

## 6 Análise dos resultados

Como pode ser observado no capítulo anterior, este trabalho alcançou um resultado intermediário nas médias gerais dos testes entre os meta-alinhadores GOAL e GOAM e superior ao meta-alinhador MaSiMe, conforme apresenta a figura 5.1.

Este capítulo apresenta uma discussão sobre resultados encontrados e pondera esses resultados em relação às demais características dos meta-alinhadores. Além disso, são analisadas as características dos meta-alinhadores apresentados no capítulo 2 e que não foram submetidos aos testes da OAEI.

### 6.1 Considerações quanto aos meta-alinhadores

O meta-alinhador GOAL, conforme visto na seção 2.4.4, utiliza as informações de precisão e cobertura na sua função de avaliação da população gerada. O indivíduo com melhor aptidão é aquele que gera uma solução com maior precisão e maior cobertura. Esse tipo de modelagem da solução se mostra muito eficaz para encontrar a melhor calibragem, conforme demonstra os resultados alcançados pelo GOAL. Por outro lado, essa abordagem se torna pouco prática, uma vez que é necessário conhecer todos os alinhamentos de referência para que se possa avaliar a aptidão de cada indivíduo.

Pode-se afirmar que o GOAL é adequado para utilização em um cenário onde um engenheiro de ontologias necessita, em algum caso de engenharia reversa, conhecer quais os pesos utilizados para conseguir um dado alinhamento de duas ontologias. Porém, num cenário onde um engenheiro de ontologias ou agentes de software necessitam criar um primeiro alinhamento entre duas ontologias, a utilização do sistema GOAL se torna inviável. A abordagem do GNoSIS+, por outro lado, é adequada para esse cenário onde não se tem conhecimento do alinhamento ideal entre duas ontologias. Como o GNoSIS+ utiliza pré-alinhamentos para calibrar as funções de similaridade, o sistema também se torna indicado para cenários onde já exista um arquivo de alinhamento entre duas ontologias e o engenheiro de ontologia deseja refazer esses

alinhamentos quando uma das ontologias sofreu alguma alteração durante o tempo. Mesmo com essa diferença de cenário de aplicação, o GNoSIS+ conseguiu atingir resultados bem próximos aos alcançados pelo GOAL. Ressalta-se que o GNoSIS+ pode ser utilizado para alinhar ontologias, mesmo sem fazer uso do meta-alinhamento.

O GNoSIS+ gerou resultados pouco melhores que o GAOM. Conforme descrito na seção 2.4.3, o GAOM aborda o problema de meta-alinhamento como um problema de otimização global de mapeamentos entre duas ontologias. Ou seja, o algoritmo genético é utilizado para alcançar uma solução aproximada ótima. Uma solução ótima para o GAOM é aquela que fornece o maior número de alinhamentos entre os conceitos das duas ontologias. Ou seja, dadas duas ontologias  $o$  e  $o'$ , a função de avaliação do GAOM tenta maximizar  $|M_{o,o'}|$ , sendo  $M_{o,o'}$  o conjunto final de correspondências de  $o$  para  $o'$  e  $|M|$  a cardinalidade do conjunto de correspondências. Essa cardinalidade será, no máximo, igual ao número de entidades de  $o$ .

Dessa forma, o GAOM consegue bons resultados nos testes em que o número de alinhamentos a serem gerados são iguais ao número de entidades da ontologia base. Porém, ao tratar o problema como uma otimização global dos mapeamentos, o algoritmo gera alinhamentos falsos positivos nos testes nos quais o número de alinhamentos a serem gerados são menores que o número de entidades de  $o$ . Esse comportamento do algoritmo genético faz com que o algoritmo alcance resultados inferiores nesses tipos de testes. Além disso, o GOAM representa os indivíduos como conjuntos de correspondências. Ou seja, o sistema não pode ser considerado um sistema de calibragem, pois não fornece para o usuário os pesos que devem ser aplicados às funções de similaridade para resolução do problema.

