

5 Avaliação do meta-alinhador

A avaliação de uma estratégia de meta-alinhamento consiste na avaliação do seu alinhamento retornado. Existem várias formas de avaliar um alinhamento como, por exemplo, (1) através de medidas de desempenho para verificar o quão satisfatória a estratégia está em termos computacionais, (2) através de métricas relacionadas ao usuário que verificam a satisfação do usuário, (3) medidas de qualidade para avaliar a qualidade dos alinhamentos identificados ou (4) medidas relacionadas à alguma tarefa para verificar o quão bom foi o alinhamento para certos casos de uso ou aplicações.

Para avaliação da abordagem, são descritos neste capítulo os testes realizados com o *benchmark* da Ontology Alignment Evaluation Initiative (OAEI) ¹, a qual fornece uma base de teste para verificar a conformidade dos alinhamentos gerados em relação aos alinhamentos esperados. A OAEI realiza campanhas anuais com o objetivo de avaliar tecnologias para combinação de ontologias. O *benchmark* foi utilizado para se ter um ponto de referência em relação aos demais algoritmos de alinhamento publicados na literatura, visto que o *benchmark* da OAEI é utilizado por diversos autores para avaliar os resultados das suas técnicas de alinhamento. Na seção 5.1, é realizada uma descrição da base de testes e dos resultados esperados. Por fim, os parâmetros utilizados para os testes e os resultados obtidos são apresentados nas seções 5.2 e 5.3, respectivamente.

5.1 Descrição dos testes

O *benchmark* utilizado para os testes consiste em uma ontologia de referência para o domínio de referências bibliográficas. A ontologia de referência, descrita em OWL-DL e serializada em formato RDF/XML, contém 33 classes nomeadas, 24 propriedades de objeto, 40 propriedades de dados, 56 indivíduos nomeados e 20 indivíduos anônimos, além de fazer referência para recursos externos para expressar informações não bibliográficas. Entre estes, estão re-

¹<http://oaei.ontologymatching.org/>

cursos do FOAF² e iCalendar³ para expressar conceitos de pessoa, organização e evento.

Cada teste consiste em analisar uma segunda ontologia (distinta em cada teste) e gerar um arquivo de alinhamento que contenha o maior número de alinhamentos descritos em um arquivo de alinhamento de referência disponibilizado para avaliar a qualidade do alinhamento. Os testes são sistematicamente gerados à partir de degradações da ontologia de referência, descartando ou alterando uma certa quantidade de informações com o objetivo de avaliar como um algoritmo se comporta quando estas informações estão ausentes.

Especificamente para o problema de meta-alinhamento de ontologias, estes testes são adequados pois cada teste exige que o meta-alinhador determine um peso maior para os algoritmos que podem ter uma relevância maior no resultado final do alinhamento e, por outro lado, descarte ou atribua um peso menor para os algoritmos que não serão determinantes para o alinhamento.

Os testes são divididos em 3 grandes categorias. Os testes iniciados por 1 (ou 1xx) não possuem alterações significantes na formalização dos conceitos; os testes iniciados por 2 (ou 2xx) consistem em alterações sistemáticas na formalização dos conceitos; e os testes iniciados por 3 (ou 3xx) consistem em comparações com ontologias externas. A tabela 5.1 resume qual a informação alterada ou suprimida da ontologia de referência. São seis elementos que podem ser alterados (OAEI, 2011):

- Nomes: nome de entidades podem ser trocados por caracteres randômicos, sinônimos, com uma convenção distinta ou traduzidos para outro idioma.
- Comentários: comentários podem ser suprimidos ou traduzidos para outro idioma.
- Hierarquia: pode ser suprimida, expandida ou reduzida.
- Instâncias: podem ser suprimidas.
- Propriedades: podem ser suprimidas ou ter suas restrições descartadas.
- Classes: podem ser expandidas, ou seja, relacionadas com várias outras classes, ou reduzidas.

