 0,94     | 0,94      | 0,94             | 30.859    |

Tabela A.2: Resultados PP Simples

| #Teste | Precisão | Cobertura | <i>f-measure</i> | Tempo (s) |
|--------|----------|-----------|------------------|-----------|
| 101    | 1,00     | 1,00      | 1,00             | 8.899     |
| 103    | 1,00     | 1,00      | 1,00             | 34.198    |
| 104    | 1,00     | 1,00      | 1,00             | 15.486    |
| 201    | 1,00     | 1,00      | 1,00             | 37.321    |
| 202    | 0,59     | 0,59      | 0,59             | 78.413    |
| 203    | 1,00     | 1,00      | 1,00             | 16.371    |
| 204    | 1,00     | 1,00      | 1,00             | 19.410    |
| 205    | 1,00     | 0,98      | 0,99             | 33.941    |
| 206    | 1,00     | 0,98      | 0,99             | 70.556    |
| 208    | 0,97     | 0,97      | 0,97             | 68.837    |
| 209    | 0,64     | 0,63      | 0,63             | 70.041    |
| 210    | 0,53     | 0,52      | 0,52             | 72.847    |
| 221    | 0,92     | 0,92      | 0,92             | 47.700    |
| 222    | 0,93     | 0,93      | 0,93             | 30.019    |
| 223    | 0,87     | 0,87      | 0,87             | 44.094    |
| 224    | 1,00     | 1,00      | 1,00             | 55.525    |
| 225    | 0,90     | 0,90      | 0,90             | 25.180    |
| 228    | 1,00     | 1,00      | 1,00             | 9.765     |
| 230    | 0,90     | 0,97      | 0,93             | 23.589    |
| 232    | 1,00     | 1,00      | 1,00             | 9.665     |
| 233    | 1,00     | 1,00      | 1,00             | 16.348    |
| 236    | 1,00     | 1,00      | 1,00             | 5.788     |
| 237    | 1,00     | 1,00      | 1,00             | 27.581    |
| 238    | 0,95     | 0,95      | 0,95             | 26.407    |
| 239    | 0,96     | 1,00      | 0,98             | 22.011    |
| 240    | 0,97     | 1,00      | 0,98             | 8.108     |
| 241    | 1,00     | 1,00      | 1,00             | 5.493     |
| 246    | 0,96     | 1,00      | 0,98             | 10.844    |
| 247    | 0,85     | 0,87      | 0,86             | 7.565     |
| 301    | 0,70     | 0,63      | 0,66             | 12.723    |
| 302    | 0,72     | 0,64      | 0,68             | 63.782    |
| 304    | 0,87     | 0,94      | 0,90             | 8.971     |
| Média: | 0,91     | 0,92      | 0,91             | 30.859    |

Tabela A.3: Resultados PP Simples + BLI

| #Teste | Precisão | Cobertura | <i>f-measure</i> | Tempo (s) |
|--------|----------|-----------|------------------|-----------|
| 101    | 1,00     | 1,00      | 1,00             | 10.063    |
| 103    | 1,00     | 1,00      | 1,00             | 4.626     |
| 104    | 1,00     | 1,00      | 1,00             | 16.653    |
| 201    | 1,00     | 1,00      | 1,00             | 27.895    |
| 202    | 0,39     | 0,39      | 0,39             | 45.758    |
| 203    | 1,00     | 1,00      | 1,00             | 17.966    |
| 204    | 1,00     | 1,00      | 1,00             | 36.185    |
| 205    | 0,90     | 0,89      | 0,89             | 14.882    |
| 206    | 1,00     | 0,98      | 0,99             | 55.953    |
| 208    | 1,00     | 1,00      | 1,00             | 68.406    |
| 209    | 0,66     | 0,65      | 0,65             | 69.392    |
| 210    | 0,87     | 0,86      | 0,86             | 70.552    |
| 221    | 1,00     | 1,00      | 1,00             | 30.097    |
| 222    | 0,93     | 0,93      | 0,93             | 16.260    |
| 223    | 1,00     | 1,00      | 1,00             | 32.822    |
| 224    | 1,00     | 1,00      | 1,00             | 20.624    |
| 225    | 1,00     | 1,00      | 1,00             | 27.792    |
| 228    | 1,00     | 1,00      | 1,00             | 10.787    |
| 230    | 0,93     | 1,00      | 0,96             | 13.676    |
| 232    | 1,00     | 1,00      | 1,00             | 11.956    |
| 233    | 1,00     | 1,00      | 1,00             | 14.417    |
| 236    | 1,00     | 1,00      | 1,00             | 5.977     |
| 237    | 0,95     | 0,95      | 0,95             | 13.573    |
| 238    | 1,00     | 1,00      | 1,00             | 32.093    |
| 239    | 0,96     | 1,00      | 0,98             | 13.413    |
| 240    | 0,97     | 1,00      | 0,98             | 11.938    |
| 241    | 1,00     | 1,00      | 1,00             | 10.594    |
| 246    | 0,96     | 1,00      | 0,98             | 6.936     |
| 247    | 0,97     | 1,00      | 0,98             | 12.212    |
| 301    | 0,87     | 0,78      | 0,82             | 7.385     |
| 302    | 0,72     | 0,64      | 0,68             | 46.823    |
| 304    | 0,88     | 0,95      | 0,91             | 17.868    |
| Média: | 0,936    | 0,938     | 0,937            | 24.862    |