Por fim, um ponto importante para análise do funcionamento dos meta-alinhadores é o tempo necessário para encontrar os melhores valores dos pesos das funções de similaridade. Abordagens que objetivam ser utilizadas para alinhamentos em tempo de execução devem necessariamente realizar o processo de calibragem de forma rápida. O desempenho do processo de calibragem está diretamente relacionado ao desempenho das funções de similaridade que estão sendo calibradas. Considerando  $t_a(e, e')$  o tempo gasto pela função de similaridade  $a$  para calcular a similaridade de duas entidades  $e$  e  $e'$ , pode-se dizer que o tempo total para um algoritmo de alinhamento decidir quais os alinhamentos corretos baseados na função de similaridade  $a$  será de  $T_a = \sum_{i=1}^{|o|} \sum_{j=1}^{|o'|} t_a(e_i, e_j)$ , sendo  $|o|$  a quantidade de entidades da ontologia  $o$ . Assim, quanto menor a frequência de avaliação das funções de similaridade

durante o processo de calibragem, maior pode ser o desempenho da calibragem.

Os meta-alinhadores GOAL, MaSiMe, GOAM e a abordagem de Acampora et al (2012) (seção 2.4.6) possuem uma abordagem similar, uma vez que suas funções de aptidão necessitam processar todos os alinhamentos para cada indivíduo (cromossomo) da população. Assim, considerando  $match_i$  o  $i$ -ésimo alinhador de um vetor de alinhadores de entrada  $VM$  fornecido para o meta-alinhador, estes sistemas necessitam de um tempo de processamento igual a  $gi \sum_{i=1}^{|VM|} T_{match_i}$  para encontrar uma boa solução para o problema, sendo  $g$  e  $i$ , respectivamente, a quantidade de gerações e a quantidade de indivíduos gerados em cada geração do algoritmo genético.

O meta-alinhador proposto por Gínsca & Iftene (2010) (seção 2.4.5, por sua vez, realiza um pré-processamento das ontologias gerando todas as medidas de similaridade que possam ser utilizadas para alinhá-las. Este processo faz com que o tempo de processamento do algoritmo genético seja reduzido, uma vez que não é necessário chamar cada alinhador para cada avaliação dos indivíduos. Já a abordagem baseada em regras de perturbação do eTuner permite que os alinhadores sejam executados poucas vezes, uma vez que a ferramenta consegue identificar as perturbações necessárias para transformar um modelo em outro antes de decidir os alinhadores que serão utilizados. Por outro lado, a abordagem de Gínsca & Iftene (2010) utiliza como função objetivo medidas de qualidade do alinhamento gerado em relação ao alinhamento de referência. Tal restrição faz com que essa abordagem, tal qual as demais, não possam ser aplicadas em cenários nos quais não se conhece qual o alinhamento entre as ontologias. Já a abordagem do eTuner realiza uma busca em uma coleção de esquemas sintéticos gerados a partir da aplicação randômica de suas regras de perturbação. Por este fato, segundo Lee et al (2007), a estratégia de busca adotada pelo eTuner é adequada para a calibragem de esquemas de tamanho médio a moderado e em tempo de projeto somente.

Por essa razão, esses meta-alinhadores não são adequados para cenários em que o tempo de resposta é importante como no problema de alinhamento de grandes ontologias<sup>1</sup> (Stoffel et al, 1997; Hahn et al, 2004; Algergawy et al, 2011; Rahm, 2011) ou na resolução *on-the-fly* de incompatibilidade de comunicação entre agentes de software (Souza et al, 2009c, 2010b, 2012b), quando a ação de um agente depende da resposta de um alinhamento entre sua ontologia de domínio com a de outro agente. Por outro lado, o GNoSIS+ pode ser aplicado nesses cenários, uma vez que os alinhadores somente são executados uma vez para cada alinhamento de entrada fornecido pelo especialista. Ou

<sup>1</sup>Very Large Ontologies – VLOs

seja, para encontrar uma solução para os pesos, o GNoSIS+ necessita de um tempo de processamento próximo a  $\sum_{i=1}^{|A|} \sum_{j=1}^{|VM|} t_{match_i}(e_{ij}, e'_{ij})$ , sendo  $|A|$  a quantidade de alinhamentos fornecidos. Vale ressaltar que foram alcançados bons resultados nos testes mesmo para  $|A| \ll |o|$ .

## 6.2

### Considerações quanto à base de teste

Os alinhamentos de referência dos testes 3xx representam dois tipos de correspondências: igualdade (denotados por =) e de especialização (denotados por <). Para essa versão do GNoSIS+, a ferramenta somente gera correspondências de igualdade. Por esse motivo, a cobertura dos testes 3xx foram menores em relação aos demais testes.