Para avaliar a conformidade dos alinhamentos gerados em relação aos alinhamentos esperados, o *benchmark* da OEAI utiliza métricas da área de Recuperação da Informação (Baeza-Yates & Ribeiro-Neto, 2011), tais como

²<http://xmlns.com/foaf/0.1/>

³<http://www.w3.org/2002/12/cal/>

Tabela 5.1: Tabela com o resumo dos testes realizados

Teste	Resumo
101	Ontologias idênticas
102	Ontologia irrelevante
103	Generalização de linguagem
104	Restrição de linguagem
201	Nome de entidades ausentes
202	Entidades com nomes aleatórios e comentários ausentes
203	Comentários ausentes
204	Convenção de nomenclatura distinta
205	Sinônimos para nome de entidades
206	Idioma distinto
207	Nomes de entidades em outro idioma
208	Convenção de nomenclatura distinta e comentários ausentes
209	Sinônimos para 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	Uma ontologia real sobre bibliografia: BibTeX/MIT
302	Uma ontologia real sobre bibliografia: BibTeX/UMBC
303	Uma ontologia real sobre bibliografia: BibTeX/Karlsruhe
304	Uma ontologia real sobre bibliografia: BibTeX/INRIA

precisão (*precision*), cobertura (*recall*), medida-F (*F-measure*) e a taxa *fall-out*.

Definição 5.1 (Precisão) Dado um alinhamento de referência A e um alinhamento obtido B , a precisão é uma medida da habilidade de fornecer alinhamentos corretos dentre os alinhamentos obtidos, calculada como:

$$\frac{|A \cap B|}{|B|} \quad (5-1)$$

Definição 5.2 (Cobertura) Dado um alinhamento de referência A e um alinhamento obtido B , a cobertura é uma medida da habilidade de fornecer alinhamentos corretos dentre o total de alinhamentos de referência, calculada como:

$$\frac{|A \cap B|}{|A|} \quad (5-2)$$

Definição 5.3 (Medida-F) Dada uma precisão P e uma cobertura R , a medida-F é a média harmônica entre P e R dada por

$$\frac{2PR}{P + R} \quad (5-3)$$

Definição 5.4 (Fall-out) Dado um alinhamento de referência A e um alinhamento obtido B , o *fall-out* é a razão entre fornecer alinhamentos incorretos dentre o total de alinhamentos obtidos, calculada como:

$$\frac{|B| - |A \cap B|}{|B|} \quad (5-4)$$

5.2

Configuração utilizada para realização dos testes

Para realização dos testes com o *benchmark* da OAEI, os dois módulos principais do sistema foram configurados: o módulo de criação de funções compostas e o módulo de calibragem dos pesos. Os testes foram realizados em uma máquina com processador Intel Core 2 Duo 2.1Ghz e 4Gb RAM, em ambiente Ubuntu 11.10 64 bits.

5.2.1

Configuração do módulo de criação de funções compostas

O sistema foi configurado para utilizar funções de similaridade simples pré-programadas, ao invés de fazer uso de algoritmos de alinhamento mais eficientes propostos pela literatura. Foram utilizadas funções mais simples onde cada função avalia a similaridade a partir de heurísticas simples como nome de conceitos, interseção de indivíduos, relação entre classes a partir de propriedades, etc. Cada função implementa uma heurística que analisa um único tipo de entidade das ontologias. Assim, torna-se mais clara a avaliação da adaptação da calibragem de pesos para cada função, visto que os testes da OAEI são realizados com depurtações sistemáticas no modelo onde, para resolver cada depurtação, é necessária maior ênfase na análise de certos tipos de entidades, ao invés de outros. Caso fossem utilizadas funções de similaridade mais complexas, isto é, que analisassem diferentes tipos de entidades ao mesmo tempo, pode ficar pouco evidente o porquê de cada função receber o peso que foi atribuído pelo algoritmo genético. Além disso, ao se utilizar funções simples, o tempo de processamento do sistema diminui e pode-se demonstrar que, uma vez tendo em mãos um conjunto de algoritmos que cubram as possíveis mudanças que o modelo pode sofrer, a relevância do meta-alinhador está em descobrir como utilizar esses algoritmos em conjunto para obter o melhor resultado.