## Bibliografia

- AARTS, E. H.; LENSTRA, J. K. *Local search in combinatorial optimization*. [S.l.]: New Jersey: Princeton University Press, 2003.
- AKBARI, I.; FATHIAN, M.; BADIE, K. An improved mlma+ and its application in ontology matching. In: IEEE. *Innovative technologies in intelligent systems and industrial applications, 2009. CITISIA 2009*. [S.l.], 2009. p. 56–60.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. In: . [S.l.]: Scientific American, 2001.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data—the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, p. 205–227, 2009.
- BOCK, J.; HETTENHAUSEN, J. Discrete particle swarm optimisation for ontology alignment. *Information Sciences*, Elsevier, v. 192, p. 152–173, 2012.
- EUZENAT, J.; SHVAIKO, P. et al. *Ontology matching*. [S.l.]: Springer, 2007. v. 18.
- FARIA, D. et al. Oaei 2016 results of aml. In: *OM@ ISWC*. [S.l.: s.n.], 2016. p. 138–145.
- FARINELLI, F.; ALMEIDA, M. Interoperabilidade semântica em sistemas de informação de saúde por meio de ontologias formais e informais: um estudo da norma openehr. *XVII Encontro Nacional de Pesquisa em Ciência da Informação*, v. 17, n. 1, 2014.
- GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge acquisition*, Elsevier, v. 5, n. 2, p. 199–220, 1993.
- HOOS, H.; STUTZLE, T. *Empirical analysis of randomized algorithms*. [S.l.]: Handbook of Approximation Algorithms and Metaheuristics, 2007. 14.1–14.7 p.
- JOSLYN, C. A.; PAULSON, P.; WHITE, A. Measuring the structural preservation of semantic hierarchy alignments. In: CEUR-WS. ORG. *Proceedings of the 4th International Conference on Ontology Matching—Volume 551*. [S.l.], 2009. p. 61–72.
- LAMBRIX, P.; TAN, H. Sambo—a system for aligning and merging biomedical ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web*, Elsevier, v. 4, n. 3, p. 196–206, 2006.
- LAUMANN, M.; RUDOLPH, G.; SCHWEFEL, H.-P. A spatial predator-prey approach to multi-objective optimization: A preliminary study. In: SPRINGER. *Parallel Problem Solving from Nature—PPSN V*. [S.l.], 1998. p. 241–249.
- MARTINEZ-GIL, J.; ALDANA-MONTES, J. F. An overview of current ontology meta-matching solutions. *The Knowledge Engineering Review*, Cambridge University Press, v. 27, n. 4, p. 393–412, 2012.
- OTERO-CERDEIRA, L.; RODRÍGUEZ-MARTÍNEZ, F. J.; GÓMEZ-RODRÍGUEZ, A. Ontology matching: A literature review. *Expert Systems with Applications*, Elsevier, v. 42, n. 2, p. 949–971, 2015.

- SHI, Y.; EBERHART, R. A modified particle swarm optimizer. In: IEEE. *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on.* [S.l.], 1998. p. 69–73.
- SHVAIKO, P.; EUZENAT, J. Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 25, n. 1, p. 158–176, 2013.
- SOUZA, J. F. Uma abordagem heurística uni-objetivo para calibragem em meta-alinhadores de ontologias. 2012.
- SOUZA, J. F. et al. Análise de abordagens populacionais para meta-alinhamento de ontologias. In: *iSys-Revista Brasileira de Sistemas de Informação*. [S.l.: s.n.], 2014. p. 75–97.
- TILAHUN, S. L.; ONG, H. C. Prey-predator algorithm: A new metaheuristic algorithm for optimization problems. *International Journal of Information Technology & Decision Making*, World Scientific, v. 14, n. 06, p. 1331–1352, 2015.
- WANG, P.; WANG, W. Lily results for oaei 2016. In: *OM@ ISWC*. [S.l.: s.n.], 2016. p. 178–184.
- XUE, X.; TANG, Z. An evolutionary algorithm based ontology matching system. 2017.
- ZHANG, X. et al. Rimom results for oaei 2009. In: CEUR-WS. ORG. *Proceedings of the 4th International Conference on Ontology Matching-Volume 551*. [S.l.], 2009. p. 208–215.