Outro ponto importante a ser considerado são as funções de similaridade utilizadas durante os testes. Diferentes funções de similaridade podem gerar resultados distintos nesses testes. Optou-se por utilizar sete funções de similaridade bem simples mas que abrangessem os diferentes tipos de mudanças que uma ontologia pode sofrer. Dessa forma, fica mais evidente a necessidade de uma calibragem correta para verificar a relevância de cada função de similaridade no cômputo do alinhamento entre dois modelos. O sistema GOAL e MaSiMe, por sua vez, optaram em utilizar quatro algoritmos: SIFO (Martinez et al, 2008), Levenshtein (1966), Stoilos et al (2005) e QGrams (Ukkonen, 1992). Talvez a falta de um algoritmos para analisar indivíduos tenha feito com que o GOAL e MaSiMe não alcançassem a mesma cobertura que o GNoSIS+ nos testes 205 e 301, os quais possuem maior diferença entre os termos de entidades e falta de comentários. Além disso, os testes com maiores depurações na estrutura da ontologia não foram divulgados pelo GOAL e MaSiMe, o que inviabilizou a comparação com os resultados do GNoSIS+. O algoritmo GOAM, por sua vez, se utiliza de apenas duas funções de similaridade, uma para tratar *características intensionais* (nomes, propriedades e instâncias) e outra para tratar *características extensionais* (relações entre conceitos), ambas definidas pelo próprio autor.

Essa diferença entre as funções de similaridade utilizadas por cada meta-alinhador compromete a comparação entre os resultados. Uma melhor abordagem para comparação dos resultados seria utilizar as mesmas funções de similaridade nos três meta-alinhadores. Porém, o código desses trabalhos não foram fornecidos pelos seus autores.

Quanto ao algoritmo genético, utilizou-se o mesmo número de gerações que o GOAL. Vale ressaltar que, embora o AG converja para uma solução do

sistema, essa não é a melhor solução. Para conseguir uma melhor solução, pode-se verificar dois elementos: o número de gerações do AG e o número de exemplos fornecidos. Ao aumentar o número de gerações do algoritmo genético, pode-se alcançar uma melhor solução. Porém, para resolver o problema proposto pelo *benchmark* não é necessário alcançar uma melhor solução, visto que a diferença entre a solução em gerações seguintes e a solução ótima chega a ser de menos de dois décimos, o que não foi significativo no cômputo dos alinhamentos. Por outro lado, alimentando o algoritmo com mais exemplos de alinhamentos conhecidos fará com que o número de soluções possíveis seja melhor e, com isso, uma melhor medida-F poderia ser alcançada. Contudo, aumentando-se consideravelmente o número de exemplos faria com que o algoritmo aumentasse também o tempo de processamento, o que não era o objetivo deste trabalho. Com os testes realizados, pôde-se demonstrar que a proposta atinge bons resultados mesmo com um número muito pequeno de exemplos fornecidos pelos especialistas.

### 6.3

#### Considerações finais

Pode-se comparar meta-alinhadores levando em consideração outros aspectos, além da corretude do seu resultado e do seu desempenho. Optou-se por comparar os meta-alinhadores seguindo os critérios e a estratégia proposta por Martinez-Gil & Aldana-Montes (2012). Dada a dificuldade em comparar esse tipo de ferramenta, o autor propõe uma categorização *fuzzy* com uma escala de 1 a 5. Mesmo sendo uma avaliação superficial, é possível comparar as ferramentas à partir dos resultados apresentados na literatura e nas características apontadas pelos autores. São cinco os critérios a serem comparados:

- Resultados: Este critério diz respeito à média harmônica dos resultados encontrados no *benchmark* da OAEI. Embora nem todos os pesquisadores disponibilizaram todos os seus resultados, com isso, possam existir discrepâncias na média final, o objetivo aqui não é realizar um *ranking* de ferramentas, mas entender o escopo de cada abordagem.
- Desempenho: Este critério está relacionado ao tempo necessário para obter os resultados. Como os pesquisadores não disponibilizam as ferramentas para testes de terceiros, não é intenção deste trabalho medir o tempo de cada ferramenta. Contudo, é de conhecimento que o tempo necessário para chegar a um resultado depende da estratégia utilizada (força bruta, aproximação heurística, etc), além da forma de representação e avaliação da solução.