Os testes foram realizados com 7 funções de similaridade. As funções foram utilizadas para analisar, em ordem, nome de entidades, instâncias, comentários, parentescos, especializações, domínio e contradomínio de propriedades e propriedade de classes. As funções para analisar hierarquias e as funções para analisar propriedades foram agrupadas em funções compostas chamadas, respectivamente, de `FuncoesHierarquia` e `FuncoesPropriedades`. A configuração utilizada está disponível para leitura no apêndice B.

5.2.2

Configuração do módulo de calibragem dos pesos

Os testes foram realizados com o módulo de calibragem configurado para gerar 300 indivíduos na geração inicial e rodar por 5 gerações. O problema foi discretizado para uma granularidade de 0,025. As taxas do algoritmo genético foram:

- Porcentagem de seleção: 50%
- Taxa de cruzamento: 80%
- Taxa de mutação: 10%

- Porcentagem de manutenção dos melhores indivíduos: 30%
- Porcentagem de manutenção dos piores indivíduos: 20%
- Periodicidade da busca local: 1 vez a cada 5 gerações. Ou seja, a busca local foi realizada somente para avaliação da solução final do AG.

Para alimentar o módulo de calibragem foi utilizado um arquivo de pré-alinhamento, o qual armazena alinhamentos já conhecidos. Para cada teste, foram utilizados arquivos de alinhamento contendo 3 alinhamentos com grau de similaridade igual a 1. Um exemplo de arquivo de pré-alinhamento e o arquivo de configuração do módulo de calibragem estão disponível para leitura no apêndice B.

5.3

Resultados obtidos

Nesta seção, são apresentados os resultados alcançados pelo GNoSIS+ e os resultados divulgados pelos meta-alinhadores GAOM, GOAL e MaSiMe. Os demais meta-alinhadores apresentados na seção 2.3 não utilizaram a base de teste da OEAI ou não divulgaram seus resultados.

A tabela 5.2 apresenta os resultados do GNoSIS+ para cada teste.

A tabela 5.3 apresenta a média harmônica alcançada total, agrupada por tipo de teste.

A tabela 5.4 faz uma comparação entre os resultados do GNoSIS+ e os alcançados pelo meta-alinhador GOAL. A tabela 5.5, por sua vez, faz uma comparação entre os resultados do GNoSIS+ e os alcançados pelo meta-alinhador MaSiMe. Para os demais meta-alinhadores, a comparação não é realizada teste a teste, pois os autores dessas abordagens não divulgaram seus resultados detalhados. As células marcadas como N/D indicam que os resultados do teste não foram divulgados pelo(s) autor(es).

A tabela 5.6 compara a média harmônica dos meta-alinhadores GNoSIS+, GOAL, GAOM e MaSiMe.

A figura 5.1 apresenta um gráfico triangular com a cobertura e precisão totais dos quatro meta-alinhadores.

Tabela 5.2: Resultados encontrados no *benchmark*

Teste	Precisão	Cobertura	Fall-out	Medida-F
101	1,00	1,00	0,00	1,00
102	N/A	N/A	N/A	N/A
103	1,00	1,00	0,00	1,00
104	1,00	1,00	0,00	1,00
201	1,00	1,00	0,00	1,00
202	0,63	0,63	0,37	0,63
203	1,00	1,00	0,00	1,00
204	1,00	1,00	0,00	1,00
205	1,00	0,98	0,00	0,99
206	0,97	0,98	0,03	0,97
207	0,98	0,97	0,02	0,97
208	0,97	0,97	0,03	0,97
209	0,81	0,80	0,19	0,81
210	0,87	0,87	0,13	0,87
221	1,00	1,00	0,00	1,00
222	1,00	1,00	0,00	1,00
223	1,00	1,00	0,00	1,00
224	1,00	1,00	0,00	1,00
225	1,00	1,00	0,00	1,00
228	1,00	1,00	0,00	1,00
230	0,94	1,00	0,06	0,97
232	1,00	1,00	0,00	1,00
233	1,00	1,00	0,00	1,00
236	1,00	1,00	0,00	1,00
237	1,00	1,00	0,00	1,00
238	1,00	1,00	0,00	1,00
239	1,00	1,00	0,00	1,00
240	1,00	1,00	0,00	1,00
241	1,00	1,00	0,00	1,00
246	1,00	1,00	0,00	1,00
247	1,00	1,00	0,00	1,00
301	0,87	0,79	0,13	0,83
302	1,00	0,71	0,00	0,83
303	0,75	0,84	0,25	0,79
304	0,88	0,95	0,12	0,91

Tabela 5.3: Média harmônica dos resultados encontrados no *benchmark*

Teste	Precisão	Cobertura	Fall-out	Medida-F
1xx	1,00	1,00	0,00	1,00
2xx	0,96	0,96	0,04	0,96
3xx	0,86	0,83	0,14	0,85
Todos	0,96	0,96	0,04	0,96