- Treinamento: Uma das desvantagens das ferramentas de meta-alinhamento de ontologias é a quantidade de esforço necessário para treinar a ferramenta. Embora essa seja uma característica inerente de ferramentas de aprendizado de máquina, também é considerado o tempo necessário para otimizar as ferramentas heurísticas.
- Comportamento do usuário: A importância do especialista que opera meta-alinhadores está na capacidade de ajustar corretamente a relevância de cada alinhador para alcançar um melhor resultado. Este critério tenta medir se é possível simular o comportamento do especialista que treina a ferramenta. Ou seja, se a ferramenta é capaz de se comportar como o usuário e de retornar com a relevância desejada para cada alinhador.
- Classificação (*Ranking generation*): É possível construir *rankings* para alinhadores utilizando técnicas de meta-alinhamento. Isto é possível pois muitas técnicas de meta-alinhamento são capazes de identificar uma aptidão numérica para cada combinação de alinhadores. Assim, pode-se gerar um *ranking* de abordagens alternativas.

A figura 6.1 apresenta o resultado da comparação dos três meta-alinhadores.

A modelagem do problema de meta-alinhamento definida nos meta-alinhadores GOAM, eTuner e na abordagem de Acampora et al (2012) não resultam um conjunto de pesos. Por esse motivo, estas abordagens receberam um valor menor para o critério classificação. Já as demais abordagens retornam um conjunto de pesos, sendo possível identificar a relevância de cada alinhador no alinhamento final. O GNoSIS+, diferente dessas abordagens, acrescenta a possibilidade de permitir, na entrada do problema, que composições de funções sejam definidas, o que facilita na identificação da relevância dos alinhadores.

A nota de resultados dos meta-alinhadores foi definida de acordo com os resultados alcançados no *benchmark* da OAEI. A exceção está nas abordagens de Acampora et al (2012), Gînsca & Iftene (2010) e na ferramenta eTuner, pois os mesmos não avaliaram suas propostas com o *benchmark* da OAEI ou não divulgaram os resultados alcançados. Para essas abordagens, foi então atribuída a nota máxima de resultados.

Em relação à simulação do comportamento do usuário, o eTuner, MaSiMe, GOAM e a abordagem de Acampora et al (2012) não simulam o comportamento de um usuário especialista, isto é, não é possível parar a simulação para retornar os valores que melhor configuram o sistema até aquele momento. O eTuner, contudo, permite simular o comportamento das regras de