PUC-Rio - Certificação Digital Nº 0821406/CA

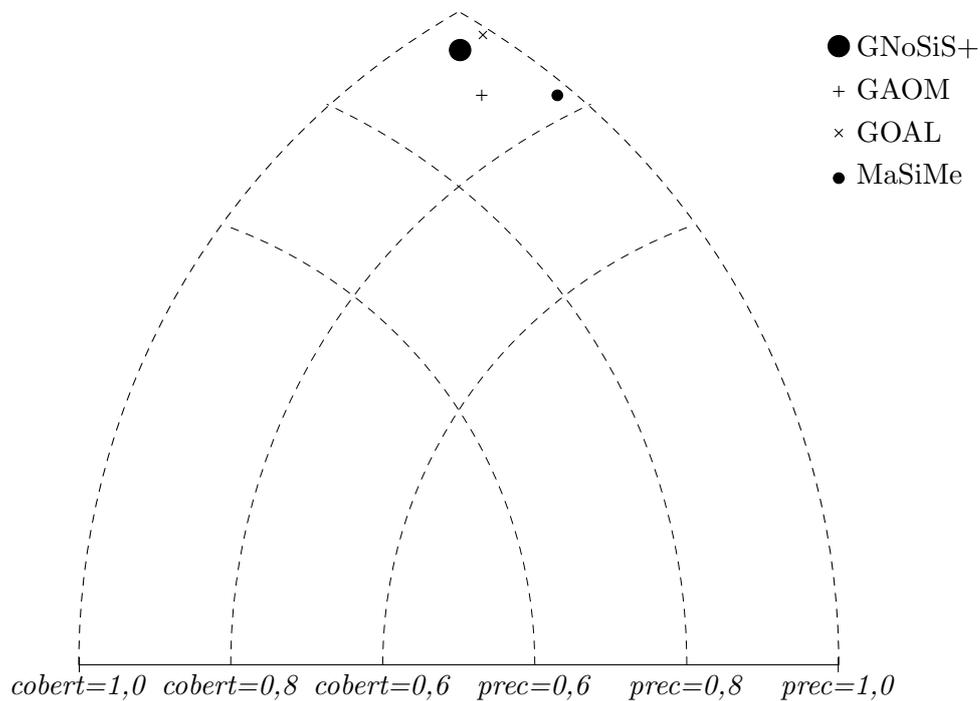


Figura 5.1: Gráfico triangular com cobertura e precisão totais

Tabela 5.4: Comparação dos resultados com o meta-alinhador GOAL

Teste	Precisão		Cobertura		Fall-out		Medida-F	
	GNoSIS+	GOAL	GNoSIS+	GOAL	GNoSIS+	GOAL	GNoSIS+	GOAL
101	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
102	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
103	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
104	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
201	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
202	0,63	1,00	0,63	1,00	0,37	0,00	0,63	1,00
203	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
204	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
205	1,00	1,00	0,98	0,71	0,00	0,06	0,99	0,83
206	0,97	1,00	0,98	1,00	0,03	0,00	0,97	1,00
207	0,98	N/D	0,97	N/D	0,02	N/D	0,97	N/D
208	0,97	N/D	0,97	N/D	0,03	N/D	0,97	N/D
209	0,81	N/D	0,80	N/D	0,19	N/D	0,81	N/D
210	0,87	N/D	0,87	N/D	0,13	N/D	0,87	N/D
221	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
222	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
223	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
224	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
225	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
228	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
230	0,94	N/D	1,00	N/D	0,06	N/D	0,97	N/D
232	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
233	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
236	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
237	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
238	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
239	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
240	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
241	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
246	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
247	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
301	0,87	0,90	0,79	0,69	0,13	0,07	0,83	0,78
302	1,00	N/D	0,71	N/D	0,00	N/D	0,83	N/D
303	0,75	N/D	0,84	N/D	0,25	N/D	0,79	N/D
304	0,88	N/D	0,95	N/D	0,12	N/D	0,91	N/D

Tabela 5.5: Comparação dos resultados com o meta-alinhador MaSiMe

Teste	Precisão		Cobertura		Fall-out		Medida-F	
	GNoSIS+	MaSiMe	GNoSIS+	MaSiMe	GNoSIS+	MaSiMe	GNoSIS+	MaSiMe
101	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
102	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
103	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
104	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
201	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
202	0,63	1,00	0,63	1,00	0,37	0,00	0,63	1,00
203	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
204	1,00	1,00	1,00	0,91	0,00	0,00	1,00	0,95
205	1,00	1,00	0,98	0,19	0,00	0,00	0,99	0,33
206	0,97	1,00	0,98	0,19	0,03	0,00	0,97	0,33
207	0,98	N/D	0,97	N/D	0,02	N/D	0,97	N/D
208	0,97	N/D	0,97	N/D	0,03	N/D	0,97	N/D
209	0,81	N/D	0,80	N/D	0,19	N/D	0,81	N/D
210	0,87	N/D	0,87	N/D	0,13	N/D	0,87	N/D
221	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
222	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
223	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
224	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
225	1,00	1,00	1,00	1,00	0,00	0,00	1,00	1,00
228	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
230	0,94	N/D	1,00	N/D	0,06	N/D	0,97	N/D
232	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
233	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
236	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
237	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
238	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
239	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
240	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
241	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
246	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
247	1,00	N/D	1,00	N/D	0,00	N/D	1,00	N/D
301	0,87	0,93	0,79	0,23	0,13	0,06	0,83	0,37
302	1,00	N/D	0,71	N/D	0,00	N/D	0,83	N/D
303	0,75	N/D	0,84	N/D	0,25	N/D	0,79	N/D
304	0,88	N/D	0,95	N/D	0,12	N/D	0,91	N/D

Tabela 5.6: Comparação entre a média harmônica dos meta-alinhadores GNoSIS+, GAOM, GOAL e MaSiMe

		GNoSIS+	GAOM	GOAL	MaSiMe
Testes 1xx	Precisão	1,00	1,00	1,00	1,00
	Cobertura	1,00	1,00	1,00	1,00
	Medida-F	1,00	1,00	1,00	1,00
Testes 2xx	Precisão	0,96	0,92	N/D	N/D
	Cobertura	0,96	0,80	N/D	N/D
	Medida-F	0,96	0,86	N/D	N/D
Testes 3xx	Precisão	0,86	0,89	N/D	N/D
	Cobertura	0,83	0,82	N/D	N/D
	Medida-F	0,85	0,85	N/D	N/D
Média total	Precisão	0,96	0,94	0,99	0,99
	Cobertura	0,96	0,87	0,96	0,82
	Medida-F	0,96	0,90	0,97	0,83