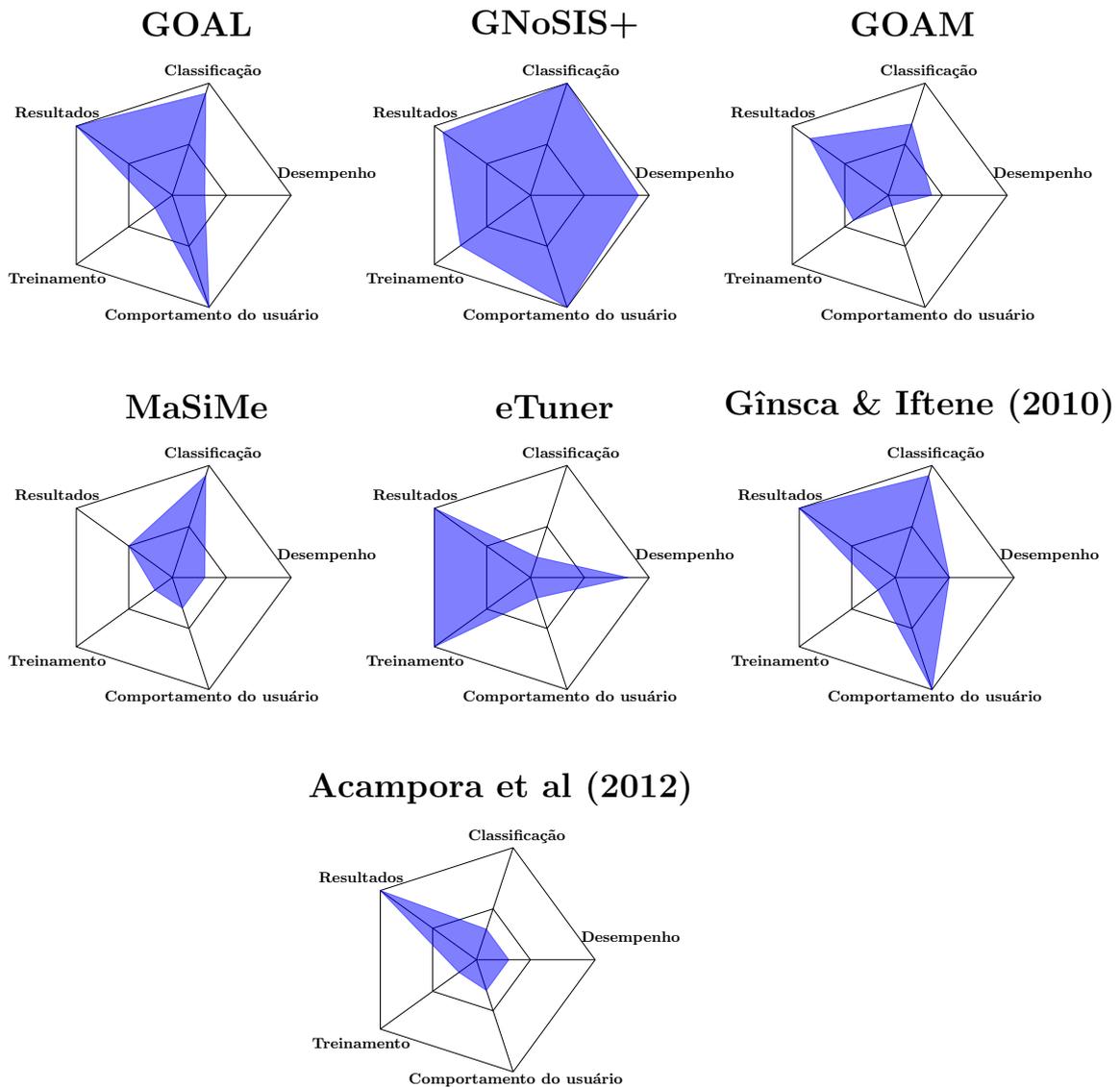


Figura 6.1: Comparação entre os meta-alinhadores heurísticos

perturbação. Os demais sistemas permitem a simulação de comportamento do usuário.

O desempenho dos meta-alinhadores está relacionado com a forma com que o meta-alinhador está modelado para avaliar a qualidade de suas soluções intermediárias. Os meta-alinhadores GOAL, GOAM, MaSiMe e a proposta de Acampora et al (2012) necessitam computar ou avaliar todos os alinhamentos para cada indivíduo da população em cada geração. Esse processo é extremamente custoso. A abordagem de Gínsca & Iftene (2010), por sua vez, tenta minimizar esse problema realizando um pré-processamento de todas as medidas de similaridades possíveis entre todas as combinações de pares de entidades. Contudo, a função objetivo dessa proposta necessita avaliar, para cada indivíduo, a qualidade da sua representação através de medidas de qualidade como

precisão e cobertura, a qual demanda também certo processamento, principalmente em modelos com muitas entidades. O eTuner e o GNoSIS+, por sua vez, possuem um bom desempenho. O GNoSIS+ possui uma função objetivo que não utiliza medidas de qualidade de alinhamento e, por isso, é rápida. Além disso, o GNoSIS+ necessita executar cada alinhador somente uma vez para cada correspondência fornecida pelo usuário especialista. Uma vez computados esses valores (e considerando que o número de correspondências fornecidas é muito menor que o número de entidades das ontologias), o GNoSIS+ se comporta como a abordagem de Gínsca & Iftene (2010), a qual não necessita recomputar as similaridades. O eTuner, por fim, também não utiliza medidas de qualidade de alinhamentos para calcular seu alinhamento final e não necessita executar os alinhadores para cada esquema sintético produzido, somente para um subconjunto pequeno de esquemas sintéticos.

Em relação à necessidade de treinamento, é considerado o treinamento necessário para preparar os dados para o início do meta-alinhamento, como a necessidade de pré-processamento, a necessidade de dados de referência para guiar o algoritmo ou a verificação dos modelos durante o processo. Os meta-alinhamentos GOAL e MaSiMe, bem como as abordagens de Acampora et al (2012) e Gínsca & Iftene (2010) necessitam de alinhamentos de referência para guiar o comportamento do algoritmo. O comportamento do meta-alinhador GOAM é guiado por uma função objetivo que verifica uma solução e as duas ontologias de entrada para cada indivíduo da população. O GNoSIS+ e o eTuner não necessitam de modelos de referência para guiar o comportamento do algoritmo e não necessitam de uma fase de pré-processamento. Assim, estas abordagens receberam nota máxima neste critério. Contudo, os autores do eTuner ressaltam que sua abordagem somente é aplicável em tempo de projeto, pois a quantidade de entidades das ontologias pode diminuir consideravelmente o desempenho do sistema.